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

Pedro Barroso[†]

Konark Saxena[‡]

A subsequent version of this paper is forthcoming at the RFS

First version: December 2015

This version: January 2019

[*]The present work has benefited from comments and suggestions of Amit Goyal, Andrew Detzel, Daniel Smith, David Feldman, Filippo Massari, Grigory Vilkov, Iliyan Georgiev, José Corrêa Guedes, Jun Tu, Li Yang, Lorenzo Garlappi, Raymond Kan, Paul Karehnke, Rossen Valkanov, Pedro Santa-Clara, Victor DeMiguel, and Xing Jin. We thank in particular the discussants Dawei Fang, Luis Vicente, Martin Young, and Ralf Kellner for their suggestions and Raymond Kan, and Michael Wolf for sharing their Matlab code. We also thank numerous seminar participants at Australian National University, Católica-Lisbon School of Business and Economics, Deakin University, Griffith University, Nova SBE, Queensland University of Technology, Schroders' Official Institutions Seminar (Sydney), Singapore Management University, University of Queensland, University of Technology Sydney, and the European Financial Management Association, Financial Management Association (FMA) Asia, FMA Europe, and Portuguese Finance Network conferences in 2016. Pedro Barroso acknowledges the support from FCT – Portuguese Foundation of Science and Technology for the project UID/GES/00407/2019. Note: This is only a pre-print version of the paper. The final version of this article is available at the website of the Review of Financial Studies at https://doi.org/10.1093/rfs/hhab041.

[†]Universidade Católica Portuguesa, Católica-Lisbon School of Business and Economics, Portugal. E-mail: pedro.barroso@ucp.pt

[‡]University of New South Wales. E-mail: k.saxena@unsw.edu.au

Abstract

Portfolio optimization often struggles in realistic out-of-sample contexts. We de-construct this stylized fact, comparing historical forecasts of portfolio optimization inputs with subsequent out of sample values. We confirm that historical forecasts are imprecise guides of subsequent values but also find the resulting forecast errors are not entirely random. They have predictable patterns and can be partially reduced using their own history. Learning from past forecast errors to calibrate inputs (akin to empirical Bayesian learning) results in portfolio performance that reinforces the case for optimization. Furthermore, the portfolios achieve performance that meets expectations, a desirable yet elusive feature of optimization methods.

JEL classification: G11; G12; G17.

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

1 Introduction

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

George Santayana, *The Life of Reason*

Laying the foundations of portfolio theory, Markowitz (1952) shows how an investor with mean-variance utility should form portfolios given the expected returns and covariance matrix of a set of assets. 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)). Several articles propose optimized approaches to account for estimation error when constructing portfolios (see e.g. Jorion (1986), Frost and Savarino (1986)), and Kan and Zhou (2007)). However, DeMiguel et al. (2009b) show that in a demanding out-of-sample (OOS) environment such optimized portfolios often struggle.[1] OOS testing thus offers a meaningful challenge to the literature on portfolio management.

The motivation for this paper starts from observing that, in these OOS tests, historical OOS forecast errors (realizations minus forecasts; hereafter simply "OOS errors") of portfolio optimization inputs such as future means, variances, and correlations are typically 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. In this study, we examine this possibility and propose

[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 our 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.

a flexible correction to portfolio optimization inputs that has the potential to attenuate many of the well known limitations of portfolio optimization. The correction only uses past OOS errors that could be known in real time, so it is still an OOS method. It is based on a simple intuition: learn from past OOS forecast errors to improve subsequent forecasts. Once given the corrected inputs, the plain Markowitz method works quite well OOS.[2] Applied consistently to the inputs of Markowitz optimization, the correction exhibits robust performance in a wide range of tests, compared to a representative set of state-of-the-art methods.

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 optimization inputs.

[Insert figure 1 near here]

Suppose an investor estimates in a historical sample of 60 months the variances, mean returns, pairwise correlations and covariances. As the figure shows, all of these inputs regress considerably to the mean in the subsequent out of sample period.[3] Some of these patterns are not surprising. The mean regression of expected returns, for instance, is known at least since Bondt and Thaler (1985). But surprisingly, a common approach to portfolio optimization relies exactly on plugging historical estimates directly in the optimization problem – the "plug-in" approach. Graphically, this amounts to expect the ex post values to lie on the 45 degree line with respect to 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 are even negatively related (mean returns). These stylized facts

[2] Other methods that improve on 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. We simply show that this classic implementation of portfolio optimization, with all of its known limitations, performs reasonably well with inputs corrected for past OOS errors.

[3] In unreported results, available upon request, we do the same for an investor using the Fama and French (1993) factor model. The pattern for betas, alphas, and residual covariances also shows strong mean reversion.

suggest a correction specific to each input based on its past reliability. We call this the Galton correction after Sir Francis Galton, who first proposed the concept of regression to the mean (Galton (1894)). Our method minimizes the OOS forecast error of subsequent realizations of inputs using their historical estimates. It infers a workable correction from a grand historical 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. This correction provides a simple and flexible method to estimate the covariance matrix and expected returns of a given set of stocks.

While an extensive research documents the existence of large OOS errors, to the best of our knowledge this is the first paper that makes explicit use of those errors to improve the portfolio allocation decision. Using these errors to correct portfolio optimization inputs alleviates considerably the most discouraging stylized results of portfolio optimization.

This correction has an intuitive empirical Bayes interpretation. It corresponds to a prior (a preliminary guess) that all (unobserved) portfolio optimization inputs are identical across assets. That is, all assets have identical means, variances, covariances and correlations. Because no asset is more attractive than any other asset under this prior, its corresponding optimal portfolio is the "1 over N" portfolio. If the data supports this prior, in the sense that (historical) ex post realizations are near to the grand mean of historical estimates, then the forecast of portfolio optimization inputs is shrunk towards their grand cross-sectional mean with a greater intensity. Accordingly, the portfolio weights are shrunk towards "1 over N". If this "1 over N" portfolio is a good initial guess for the optimal portfolio, then shrunk forecasts are expected to substantially lower Mean Squared Forecast Error (MSFE)(James and Stein 1961). The influential results of DeMiguel et al. (2009b) seem to suggest that this "1 over N" prior is a prudent initial guess. Our method essentially uses (approximate) posterior forecasts of portfolio optimization inputs using a prior distribution that is consistent with the "1 over N" prior.

The main difference of this approach from traditional Bayesian approaches is the way we empirically compute the shrinkage parameters by minimizing the sum of squared

OOS errors, instead of using closed-form expressions for the posteriors and estimating hyper-parameters to evaluate these expressions. A key benefit of our approach is that it assumes very little about the data. If the true data generating process is known, more sophisticated approaches are likely to work better. But given uncertainty and instability in real data, we expect our approach to be more robust to underlying assumptions (see section 4.1 for details).

Different choices for the historical window, the learning period for the correction and other parameters of the analysis do not dismiss the usefulness of the Galton correction. Also, evidence from 1,000 OOS horse races, spanning half a century of data, with stocks randomly drawn from the set of largest caps confirm that the Galton outperforms other methods, both in terms of Sharpe ratios and accuracy of risk estimates, more than expectable by chance alone.

The benefits of the Galton corrected covariance matrix are particularly significant for building portfolios able to deliver expected performance, both in terms of returns and volatility. The implications are especially meaningful for extreme quantiles of the distribution. For instance, in an OOS exercise with the 50 largest stocks by capitalization, the plain Markowitz optimal risky portfolio has losses exceeding the 1% Value at Risk (VaR) almost half of the months. This provides an extreme example of bias in historical risk estimates. 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.

This problem is especially acute for plain Markowitz. But the difficulty in delivering expected performance is not specific to this method. We compare the Galton with a broad set of optimization methods. A ubiquitous problem is that they posit portfolios that fail to deliver expected performance. Their expectations tend to exhibit a pronounced optimistic bias, overestimating returns and underestimating volatility.

In sharp contrast, the Galton-corrected covariance matrix produces risk estimates close to actual OOS risk. For example, for the 50 largest stocks, losses exceeding the 1% VaR only happen in 0.83% of OOS observations. The correction to expected returns also reduces forecast errors in real time. Hence, calibrating from past mistakes helps having expectations closer to ex post performance.

Motivated by the abundant evidence on stock return predictability in the cross section using characteristics, we also examine the OOS persistence of optimization inputs in four sets of 25 double-sorted portfolios on stock characteristics (size, book-to-market, operating profitability, investment, and beta). We find predictability patterns differ across sets. Generally, past returns are more informative about future returns than for individual stocks. As a result, optimized portfolios perform better with these sets of assets, with higher Sharpe ratios OOS. The Galton is no exception and shows consistent above average performance in all sets of assets.

The paper is organized as follows. Section 2 briefly discusses the closely related literature. Section 3 takes a closer look at the evidence on regression to the mean in the inputs of Markowitz optimization. Section 4 provides a theoretical motivation for the Galton correction. Section 5 describes the estimation procedure for Galton coefficients. Section 6 shows the OOS performance of optimized portfolios with individual stocks. Section 7 examines robustness across parameters, variations of the method, and tests with portfolios formed on characteristics. It also assesses statistical significance from 1,000 horse races with Galton and examines GMV portfolios. Section 8 concludes.

2 Related literature

This work is related to a recent literature that proposes robust portfolio 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. A monkey throwing darts may pick a portfolio of stocks with a decent Sharpe ratio. Still, it would be reckless to trust the same monkey with the task of forecasting and managing its risk. Comparatively our work also studies how well various optimization methods are able to forecast the risk of their corresponding optimal portfolios.

Correcting portfolio optimization inputs for past OOS forecast errors can be seen as a form of shrinkage. This is because the most recurrent pattern in our correction is regression to the mean. That is, it adjusts the forecast towards the cross-asset or global average of historical estimates. Hence the cross-sectional differences in the corrected inputs are smaller than the initial estimates. For example, the shrinkage forecast of the one-period ahead expected return of asset i (denoted as $r_{i,t+1|t}$) can be obtained by adjusting its historical average return ($\mu_{i,t}$) towards the cross-sectional grand mean ($\sum_i \mu_{i,t}/N$) across N assets. Therefore, our approach can also be motivated using the influential Stein (1964) result that an estimator obtained by "shrinking" the sample means toward a common value reduces expected mean squared forecast error (MSFE) compared to the sample mean. This approach is related to several shrinkage approaches proposed in the finance literature such as Jorion (1986), Best and Grauer (1992), Ledoit and Wolf (2004a), Ledoit and Wolf (2004b), and Disatnik and Benninga (2007)).

Shrinkage forecasts can also be interpreted as posteriors in an empirical Bayes framework (see e.g. Efron and Morris (1972)). So our approach is also related to the 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), and Frost and Savarino (1986)). Structural models where investors have short sale constraints or multiple priors and aversion to ambiguity also lead to shrinkage of the above form (Jagannathan and Ma (2003) and Garlappi et al. (2006)). Generally, this literature shows that one effective way to reduce the estimation error in expected returns or the covariance matrix is to shrink the estimates to some target.

The idea of using past out-of-sample errors to correct estimates is also akin to cross-validation (e.g., Stone (1974), Shao (1993)) and jackknife estimators (e.g., Efron (1983)). Typically, these approaches examine different methods of splitting the sample, inferring estimates from one part of the sample and testing them on the other. We 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 important ways. Given a specific sample of, say, 5 years of monthly returns for N stocks, the innovation of this method is not to split this sample or throw observations out for out-of-sample testing. Instead, our testing sample is a very specific sample that is of interest of investors and exploits serial correlation in returns and clustering in volatility and correlations. Typical cross-validation approaches generally treat observations in different times as independently distributed and thereby throw away the predictable component of optimization inputs in the subsequent period. Our out-of-sample testing design preserves the order of realized observations. It is also closer to the investors problem of estimating inputs of optimization in real time. These are important distinctions from other possible ways to split data.

In our study the Galton can be interpreted as a correction on top of another forecasting method – the historical estimate plug-in forecasts. However, because our approach is flexible, it can also be used to correct the OOS forecast errors of other forecasting methods such as the single index model. For any such method, our approach suggests using regressions of subsequent realizations of inputs on past forecasts to obtain their corresponding shrinkage parameters. Therefore the approach has wider applications and could, in principle, complement other forecasting methods in the literature.

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 mean-variance (MV) optimal risky portfolio is:

$$\boldsymbol{w}_t = \frac{\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}}{\mathbf{1}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}} \tag{1}$$

where $\boldsymbol{\mu}$ is a N-by-1 vector of mean returns, $\mathbf{1}$ is a N-by-1 vectors of ones, N is the number of assets, and $\boldsymbol{\Sigma}$ is the covariance matrix.[4] 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 at each point in time, this amounts to estimating the inputs $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ 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, we call this the historical (or plug-in or plain Markowitz) method and denote the respective estimates as $\boldsymbol{\mu}_{\mathsf{H}}$ and $\boldsymbol{\Sigma}_{\mathsf{H}}$. In particular, if the return vector of N assets at time t is $\boldsymbol{r}_{\mathsf{H},t}$, then the historical estimates are $\boldsymbol{\mu}_{\mathsf{H},t} = \sum_{s=t-\mathsf{H}+1}^{t} \boldsymbol{r}_{\mathsf{H},s}/\mathsf{H}$ and $\boldsymbol{\Sigma}_{\mathsf{H},t} = (\boldsymbol{r}_{\mathsf{H},t} - \boldsymbol{\mu}_{\mathsf{H},t})(\boldsymbol{r}_{\mathsf{H},t} - \boldsymbol{\mu}_{\mathsf{H},t})'/(\mathsf{H}-1)$.

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

[4]In the results in the empirical sections we divide by $|\mathbf{1}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}|$. This prevents the cases when the negative denominator switches the sign of the relative weights in the complete portfolio.

historical samples with their ex post, out-of-sample, counterparts.

Figure 1 shows the relation between the historical sample and the ex post periods for the entire universe of US stocks. This figure focuses on the general case of estimating risk and returns without any particular risk (or factor) model. Given the large number of asset pricing models available in the literature today, we chose to focus on this agnostic approach where no model is assumed as the truth and the optimization simply relies on past correlations and variances.[5] 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 constant correlation matrix of Elton and Gruber (1973). The slope of the OOS values is clearly not zero either.

Panel C shows variances are relatively well approximated by past values. Panel D shows $\boldsymbol{\mu}_{\mathsf{H}}$ is actually negatively correlated with true expected returns. This result offers a simple explanation for the fact that plain Markowitz performs badly not only in terms of Sharpe ratio but also in expected returns. The method suffers from pervasive estimation error issues but this should not explain its consistently low returns. The plot in panel D shows the method tends to overweight (underweight) stocks with low (high) expected returns, suggesting a plausible explanation of its low returns.

4 Correcting out-of-sample forecast errors

The previous section shows there are large differences between historical estimates of optimization inputs and the values they assume OOS on average. But those differences

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

are also consistent and predictable to some extent. For instance, above average historical correlations tend to become smaller OOS[6]. This leads to the possibility that correcting past OOS forecast errors in inputs can lead to more robust portfolio optimization.

To see this, note that one way to look at the evidence in figure 1 is as a hypothesis test of whether historical estimates of covariances, correlations, variances, and expected returns predict ex-post realizations with a slope of one. Another way to look at the same evidence is to ask: can we improve forecast accuracy by adjusting these historical estimates using a linear correction with a slope coefficient smaller than one? This slope coefficient can be estimated using a regression of ex-post realizations on historical estimates. If this pattern is persistent over time, it is likely that we can reduce MSFE by choosing forecasts on a regression line with a slope smaller than one. This observation motivates our proposed correction for optimization inputs.

Consider the following reversion to the mean (or "Galton") corrected forecast for mean returns: multiply the average historical return ($\mu_{i,\mathsf{H},t}$) by a slope coefficient $g_{1,m,t}$, estimated based on a historical regression of ex post realizations on historical estimates, and add a constant $g_{0,m,t}$ to ensure the mean forecast matches the mean ex post realization in the historical sample:

$$\mu_{i,t+1|t} = g_{0,m,t} + g_{1,m,t}\mu_{i,\mathsf{H},t}. \tag{2}$$

When OLS is used to estimate $g_{.,m,t}$, the mean squared forecast error (MSFE) of ex post returns ($r_{i,t+1}$), $\sum_i^N (r_{i,t+1} - \mu_{i,t+1|t})^2/N$, is expected to be lower than all other arbitrary choices of $g_{.,m,t}$ because OLS minimizes MSFE. Here, we appeal to the result that, even if the true model is not linear, the regression line fit by least squares is an optimal linear predictor for the dependent variable, where optimal is defined based on the quadratic loss function (See Internet Appendix for details). In other words, the forecast $\mu_{i,t+1|t}$ in Eq 2 is expected to be a better predictor of ex post realizations compared to

[6] 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.

simply using the average historical return $\mu_{i,\mathsf{H},t}$ as an input in portfolio optimization. Similarly, using linear "Galton" corrections, $g_{.,v,t}$, $g_{.,c,t}$, and $g_{.,\rho,t}$, for other optimization inputs (variances, covariances, and correlations), analogous to Eq. 2, is likely to reduce MSFE in their forecasts.

The role of the OLS slope in the "Galton" correction is to predict how well each input forecasts its ex post realization. It optimizes on how well the assets that are forecast to be higher (lower) than average in the input vector, actually are higher (lower) than average OOS. If the historical estimate has proven to be a "bad" input in the past, the procedure will give less weight to its predictions (smaller $g_{1,.}$) and more weight to a prediction that says that all elements of the input vector are equal to their global mean. If the historical estimate has proven to be a "good" input in the past, the procedure gives the historical forecast a higher weight (larger g_1). The Galton correction optimally chooses the weight g_1 to make this trade-off with the goal of minimizing forecast risk.

For example, the regression slope in panel D of figure 1 is flat or negative for the average returns of individual securities. Our corrected forecast will account for this and suggest that the expected return input vector for portfolio optimization should have small or no differences across assets. These inputs will be more accurate (lower MSFE) compared to putting more weight on historical top performers and being exposed to mean reversion as well as more concentrated bets on those assets. Similarly, panel A of figure 1 suggests that the regression slope is the largest for variance estimates. So, there is more information in differences in historical variance estimates, but this information is still less than that suggested by historical data. Our corrected forecast will account for these discrepancies and suggest corrections for expected returns, variance, covariance, and correlation inputs to portfolio optimization.

This method, based on minimizing MSFE, should provide better forecasts for portfolio optimization inputs compared to historical estimates. Next, we build on the arguments presented so far and interpret our "Galton" forecasting approach as an empirical Bayes technique.

4.1 An empirical Bayes interpretation

A Bayesian interpretation of our approach is that we assume a prior that all (unobserved) portfolio optimization inputs are identical across assets. In other words, the prior or initial guess of the Galton estimator is that all assets have identical means, variances, covariances and correlations. Because no asset is more attractive than any other asset under this prior, its corresponding optimal portfolio is the "1 over N" portfolio.

If the data supports this prior, in the sense that (historical) ex post realizations are near to the grand mean of historical estimates, then the forecast of portfolio optimization inputs is shrunk towards their grand mean with a greater intensity. That is, in Eq. 2, $g_{1,\cdot}$ is smaller. Accordingly, the portfolio weights are shrunk towards "1 over N". If the guess is contradicted, then not much shrinkage is done.

Stein (1964) shows that, under very general assumptions, such a shrinkage approach can expect to forecast future averages with lower MSFE, no matter what the true expected values of the inputs are. However, this procedure does *substantially* better only if many of the true values lie near each other, so that the initial guess involved is confirmed. So in our context, the Galton forecasts are expected to substantially lower MSFE, if the "1 over N" portfolio is a good initial guess for the optimal portfolio. The influential results of DeMiguel et al. (2009b) seem to suggest that this is the case.

We enhance this intuitive rationale using a Bayesian framework.

4.1.1 Normal approximation of the prior and posterior

Consider the portfolio choice problem where neither the mean nor the covariance matrix is known. In a Bayesian setting, a posterior for these inputs can be used as inputs to the optimization problem. For example, Frost and Savarino (1986) follow a Bayesian approach and assert a Normal-Wishart conjugate prior where the means are normally distributed and the covariance matrix has a Wishart distribution. To obtain closed-form solutions to posteriors, a natural approach is to use such a conjugate prior.[7] However, the structure associated with conjugate priors often leads to non-trivial restrictions in the solution to the problem. In realistic settings, for example when returns are autocorrelated or variances are clustered, conjugate priors are often not available and it is difficult to obtain closed form solutions for posteriors. It is also difficult to obtain closed form solutions when priors are on correlations instead of covariances (see Internet Appendix for a discussion). In our opinion, these difficulties are more of a concern when analytical expressions for posteriors are desired.

Our approach is less restrictive as we do not seek closed form analytical expressions to calculate posteriors. We directly estimate the shrinkage parameters from the data and therefore do not need analytical expressions for these parameters. Our purpose in this subsection is simply to present an empirical Bayes foundation for the Galton method. Hence, we focus on the intuition underlying our approach. We argue that central limit theorems justify using normality as an approximation for priors and posteriors. Assuming the central limit theorem applies, we approximate the prior distribution of means, variances, covariances, and correlations by a normal distribution. The posterior distribution, for large samples, can also be approximated by a normal distribution when the regularity conditions of the *Bayesian* central limit theorem are met. These approximations circumvent several technical difficulties in obtaining closed-form posterior distributions.

Our approximation approach appeals to the robustness and flexibility of normal approximations of priors and posteriors at the cost of the preciseness of fully specified distributions (which may not be consistent with the data). The cost of this more flexible approach, however, is approximation error. Our empirical results suggest, in portfolio optimization contexts using individual security data, the benefits of using normal approximations are greater than associated approximation error costs.

Armed with this normality approximation, the "Galton" estimation approach can be

[7]Choosing a conjugate prior leads to a posterior distribution belonging to the same family as the prior.

interpreted as an empirical Bayesian method. In particular, consider a two-stage Gaussian model for all moments: means, variances, covariances, and correlations. Let the moment under consideration be a vector $\boldsymbol{X}_t$ generated from an (approximate) two-stage Gaussian model:

$$\boldsymbol{X}_t|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x \sim \mathcal{N}_N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x), \tag{3}$$

$$\boldsymbol{\mu}_x \sim \mathcal{N}(\mu_{0,x}\mathbf{1}_N, \tau_x^2\boldsymbol{\Sigma}_x), \tag{4}$$

where $\boldsymbol{X}_t$ is drawn from a multivariate normal distribution with mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Sigma}_x$. Equation 4 is the Bayesian informative prior that the elements of the vector $\boldsymbol{\mu}$, the (unobserved) expected values of the moments, are themselves normally distributed with mean $\mu_{0,x}$ and covariance matrix $\tau_x^2\boldsymbol{\Sigma}_x$. As discussed above, this is a preliminary guess that the expected values of the moments $\mu_{x,i}$ are typically near to their grand mean $\mu_{0,x}$.

We will obtain a posterior for $\boldsymbol{\mu}_x$ that is a linear function of $\boldsymbol{X}_{\mathsf{H},t}$. If $\mu_{0,x}$, and τ_x are known, then we obtain Stein shrinkage as a Bayesian posterior for the expected value vector $\boldsymbol{\mu}_x$, given the historical realization $\boldsymbol{X}_t$:[8]

$$\boldsymbol{\mu}_x|\boldsymbol{X}_t = \mu_{0,x}\mathbf{1} + g_{1,x}(\boldsymbol{X}_t - \mu_{0,x}\mathbf{1}), \tag{5}$$

where $g_{1,x} = \tau_x^2/(1+\tau_x^2)$. The Bayesian posterior, given the data, shrinks the historical estimates in $\boldsymbol{X}_t$ towards their global mean $\mu_{0,x}$, based on τ_x from Eq. 4, which determines the cross-sectional variance of $\boldsymbol{\mu}_x$. The smaller (larger) the variability across the elements of $\boldsymbol{\mu}_x$, the larger (smaller) the shrinkage towards its global mean. This helps align the variance of the posteriors with the a priori variability in $\boldsymbol{\mu}_x$, while reducing the variance due to estimation error.

In purely Bayesian inference, the hyper-parameters $\mu_{0,x}$ and τ_x is assumed known *a*

[8]It is straight forward to derive this using the results in Efron and Morris (1972) (Eq. 4.4).

priori. In contrast, an empirical Bayes approach estimates these parameters, denoted as $\hat{\mu}_{0,x}$, and $\hat{\tau}_x$, from the data. Then parameter estimates of $\hat{g}_{1,m} = \hat{\tau}_m^2/(1+\hat{\tau}_m^2)$ and of the global mean $\hat{\mu}_{0,x} = \mathbf{1}'\boldsymbol{X}_{H,t}/N$ are plugged into the posterior given by Eq. 5 to obtain an estimated posterior mean.

Instead of indirectly inferring shrinkage parameters from hyper-parameters $\mu_{0,x}$ and τ_x, our Galton approach directly estimates the shrinkage parameters using historical data. Despite this difference in estimation, our method can still be interpreted as an empirical Bayes method where the estimated shrinkage parameters $g_{1,x}$ can be used to back-out the hyper-parameters $\mu_{0,x}$ and τ_x. It shares the benefits of an empirical Bayes approach, and when its assumptions hold, it is expected to improve forecasts of portfolio optimization inputs.

4.2 Shrinkage versus Galton forecasts

In various empirical Bayesian approaches, appropriate hyper-parameters determine the shrinkage parameters $g_{1,x}$. Such approaches first estimate hyper-parameters from historical data and then apply analytically derived expressions to infer the shrinkage intensity from these estimated hyper-parameters. For example, in the Frost and Savarino (1986) posterior, the covariance matrix shrinkage parameters are $g_{1,v} = g_{1,c} = \mathsf{H}/(\mathsf{H}+v_0)$, where v_0 is a hyper-parameter that determines the intensity of shrinkage. The shrinkage parameter is calculated based on estimates of v_0 that further depend on the assumed structure of the problem.

When the data is not generated from a two-stage Gaussian model, Eq. 5 is only an approximation to the posterior forecast using a normal approximation for the prior distribution. Still, as an approximation, these Bayesian posterior forecasts are consistent with several prior and posterior distributions. The generality of our normal distribution approximation for priors and posteriors suggests that it is likely to be more robust to model assumptions, compared to competitor methods that derive analytical expressions for posteriors. However, it is likely to under perform in environments where the assumptions of these competing models are consistent with the data.

In a simulation environment where the underlying assumptions of the hyper-parameter approach hold, it is likely to yield superior estimates of $g_{1,m}$, $g_{1,v}$, $g_{1,c}$ and $g_{1,\rho}$ compared to the agnostic Galton approach. The hyper-parameter approach will be more efficient and converge faster. This is because the hyper-parameters incorporate knowledge of the particular structure of the data into the estimation of these parameters. However, in real data, such assumptions are likely violated. We expect the agnostic Galton approach to be more robust in real world environments where assumptions about underlying distribution(s) are likely to be violated. As the Galton approach relies on robust large sample approximations based on the central limit theorem, instead of a fully specified structure, it is likely to be approximately correct under numerous specifications.

4.3 Portfolio choice using Galton corrected inputs

Depending on whether $\boldsymbol{X}_t$ represents means, volatility, covariances, or correlations, we denote the respective Galton forecasts for asset i ($X_{i,t+\mathsf{E}|t}$) as $\mu_{i,\mathsf{G},t}$, $\sigma_{i,\mathsf{G},t}$, $\sigma_{i,j,\mathsf{G},t}$, or $\rho_{i,j,\mathsf{G},t}$, where the G subscript stands for "Galton". The Galton correlation matrix, $\boldsymbol{\rho}_{\mathsf{G},t}$, has in each entry the corrected pairwise correlation $\rho_{i,j,G,t}$ and ones in the diagonal. Similarly, we obtain the N-by-1 vector of the Galton volatility vector, $\boldsymbol{\sigma}_{\mathsf{G},t}$, and mean return vector, $\boldsymbol{\mu}_{G,t}$. The Galton corrected covariance matrix is:

$$\boldsymbol{\Sigma}_{\mathsf{G},t} = diag(\boldsymbol{\sigma}_{\mathsf{G},t})\boldsymbol{\rho}_{\mathsf{G},t}diag(\boldsymbol{\sigma}_{\mathsf{G},t}) \tag{6}$$

where $diag(.)$ stands for a function that maps a N-by-1 vector into a N-by-N matrix whose diagonal elements correspond to those of the argument while the others are zero.

Using these forecasts, we compute corrected inputs for the Markowitz optimization.

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

$$\boldsymbol{w}_{\mathsf{G},t}^{MV} = \frac{\boldsymbol{\Sigma}_{\mathsf{G},t}^{-1}\boldsymbol{\mu}_{\mathsf{G},t}}{\mathbf{1}\boldsymbol{\Sigma}_{\mathsf{G},t}^{-1}\boldsymbol{\mu}_{\mathsf{G},t}} \tag{7}$$

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

$$\boldsymbol{w}_{\mathsf{G},t}^{GMV} = \frac{\boldsymbol{\Sigma}_{\mathsf{G},t}^{-1}\mathbf{1}}{\mathbf{1}\boldsymbol{\Sigma}_{\mathsf{G},t}^{-1}\mathbf{1}} \tag{8}$$

The weights of these "Galton" portfolios are simply the result of plain Markowitz optimization applied to corrected inputs.

4.3.1 Positive semi-definiteness and using correlation versus covariance shrinkage

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 naturally met for the correction we propose in Eq. 6. This happens because our methodology (described in the next section) imposes $g_{0,\rho,t} = \rho_{0,t}(1 - g_{1,\rho,t})$, where $\rho_{0,t}$ is the grand mean of pairwise correlations observed up to time t. As a result, the Galton correlation matrix $\boldsymbol{\rho}_{\mathsf{G},t} = \rho_{0,t}(1 - g_{1,\rho,t})\boldsymbol{\iota}\boldsymbol{\iota}' + g_{1,\rho,t} * \boldsymbol{\rho}_{\mathsf{H},t}$. As $g_{1,\rho,t}$ is restricted to be a number between 0 and 1 (see section 3), this is the convex combination of two positive semi-definite matrices. As a result $\boldsymbol{\rho}_{\mathsf{G},t}$ is a positive semi-definite matrix itself. Furthermore, we restrict $\boldsymbol{\Sigma}_{\mathsf{G},t}$ to be strictly positive to machine precision.

Notice that, instead of using Eq. 6, we could consider using a covariance matrix ($\boldsymbol{\Sigma}_{\mathsf{G},t}$) with the Galton variances $\sigma_{i,\mathsf{G},t}$ in the diagonals, and the Galton covariances $\sigma_{i,j,\mathsf{G},t}$ in the off-diagonals. However, to the best of our knowledge, there is no guarantee that such a matrix will be positive semi-definite. Therefore, we use the covariance matrix from Eq. 6 in our main tests. In robustness tests, we also consider the covariance matrix forecast constructed from shrunk covariances and find similar results.

4.4 The Galton forecast optimizes across a platter of portfolios

It is well known that if a mean-variance optimizing investor knows the true parameters, she should invest only in the riskless asset and the tangency portfolio. However, these parameters are unknown in practice and this generates a loss in expected out-of-sample performance. Kan and Zhou (2007) show that choosing a combination of the riskless asset, the sample tangency portfolio, and the sample minimum variance portfolio is likely to diversify estimation risk and improve out-of-sample performance. They refer to this as three-fund separation, where an investor's complete portfolio can be chosen as a combination of these three portfolios. Tu and Zhou (2011) optimally combine the "1 over N" rule with one of four other strategies. Such approaches typically derive optimal weight combinations analytically and then implement them using estimated hyper-parameters.

Our Galton procedure is related to these approaches, though it does not estimate hyper-parameters to combine a set of exogenously specified portfolio rules. Instead, the Galton procedure combines a set of portfolio rules by finding a combination that minimizes forecast error for means, variances, and correlations. It allows the data to decide how to combine one of several candidate portfolio rules. Three familiar rules that the Galton procedure optimizes over are the (i) equally-weighted ("1 over N") Talmud portfolio, (ii) the Markowitz ex-post tangency portfolio, and (iii) the sample global minimum variance portfolio. In addition, it also optimizes over a portfolio based on past return realizations, which can include either return momentum or return reversal based on the empirical relationship between past returns and future returns in the time horizon being considered.

To see this, let $g_{.,m}$, $g_{.,v}$, and $g_{.,\rho}$ represent the Galton coefficients for the means, variances, and correlations respectively. Different extreme values of $g_{1,m}$, $g_{1,v}$, and $g_{1,\rho}$ will produce recognizable portfolio rules. The Galton procedure optimizes over these to minimize forecast risk for means, variances, and correlations.

We obtain the equally-weighted ("1 over N") Talmud portfolio when $g_{1,m} = g_{1,v} = g_{1,\rho} = 0$. Here, the Galton procedure basically says that it can't learn anything from the historical data, so it chooses a portfolio assuming all assets are identical.

We obtain the Markowitz mean-variance portfolio when $g_{1,m} = g_{1,v} = g_{1,\rho} = 1$. Here, the Galton procedure says that the historical data is the best forecast of the future. So Galton picks a portfolio assuming that the historical estimates are the same as the population estimates.

When $g_{1,m} = 0, g_{1,v} = g_{1,\rho} = 1$, we obtain the sample global minimum variance portfolio. In this case, the Galton procedure says that the historical data for the means is very noisy, but the historical data for the sample variances and covariances are very reliable. Therefore, lets choose a portfolio whose weights do not depend on mean returns.

An interesting momentum portfolio rule arises when $g_{1,m} > 0, g_{1,v} = g_{1,\rho} = 0$. This portfolio says that the weights should depend completely on past returns over H periods. This will be the case when the estimates of past variances and covariances do not predict their future realizations, but past returns are very good predictors of future returns. The Galton procedure evaluates whether this is the optimal portfolio rule based on the ability of past returns to predict future returns. A reversal portfolio is also considered by Galton, when $g_{1,m} < 0, g_{1,v} = g_{1,\rho} = 0$. Here, the Galton procedure evaluates whether high past returns over H predict low future returns.

A key advantage of the Galton procedure is that choosing only 3 shrinkage parameters allows for a rich range of possible portfolio rules. Further, these parameters can be readily estimated from the data without the need for additional assumptions regarding hyper-parameters that determine the amount of shrinkage. This balance between flexibility and structure is a key differentiating feature of the Galton procedure.

5 Estimating the Galton shrinkage parameters

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 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_i (variance, pairwise correlation, or mean return) for asset i, let $X_{i,\mathsf{H},t}$ 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 $X_{i,\mathsf{E},t+\mathsf{E}}$. The baseline values for these periods in this paper are $\mathsf{H} = 60$ and $\mathsf{E} = 12$.

We are interested in obtaining a forecast of $X_{i,\mathsf{E},t+\mathsf{E}}$ that is linear in the historical estimate $X_{i,\mathsf{H},t}$:

$$X_{i,t+\mathsf{E}|t} = \hat{g}_0 + \hat{g}_1 X_{i,\mathsf{H},t}. \tag{9}$$

In particular, we are after a forecast that is the historical estimate, $X_{i,\mathsf{H},t}$, corrected by how close (or how far) all known past historical estimates were of their subsequent OOS realizations. Other more sophisticated methods would likely produce better corrections, but we refrain from that pursuit and focus instead on a straightforward linear function for its simplicity. Further, we are agnostic about the data generating process. As this forecast only uses information available until time t, it can be used for OOS tests.

We consider the following population regression model:

$$X_{i,\mathsf{E},t+\mathsf{E}} = g_0 + g_1 X_{i,\mathsf{H},t} + \epsilon_{i,t} \tag{10}$$

for $i = 1, ..., \mathsf{N}_t$, 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 exactly correct, the best estimate of $X_{i,\mathsf{E},t+\mathsf{E}}$ is $X_{i,\mathsf{H},t}$ and so $g_0 = 0$ and $g_1 = 1$. If regression to the mean is total, the best estimate of $X_{i,\mathsf{E},t+\mathsf{E}}$ is the cross section average, so $E(X_{i,\mathsf{E},t+\mathsf{E}}) = g_0$ and $g_1 = 0$.

5.1 Fama-MacBeth estimation of the Galton coefficients

We estimate the Galton coefficients from a historical panel of the $X_{i,t}$ vector realizations over T periods. This panel data can be unbalanced as we do not require that the length of the vector $\boldsymbol{X}_t$ to be the same in each period. That is, individual stocks can enter and exit our sample without impacting the procedure. We use a "pseudo out-of-sample" method in all empirical tests. That is, we obtain $\hat{g}_{0,t}$ and $\hat{g}_{1,t}$, our estimate of g_0 and g_1 using data only till time t.

For each forecasting time period t, we use t observations of $\boldsymbol{X}_t$ to estimate the coefficients. The panel procedure can be considered a series of cross-sectional procedures when the forecast target is the realization $\boldsymbol{X}_{\mathsf{E},t}$ and the input vector $\boldsymbol{X}_{\mathsf{H},t-\mathsf{E}}$ uses data H periods prior to $t - \mathsf{E}$ to construct forecasts of $\boldsymbol{X}_{\mathsf{E},t}$.[9]

Of all linear predictions that use $\boldsymbol{X}_{\mathsf{H},t-\mathsf{E}}$ to predict $\boldsymbol{X}_{\mathsf{E},t}$, we are interested in the linear forecast that minimizes expected mean squared forecast error (MSFE):

$$\text{MSFE} = \text{E}[(\boldsymbol{X}_{\mathsf{E},t+\mathsf{E}} - \boldsymbol{X}_{t+\mathsf{E}|t})^2]. \tag{11}$$

This is estimated by the least square regression line of $\boldsymbol{X}_{\mathsf{E},t}$ on $\boldsymbol{X}_{H,t-\mathsf{E}}$.[10] In principle, if the entire historical panel is used, then estimates of these coefficients can be obtained by stacking up t sets of historical vectors $\boldsymbol{X}_{\mathsf{E},t}$ and $\boldsymbol{X}_{\mathsf{H},t-\mathsf{E}}$ to obtain a dataset with more

[9]Note that we use a rolling window approach, keeping the horizons H and E fixed. Using expanding windows might imply different values for the Galton coefficients, as shrinkage parameters are typically sensitive to sample length.

[10]The main condition for this result is that the data generating process is such that the law of large numbers applies to the least squares estimator (see Greene (2003), section 4.2 and theorem 4.1 for details).

data points. That is, create two stacked vectors $\boldsymbol{X}_{\mathsf{E},.}$ and $\boldsymbol{X}_{\mathsf{H},.}$, and evaluate one pooled regression line to estimate $\hat{g}_{0,t}$ and $\hat{g}_{1,t}$.[11] This approach is likely to provide lower risk estimators compared to the single cross-section approach as it exploits the observations in the time-series dimension to obtain better estimates of g_0 and g_1. Yet, this relies on the assumption of uncorrelated errors across the stacked historical vectors.

When errors are correlated, OLS is still consistent, but its efficiency is reduced. This is a concern since return data typically have high cross-sectional correlations. If one asset's return is unusually high this period, another asset's return is also likely to be high; if one asset's variance is unusually high this period, another asset's variance is also likely to be high. Since this is an important feature of the data, accounting for these variations improves estimates of g_0 and g_1. This can be achieved in a parsimonious manner using the Fama-MacBeth approach (see e.g. Cochrane (2001), Chapter 12). This means that the forecast risk given by Eq. 11 can be further reduced by accounting for cross-sectional correlation in errors using the Fama-MacBeth approach, thereby improving estimates of g_0 and g_1 and reducing estimation variability.

For each period $s \leq t$ in the sample, we run the cross-section regression:

$$X_{i,\mathsf{E},s} = g_{0,s}^{fm} + g_{1,s}^{fm} X_{\mathsf{H},s-\mathsf{E}} + e_{s,i} \tag{12}$$

for the superscript fm stands for Fama-MacBeth and $i = 1, ..., N_s$, where N_s is the number of stocks (for variances or mean returns) or pairs of stocks (for correlations) available in the sample at time s. The Fama-MacBeth estimates of g_0 and g_1 are $\hat{g}_{0,t} = \sum_{s=1}^{t} \hat{g}_{0,s}^{fm}/t$ and $\hat{g}_{1,t} = \sum_{s=1}^{t} \hat{g}_{1,s}^{fm}/t$.

This procedure improves estimates by allowing time-fixed effects to be absorbed by $\hat{g}_{0,s}^{fm}$ and $\hat{g}_{1,s}^{fm}$. The time-series average of $\hat{g}_{0,s}^{fm}$ and $\hat{g}_{1,s}^{fm}$ have lower (in-sample) estimation error compared to coefficients from any single cross-section or a pooled regression with all

[11]Note that, when the historical vectors estimated over H periods are stacked on one-another, there will be considerable overlap in the data across periods. However, the consistency of OLS estimates will not be impacted by this autocorrelation due to overlapping data.

time periods before t. To see this, note that the mean squared error (MSE, in-sample) of Fama-MacBeth is the same as that from a "time-fixed effects" model:

$$\mathrm{MSE}(g_0, g_1) = \sum_{s=1}^{t} \frac{\mathrm{E}[\boldsymbol{X}_{\mathsf{E},s} - (g_0 + \eta_{0,s})\iota - (g_1 + \eta_{1,s})\boldsymbol{X}_{\mathsf{H},s-\mathsf{E}})]^2}{t}, \tag{13}$$

where $\eta_{0,s}$ and $\eta_{1,s}$ are time fixed effects for the intercept and slope; $\sum_{s=1}^{t} \eta_{0,s} = \sum_{s=1}^{t} \eta_{1,s} = 0$; $\hat{g}_{0,s} = \hat{g}_0 + \eta_{0,s}$ and $\hat{g}_{1,s} = \hat{g}_1 + \eta_{1,s}$.[12]. The main advantage of including these fixed effects is to get better (lower error variance) estimates of g_0 and g_1 by controlling for the impact of time-period specific estimation error captured by $\eta_{0,s}$ and $\eta_{1,s}$. Note that, this equation should not be interpreted as reducing forecast risk by including future fixed effect coefficients $\eta_{0,t+1}$ and $\eta_{1,t+1}$, which are not observed at t.

The Galton coefficients are updated each period based on the cross-sectional regression in that period. That is, $\hat{g}_{0,t} = [(t-1)\hat{g}_{0,t-1} + \hat{g}_{0,t}^{fm}]/t$ and $\hat{g}_{1,t} = [(t-1)\hat{g}_{1,t-1} + \hat{g}_{1,t}^{fm}]/t$. These Galton coefficients constantly learn from data, even though the set of stocks in the current period might be totally different from the set of stocks a few decades ago.

The history of cross sectional estimates $\hat{g}_{0,t}$ and $\hat{g}_{1,t}$ can be used to form the Galton forecast of the portfolio optimization input:

$$X_{i,t+\mathsf{E}|t} = \hat{g}_{0,t} + \hat{g}_{1,t} X_{i,\mathsf{H},t}. \tag{14}$$

Assuming stationarity (unconditional moments exist), the Galton forecast should become more accurate (closer to unconditional values) 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 $X_{i,t+\mathsf{E}|t}$ will occur at time $\mathsf{H}+\mathsf{L}+\mathsf{E}+1$. We pick an arbitrary initial period (L) of 120 months for the correction. For the chosen

[12]In practical implementations of the time fixed effect regression, one of the time-fixed effects must be dropped to avoid perfect multicollinearity amongst the regressions. This however will not affect the MSE as there is no loss in predictive power when a perfectly spanned variable is omitted. To make the relation between $g_{.,t}$ and $g_{.}$ transparent, Eq. 13 includes all fixed effects.

values of $\mathsf{H} = 60$, $\mathsf{L} = 120$, and $\mathsf{E} = 12$, the starting period amounts to 193 months.[13]

5.2 Estimated Galton coefficients

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, we 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 ten 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 10.30, 17.48, and 11.43 respectively. The hypothesis of a slope of zero is rejected at the 5% level in at least 98% of the cross section regressions for covariance, correlation, and variance. This strongly suggests optimization could do better than just using the constant correlation matrix of Elton and Gruber (1973).[14] Effectively, only a few monthly regressions should be enough to suspect of the existence of some predictability in optimization inputs.

[13]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.

[14]For correlations, the R-square of the regression is quite low (only about 3%). This shows that making inference with a relatively small number of assets, as in Elton and Gruber (1973), the constant correlation matrix provides a good approximation. By contrast, the high statistical significance of the slope coefficient in our results benefits from a very large cross section of pair-wise correlations.

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, -17.94, -44.69, and -9.46. 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 -4.88.

For covariances, correlations, variances, and mean returns, at least 88% of 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 the existence of regression to the mean.

The coefficients in table 1 are estimated at the end of the sample. The Internet Appendix plots the expanding window Galton coefficients. These show modest variation after the initial learning period and then become increasingly stable.

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 for all pairs of stocks. For mean returns there is very little forecasting ability using past returns.

6 The OOS performance

Our Galton approach is not meant to span existing estimators as its unique, flexible, characteristics offer a potential complement to existing methods. But it is pertinent to compare its performance with important optimization methods studied in the literature. We implement these optimized methods both for the mean-variance portfolio (MV) and, where appropriate, the global minimum variance portfolio (GMV).

For the MV portfolios we include standard implementations of the methods, estimating expected returns with samples means, as well as our implementations, where we estimate expected returns using the James and Stein (1961) positive part shrinkage estimator. We examine MV portfolios of only risky assets (the tangency portfolio) as well as MV portfolios that invest in the riskless asset in addition to risky assets (the complete portfolio). We refer to the portfolio with only risky assets as scaled because its weights in risky assets are adjusted so that they add up to one. The weights in risky assets of the complete portfolio are not scaled and do not add up to one. We refer to this portfolio as unscaled.

In total, the comparison universe for our two "Galton" MV methods (scaled and unscaled) comprises fifteen optimized and two naïve portfolios. Beyond this, we examine GMV portfolios and hybrid strategies in section 7.

6.1 Description of the methods compared

Given the importance in the literature of naïve portfolios as simple methods to avoid estimation risk altogether (e.g. DeMiguel et al. (2009b)), we include the value-weighted (VW) and equal-weighted (Talmud or EW for short) portfolios in our comparison universe.

We also include the Jorion (1986) method, which assumes that the mean vector is unknown but the covariance matrix is known. It provides a Bayesian posterior which effectively shrinks the expected returns towards the expected return on the sample GMV.

We also include portfolio mixing strategies suggested by Kan and Zhou (2007) (KZ) and Tu and Zhou (2011) (TZ). KZ show that in the presence of estimation error the well-known result of the two-fund separation theorem is no longer optimal. They propose instead an optimized combination of three "funds": they add the sample GMV to the sample MV and the risk-less rate. TZ propose optimal combinations of naïve portfolios with optimized strategies. These strategies trade-off bias and variance. Naïve rules (such as the $1/N$) have zero variance but non-zero bias. On the other end, optimal rules, such as Markowitz, have zero bias but considerable small-sample variance. They argue that combinations of both extremes should dominate either extreme. We include their combination of the $1/N$ with KZ as that shows particularly good results in their study.

We include two MV strategies using the Elton and Gruber (1973) (EG) covariance matrix the and the Ledoit and Wolf (2004a) (LW) shrinkage approach. The EG covariance matrix is:

$$\mathbf{\Sigma}_{\mathsf{EG},t} = diag(\boldsymbol{\sigma}_{\mathsf{H},t})\boldsymbol{\rho}_{\mathsf{EG},t}diag(\boldsymbol{\sigma}_{\mathsf{H},t}) \tag{15}$$

where $\boldsymbol{\sigma}_{\mathsf{H},t}$ is the N-by-1 vector of estimated volatilities from the historical sample, and $\boldsymbol{\rho}_{\mathsf{EG},t}$ consists of a matrix where the diagonal elements are all ones and the non-diagonal elements are a constant, equal to the average of pairwise correlations in the rolling historical sample. LW provide a covariance matrix that shrinks the sample covariance matrix towards the constant-correlation model. They obtain their shrinkage intensity by minimizing a loss function based on the Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix (under their assumed structure).

The EG and LW shrinkage methods do not specify estimators for the expected return vector. The natural approach is to simply use historical averages. We report results using this approach. However, we recognize that historical averages are very noisy estimates of true expected returns (e.g. Jobson and Korkie (1980)). As such it can be seen as a fool's errand to only use these. To compare the performance of "Galton" with better performing MV portfolios, we consider portfolios combining EG, LW, and the sample covariance matrix with the positive part James-Stein shrinkage estimator for mean returns ($\boldsymbol{\mu}_{\mathsf{JS}}$)[15]:

$$\boldsymbol{\mu}_{\mathsf{JS}} = \mu_0\mathbf{1} + g_{1,\mathsf{JS}}(\boldsymbol{\mu}_{\mathsf{H}} - \mu_0\mathbf{1}), \tag{16}$$

$$g_{1,\mathsf{JS}} = \max\left(1 - \frac{(N-3)\sigma_0^2}{\|\boldsymbol{\mu}_{\mathsf{H}} - \mu_0\mathbf{1}\|^2}, 0\right), N \geq 4. \tag{17}$$

This is a shrinkage estimator that consists of a weighted average of the $N \times 1$ vector of

[15]We used this shrinkage estimator instead of others as it is a standard in the statistics literature and we did not want to appeal to the underlying Bayesian priors of such mean shrinkage estimators proposed in the finance literature such as Jorion (1986). Such Bayesian priors on the mean are likely to be inconsistent with the priors of various covariance shrinkage methods used in our analysis.

sample average returns ($\boldsymbol{\mu}_{\mathsf{H}}$), taken over the historical sample, with its own cross-sectional average(μ_0). σ_0^2 is the grand variance of returns. To the best of our knowledge, portfolios constructed using combinations of EG, LW, or historical covariance matrices with $\boldsymbol{\mu}_{\mathsf{JS}}$ have not been studied in the literature. They may not be fully internally consistent as they have been derived under assumptions that are not necessarily compatible with each other. They are, in a sense, artificial additions to the comparison set with the aim of populating it with more challenging methods. Note that, both KZ and TZ are specifically MV methods that already implement shrinkage. So, we do not combine it with $\boldsymbol{\mu}_{\mathsf{JS}}$.

We also examine a set of optimized portfolios where the riskless rate is included in the investment universe. These address the important economic problem of an investor not only seeking the optimal risky portfolio – the one spanning the steepest capital allocation line – but also the optimal combination of this risky portfolio with the riskless asset by choosing the best point on that line. Note that, changing the scale of weights would make no difference to the Sharpe ratio if the riskless rate is constant. However, when the riskless rate is dynamic, the (unconditional) Sharpe ratio of the scaled and unscaled strategies will differ. We calculate un-scaled weights for the risky assets as $w^* = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma$ where γ is the coefficient of relative risk aversion for a mean variance investor. As a result, for this set of methods, weights in risky assets do not add up to one. The difference of this sum from one, $1 - w^*\mathbf{1}'$, is invested in the riskless rate. We adopt a γ of 3 as in Kan and Zhou (2007) and Tu and Zhou (2011).[16]

6.2 The data

Our main dataset consists of monthly returns of the entire universe of US listed stocks on the Center for Research in Security Prices (CRSP). We adopt as the evaluation period the standard asset pricing sample of 1962:01 to 2016:12. In our base line optimization, we use a 60 month window for H. As a result, the first OOS return is in 1967:01 – the

[16]These two methods in particular are originally proposed with un-scaled weights. So, this optimization with a riskless rate is a better reflection of their intended purpose.

start of the horse race period. We also use the 10 years of data from 1952:01 to 1961:12 to inform the burn-in estimate for "Galton". Specifically, we use a learning period (L) of 108 months and a subsequent window (E) of 12 months.[17]

We pick the 50 stocks with the largest market capitalization in December, 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 at the end of next December, until the end of the time series. The resulting OOS horse race spans 50 years from 1967:01 to 2016:12.

6.3 Horse race

Panel A of table 2 shows the OOS performance of MV portfolios of the 50 largest stocks in the US market. The optimization method achieving the highest Sharpe ratio OOS over this extensive 50-year OOS period is the Galton MV with a ratio of 0.47. It outperforms naïve portfolios as well as optimized portfolios. Of the first ten optimized portfolios, without investments in the riskfree asset, six have negative Sharpe ratios. Besides "Galton", the best performing method is TZ with a Sharpe ratio of 0.18. LW with (without) shrinkage in expected returns achieves a positive Sharpe ratio of 0.11 (0.12). Still, this performance is far below that of naïve portfolios with Sharpe ratios of 0.37 (VW) and 0.38 (EW). The overall performance of these optimized portfolios confirms with individual stocks the result DeMiguel et al. (2009b) obtain with industry and size / book-to-market sorted portfolios. [18]

[Insert table 2 about here]

Limitations of portfolio management go beyond low Sharpe ratios alone. To show this, we report measures of portfolio concentration and stability for each method in panel

[17] We show in table 3 that our results are robust to other choices of number of stocks and historical, subsequent, and learning windows.

[18] Plyakha et al. (2012) provides an alternative explanation for this. They argue that this rule, coupled with monthly rebalancing, implies a dynamic trading strategy that buys (sells) short-term losers (winners). This exploits the well-known short term reversal effect of Jegadeesh (1990).

A. These are the active share, the turnover of the portfolio, the average minimum and maximum weight, the portfolio concentration (the mean, over the time series, of the standard deviation of weights in the cross section), the time series standard deviation of this standard deviation (a measure capturing instability of that concentration), and the sum of negative weights.

It is apparent that optimized MV portfolios imply very high levels of turnover contrasting those of EW and VW. They are as high as 21427.91% a month for the Markowitz portfolio. They also imply extreme weights, both positive and negative, weights that deviate substantially from the market, and concentrated portfolios with considerable time-variation in that concentration.

Using James-Stein shrinkage alleviates these concerns. LW with $\boldsymbol{\mu}_{\mathsf{JS}}$ has a turnover (116.77%) and portfolio concentration (7.28) similar to the Galton (78.35% and 6.14 respectively). Nevertheless, the Galton achieves particularly sensible weight metrics. It is often closer to those of VW and EW than to optimized portfolios.

The last seven rows show the performance of optimized portfolios investing also in the riskless rate. Not scaling weights considerably improves performance for the KZ portfolio (Sharpe ratio increases from -0.27 to 0.05). Still, the Galton has the highest Sharpe ratio (0.46). The second best method is TZ with a Sharpe ratio of 0.18.

An often neglected shortcoming of portfolio management is the difficulty in designing portfolios that meet expectations. The literature on portfolio optimization rarely reports expected out-of-sample performance, and research in this field is typically satisfied with achieving satisfactory out-of-sample performance in terms of Sharpe ratios. But the gap between expectations and plausibility is too extreme to ignore. For instance, a Markowitz investor expects on average a return of 398.28 and a standard deviation of 54.08. This implies a Sharpe ratio of 7.36. Yet, the historical long run Sharpe ratio of investing in the stock market has been close to 0.4 and this is already a famous puzzle in asset pricing (Mehra and Prescott (1985)). This alone suggests the method implies unrealistic expectations that need correction.

Accurate risk forecasts are important for themselves. Institutional investors often have concentrated portfolios of effectively 50 stocks or less.[19] These concentrated holdings are positively associated with risk-adjusted performance (Kacperczyk et al. (2005)). Whatever the information set used to estimate returns – factors, characteristics, or privately produced fundamental analysis – institutional investors need reasonable covariance matrices on its own merits, not just to pick weights. This is needed to estimate value-at-risk (VaR) for extreme quantiles of the distribution, particularly for hedge funds pursuing long-short strategies or central clearing counterparties (CCPs) estimating margin requirements. For instance, article 41 of the 648/2012 regulation known by EMIR of the Council of European Union (2012) specifies that CCPs must estimate margins that are "sufficient to cover losses that result from at least 99% of the exposures movements over an appropriate time horizon." So, estimating the 1% VaR of portfolios is important for fixing the margin requirements implicit on day-to-day operations in financial markets.

Panel B of table 2 shows the average expected returns, average realized return ex post, the root mean squared forecast error (RMSFE) of the return forecast, the expected standard deviations and its respective ex post realizations. It also shows, in the last four columns, the hit rates for extreme quantiles. This allows a thorough comparison of ex post performance with ex ante expectations derived from each method (naïve portfolios are devoid of methods to form expectations so none are reported for those).

Optimized portfolios tend to systematically over-estimate expected returns. The investor in a plug-in Markowitz portfolio on average expects an annualized return of 398.28% but is surprised with ex post returns of -253.57%. The RMSFE shows forecasts are on average 431.90% away from ex post values. These are meaningful deviations. The exceptions with positive surprises are the LW with $\boldsymbol{\mu}_{\mathsf{H}}$ and the scaled TZ. Still, their respective RMSFE are quite high (1612.96 and 52.81, respectively). The RMSFE of the ten methods with scaled weights – hence more easily comparable between them – ranges

[19]By effective number of holdings we mean the inverse of the Herfindhal index, the equivalent number of holdings. This is about 20 for hedge funds and 50 for mutual funds (Agarwal et al. (2013)).

between 5.29 (Galton) to 1612.96 (LW with $\boldsymbol{\mu}_{\mathsf{H}}$). The only method with an accuracy close to Galton in this set is the LW with $\boldsymbol{\mu}_{\mathsf{JS}}$ (6.60).

Expected returns are notoriously hard to estimate. But the risk of optimized portfolios is not accurately estimated either. It is substantially under-estimated. The investor in a plain Markowitz, for example, gets on average 17 times the anticipated standard deviation (933.51% versus 54.08%). Among the other nine scaled methods, the best performer is the LW with $\boldsymbol{\mu}_{\mathsf{JS}}$. Its ex post volatility is about twice the ex ante estimate using the same covariance matrix (21.05 versus 10.17). Even the TZ method, among the best performers in terms of OOS Sharpe ratio, has ex post risk more than 10 times its ex ante estimate (219.52 versus 20.08). The best performing method predicting risk is the Galton MV with an expected standard deviation of 19.03% versus an average realization of 18.27%. This means it is, on average, moderately conservative.

For the strategies investing in the risk-free rate, the unscaled TZ method has the smallest RMSFE (4.90) and the second smallest is for the Galton (6.74). Again, the Galton has the closest estimate of ex post volatility (23.89 expected versus 23.17 ex post). The unscaled LW comes closer to this performance with ex post risk 64% higher than estimated (165.00 versus 100.49).

Related to these results, Basak et al. (2005) find the historical method, the single factor model, the Fama and French (1993) 3-factor model and Ledoit and Wolf (2004a) all systematically underestimate the OOS risk of the GMV. For the case of the historical method, Kan and Smith (2008) further deepen the result and show, theoretically and empirically, the same result holds for the entire minimum variance frontier – as opposed to the GMV alone. The problem of excessive optimism is pervasive in our results. It shows up consistently using a broad set of methods besides the historical one and it features in expected returns and volatilities. So, it seems a general feature of MV portfolios, even when using shrinkage in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This result also implies a milder statement of concerning implications for risk management: empirically, for each available covariance estimation method, there seems to be a non-empty set of portfolios (namely the estimated optimal

MV, but most likely other portfolios too) for which risk is greatly under-estimated.

The importance of estimating risk accurately is particularly relevant for left-tail risk measures such as value-at-risk (VaR). The last 4 columns of panel B report the hit rates below low (above high) quantiles of the distribution according to expected returns and standard deviation assuming a normal distribution. The results show that for some portfolios, designed to obtain an optimal risk-return trade-off in a historical sample, past hit rates are quite misleading of true OOS risk.

The historical method achieves a dismal performance where 49.33% of OOS returns are below the estimate of 1% level VaR. Other methods achieve much better results. Namely, the LW and the EG-MV with $\boldsymbol{\mu}_{\mathsf{JS}}$ have losses exceeding the 1% VaR only 8.00% and 10.83% of the time, respectivelly. Yet, these are still very uncomfortable results for any conscientious risk manager.

The Galton, with a hit rate of 0.83% for the 1% quantile achieves the most accurate performance. In fact, for this method and for all shown quantiles, the p-values fail to reject the null that ex post hit rates are the same as expected ex ante. Among the compared methods, the Galton is the only one achieving this adequate level of accuracy. This high precision while assuming a normal distribution is perhaps surprising given the well-known fact that equity returns are fat-tailed (Mandelbrot (1963)) and negatively skewed.

7 Robustness and other tests

7.1 Robustness

We choose as parameters 60-months historical window (H), 12-months of subsequent window (E), a learning period of 108 months (L), and 50 stocks (N). This base setting intends to be plausible. For instance, betas of individual stocks are usually not estimated with more than 60 months as longer historical windows may fail to capture time-varying leverage, profitability, operational risk, and growth rates of firms in the sample. Also, 50 stocks matches the equivalent number of holdings for typical mutual funds (Agarwal et al. (2013)). Yet, these are still arbitrary choices. For robustness, we tweak each of these parameters and assess the performance of the Galton in different settings.

[Insert table 3 about here]

Table 3 shows the results. We find the Sharpe ratio varies between 0.38 and 0.67 depending on the setting. As a method that exploits errors in forecasts, it naturally benefits from having a long enough sample to learn. The best results in this exercise are with an L of 180 and the worst with an L of 60. We note that, given the extensive history available for individual stocks now, there is currently not an active data constraint for using an appropriately high level of L.

The method systematically over-estimates the returns of the portfolio by a range between 0.86% (10.80-9.94) and 6.61% (14.81-8.20).[20] On the other hand, the expected standard deviations are close to ex post realizations. So, accuracy of volatility forecasts seems a robust feature of the method. This translates into reasonable hit rates. For the 1% VaR these range between 0.95% and 1.96%.

For the sake of space, we omit here the comparison with all other methods. The full set of results is available in the Internet Appendix. In these results, we notice that compared methods tend to exhibit greater variation in performance across settings. The consistent relative performance of Galton, as either one of the best methods (or the best) in each of the settings, further reinforces the case for its robustness.

7.2 Variations and hybrid strategies

We analyze some variations of Galton and hybrid strategies (e.g. $\boldsymbol{\mu}_{\mathsf{G}}$ with $\boldsymbol{\Sigma}_{\mathsf{LW}}$) to better isolate the role of shrinking $\boldsymbol{\mu}$ versus $\boldsymbol{\Sigma}$ using the Galton procedure as well as to establish closer comparisons with other methods. Table 4 shows the results for this set of exercises.

[20]Part of this effect may be caused by the size premium. Galton expected returns ignore stock characteristics and the samples in this exercise consists of the largest 50 to 100 stocks.

[Insert table 4 about here]

Our method– labeled in this table as "Galton (Correlation)"– implements a different shrinkage to variances and correlations. This is justified by the different slopes for these two inputs shown in table 1 (0.55 for variances versus 0.28 for correlations). This table shows the performance of an alternative version of Galton applying shrinkage to covariances instead of correlations – "Galton (Covariance)". As noted in section 4.3.1, applying different shrinkages to variances and covariances to estimate $\mathbf{\Sigma}_{\mathsf{G}}$ does not necessarily result in a positive definite covariance matrix. However, for the sake of comparison, we estimate the covariance matrix in this way. (We did not encounter any invertability issues in our empirical analysis.) The results show this MV portfolio has a reasonable performance in terms of Sharpe ratio (0.43), although not as good as the 0.47 of plain Galton – "Galton (Correlation)". Its ex ante estimated volatility, 19.71, is also very close to ex post volatility of 19.13. This suggests the main concept of the Galton – using past OOS forecast errors to improve estimation – can be implemented without much of a loss by applying Galton shrinkage to covariances instead of correlations. However, we prefer "Galton (Correlation)" as it guarantees that the covariance matrix is positive definite.

To better compare the relative importance of shrinking means and the covariance matrix, we examine two MV strategies: a) "Galton Mean Only" combining $\boldsymbol{\mu}_{\mathsf{G}}$ with $\mathbf{\Sigma}_{\mathsf{H}}$; and b) "Galton Cov Only" combining $\boldsymbol{\mu}_{\mathsf{H}}$ with $\mathbf{\Sigma}_{\mathsf{G}}$ (the plain Galton covariance matrix). Compared to the Sharpe ratio of Markowitz in table 2 (-0.27), applying Galton to correct only μ increases the Sharpe ratio to 0.24 while correcting only Σ achieves a Sharpe ratio of just -0.21. This shows that correcting means alone is preferable to shrinking the covariance matrix alone. This is likely due to the large errors associated with estimating means. Notwithstanding, none of these two variants of Galton can be credited with the full gains of the approach. Correcting both inputs adds to performance relative to correcting just one (no matter which one). So, there is not a unique channel driving the results.

Our method shares similarities with the shrinkage approach of Ledoit and Wolf

(2004a), a method performing well in some of our comparisons. To establish a more direct comparison with this method, we form a MV portfolio using $\boldsymbol{\mu}_{\mathsf{G}}$ and $\boldsymbol{\Sigma}_{\mathsf{LW}}$ as inputs. This test equips both Galton and the LW with the same estimate for expected returns, hence isolating a comparison of covariance matrices. We find the Sharpe ratio of this combination is 0.53, above the 0.47 of plain Galton. So LW covariance matrix performs well when combined with $\boldsymbol{\mu}_{\mathsf{G}}$. On the other hand, the $\boldsymbol{\Sigma}_{\mathsf{LW}}$ used to form the portfolio does not capture the risk of that same portfolio particularly well. Its expected volatility is 9.85 versus ex post standard deviation of 15.55. The ex ante left tail at the 1% VaR is populated with 5.67% of OOS observations (versus 0.83% for the plain Galton). The portfolio performs well but does not meet expectations. As predicting risk is an important use of a covariance matrix, $\boldsymbol{\Sigma}_{\mathsf{LW}}$ does not dominate $\boldsymbol{\Sigma}_{\mathsf{G}}$ in this comparison.

This exercise also allows a direct comparison of the marginal impact of replacing $\boldsymbol{\mu}_{\mathsf{JS}}$ with $\boldsymbol{\mu}_{\mathsf{G}}$ for an investor using Ledoit and Wolf (2004a) to model covariances. In table 2 the MV portfolio with $\boldsymbol{\Sigma}_{\mathsf{LW}}$ and $\boldsymbol{\mu}_{\mathsf{JS}}$ has a Sharpe ratio of 0.11. The combination of the same covariance matrix with $\boldsymbol{\mu}_{\mathsf{G}}$ achieves a considerably higher Sharpe ratio (0.53). So, $\boldsymbol{\mu}_{\mathsf{G}}$ seems preferable to $\boldsymbol{\mu}_{\mathsf{JS}}$ in this comparison.

7.3 Evidence from 1,000 horse races

The MVE results above are based on an unique sequence of 600 OOS monthly returns with a stock universe restricted to the 50 largest firms by capitalization. In spite of the long test period, sampling error remains a concern – especially as the composition of the stock universe shows little variation by construction. To address this, we simulate 1,000 horse races of the same length, randomly renewing the stock universe each year. This results in 50,000 stock universes with a total of 600,000 OOS returns. All returns are OOS, so within each sequence, the investor following a strategy only uses data available up to the month in question. These simulations use real stock data to generate each portfolio OOS returns and make no assumptions on the distribution.

[Insert table 5 about here]

Table 5 shows the performance summary of the 1,000 races. We bias the comparison by selectively omitting worse performing horses (of a similar type) in Table 2. We keep the best performing methods: KZ, TZ (both unscaled), and the LW, EG, sample (Markowitz) covariance matrices equipped with $\boldsymbol{\mu}_{\mathsf{JS}}$ return expectations. We also keep the naive strategies as well as the Jorion strategy.

The Galton achieves the highest Sharpe ratio (0.44) on average in the 1000 horse races among the nine methods. The second column tests the null that each method is able to produce Sharpe ratios at least as high as the Galton in half of the races.[21] This is rejected at a 1% significance level for all methods. Thus, the Galton performance in terms of Sharpe ratio seems unlikely to be the product of chance.

Columns 3 and 4 compare the expected and realized volatility of the portfolios. The methods compared have a tendency to produce overly optimistic risk estimates. The historical Markowitz, for example, has on average an ex post volatility about 40 times its ex ante estimate (198.31/4.99=39.74). The exception is the Galton where ex ante volatility is slightly higher than ex post (19.04 versus 18.99). Column 5 shows, for all methods, a clear rejection of the null that ex post volatility is not greater than expected. The exception is the Galton. In fact, as this method has, on average, higher expected risk than ex post realizations, we do not report its p-value. Interestingly, performing the reverse test, the conservative bias of the method would be statistically significant. So the Galton tends to err on the conservative side.

Columns 7 and 8 test the accuracy of each method estimating volatilities and expected returns. It is noteworthy that the unscaled TZ method has much less accuracy, on average, than reported in table 2. For instance, its RMSFE for the expected return is 1197.23 versus 4.90 in that table. In unreported results, we find this method has a median RMSFE of only 7.85 but numerous and extremely large outliers in the right tail of the

[21] The test statistic for this is $\frac{Sharpe \geq Galton - 0.5}{\sqrt{0.5^2/1000}}$ where $Sharpe \geq Galton$ is the proportion of horse races where the method achieves a Sharpe ratio at least as high as the Galton.

distribution. So, the method typically has reasonable forecasts, but at times these deviate immensely. This underscores the value in looking at many horse races to better gauge the average accuracy of a method. The Galton shows the smallest RMSFE for both volatilities and expected returns. So, this method tends to produce portfolios that deliver close to expected performance.

The final column shows the hit rates at the 5% VaR. The average hit rate for the Galton is 4.53%. This is less than one standard deviation away from the target. Thus, we cannot reject that the Galton is estimating the 5% quantiles correctly in these 1,000 horse races. For other methods, the actual hit rates are more than 2 standard deviations away from 5% for all. This shows estimated 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 CCPs.

Collectively, the evidence in these 1,000 horse races suggests the Galton tends to produce portfolios with Sharpe ratios that tend to be larger than their peers. Furthermore, the realized performance ex post is relatively close to ex ante expectations.

7.4 GMV portfolios

GMV portfolios do not require estimating expected returns. So, analyzing the Minimum Variance problem nicely isolates the contribution made by modeling covariances alone.

Kan and Zhou (2007) and Tu and Zhou (2011) are specifically MV methods, so we do not include them in this exercise. On the other hand, Jagannathan and Ma (2003) (JM) propose a GMV using $\mathbf{\Sigma}_{\mathsf{H}}$ but with a non-negativity constraint. Given the pertinence of this method here, we include it in the comparison. As a result, our comparison set has five optimized (Galton, Markowitz, EG, LW, and JM) and two naïve portfolios (EW and VW).

[Insert table 6 about here]

Table 6 shows the performance (and accuracy of estimates) of optimized GMV portfolios.

The contrast with the MV results is noteworthy. Now, four of the five optimized portfolios outperform considerably both the EW and the VW in terms of Sharpe ratio. This ilustrates the well-known advantages of avoiding estimation error issues in μ (see Jorion (1985), Jorion (1986), Jagannathan and Ma (2003), among others).[22] The case for portfolio optimization seems much more solid with the GMV.

The exception is the Markowitz-GMV with a Sharpe ratio of 0.19. Jorion (1986) shows with simulations this portfolio outperforms VW unless N is too large. With N=50 and H=60, it is likely that the number of degrees of freedom per parameter in the covariance matrix is not sufficient to make this method robust enough in our implementation.

Nevertheless, Galton has the highest Sharpe ratio in the comparison set.[23] Also, with a STD of the mean of 3.09, the Galton is less concentrated than other optimized methods and its concentration is relatively stable.

Panel B shows LW, JM and the Galton achieve low volatility OOS – in the range of 12.22 to 12.57. Moreover, the GMV portfolios in table 6 have considerably less extreme positions than MV portfolios in table 2. This shows up in less severe problems of estimating volatility. Nevertheless, all methods except Galton underestimate risk to some extent.

The Markowitz historical method is especially misleading about 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 Markowitz 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 sample GMV is of 3.66 percentage points (annualized) but the ex post standard deviation of the strategy is 25.32. There are losses greater than the 1% VaR threshold between 4.50% of the time, for JM, and 33.67%, for the Markowitz-GMV.

[22]The GMV also tends to overweight "low risk" stocks, with low volatility and / or correlation to other stocks. Several studies document high risk-adjusted returns for such portfolios (Ang et al. (2009), Frazzini and Pedersen (2014). This should contribute to the overall performance of GMV portfolios.

[23]If one interprets the Galton as a shrinkage method akin to the positive part James-Stein estimator (16), then the shrinkage coefficient should be constrained to be positive. The negative point estimate for the slope in expected returns will imply a shrinkage coefficient of zero for the mean – in this case this Galton GMV portfolio is also the Galton MV portfolio.

By contrast, for the GMV portfolio, the expected standard deviation is 15.75 percentage points while the standard deviation in the OOS period is 12.22. So risk is actually lower OOS than expected for this portfolio. None of the hit rates is significantly higher than the respective target rate. Looking at both tails, the "Galton" errs by being excessively conservative. Over this 50-year OOS sample period, it has 1.17% losses exceeding the 1% VaR.

The expected returns are, on average, close to ex post realizations. The RMSFE shows more modest forecast errors for all methods when compared to those seen in table 2. The GMV portfolios, being less extreme than their MV counterparts, have returns easier to forecast. The lowest error is for Galton (3.53) and the highest for Markowitz (7.33).

7.5 Optimizing portfolios of portfolios

We focus on individual securities in our main analysis as they have more noise to filter out and investors use them in practice to form their portfolios. Still, portfolios formed on characteristics are an important set of investable assets. Some of these sets are readily available for long historical samples and the sorting methodology produces, by construction, assets with more stable characteristics for the optimization. Hence, for robustness, we examine the OOS performance of the Galton correction with portfolios sorted on different characteristics available on Kenneth French's website.

Table 7 shows the slope of a regression of ex post realizations of optimization inputs on their ex ante estimates from the historical sample.

[Insert table 7 about here]

Generally, pairwise correlations do not regress toward the mean as much as for individual stocks. The slope is significantly below one for double-sorted portfolios on operating profitability and investment (panel B) and on size and momentum (panel D) but not for the other two sets. There is no evidence of a negative relation between historical and ex post returns in any of the sets. For three of these, the slope coefficient is significantly positive. Still, for all mean returns the slope is below one with statistical significance. So, there is a prevalent, albeit partial, regression to the mean in returns. These results suggest the sorting exercises are successful in creating a set of assets with strong risk and return structures.

But there are important differences in persistence between the sets of portfolios. For instance, returns are highly predictable in the size and momentum set. The R-square of the regression for returns is the highest with 29.97%. The slope coefficient is also closer to unity than 0, unlike other sets.

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

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

For each set of assets, we follow DeMiguel et al. (2009b) and use a rolling historical window of 120 months. On top of this, for the initial optimization, Galton requires 120 months for the learning window. Therefore, the OOS period begins 20 years after the start of the sample for each set of assets. This start of the sample is 1926:07 for size / book-to-market (ME-BTM), 1927:01 for size / momentum (ME-MOM) portfolios and 1963:07 for operating profitability / investment (OP-IN) and size / beta (ME-BETA) double-sorted portfolios.

[Insert table 8 about here]

We focus here, for space and readability, on two key statistics: the OOS Sharpe ratio and the ratio of ex post volatility to ex ante expectation. [24] Table 8 shows the results for 24 portfolios: 2 naïve portfolios, 10 optimized with scaled weights (3 of which with $\boldsymbol{\mu}_{\mathsf{JS}}$), 7 with unscaled weights, and 5 GMV portfolios.

The naïve portfolios produce a consistent performance in terms of Sharpe ratio between 0.53 and 0.59, depending on the test assets. The methods using $\boldsymbol{\mu}_{\mathsf{JS}}$ generally perform better than using $\boldsymbol{\mu}_{\mathsf{H}}$ and the naïve portfolios. This illustrates the importance of shrinking expected returns to achieve less over-fitting as shown in Jorion (1986).

Comparing across sets, Galton shows an overall consistent performance in terms of Sharpe ratio. It has an above median performance in scaled, unscaled, and GMV portfolios. It has the highest Sharpe ratio for the scaled portfolios with ME-BTM and OP-IN, the unscaled portfolios with OP-IN and ME-MOM, and the GMV portfolios for OP-IN and ME-BETA.

Notwithstanding, several other methods show interesting performance. Using unscaled weights, as in the original Kan and Smith (2008), makes a significant difference. Its Sharpe ratio for ME-BTM, for example, increases from -0.09 to 1.07. Its performance across sets is consistently above the median in the set of unscaled portfolios. Tu and Zhou (2011) show a consistent performance as well and both for unscaled and scaled portfolios. Interestingly, the plain Markowitz GMV with $\boldsymbol{\Sigma}_{\mathsf{H}}$ outperforms LW and JM GMVs for most sets, in contrast to its relative performance with individual stocks in table 6.

Another striking result is that GMV portfolios do not necessarily dominate their MV counterparts (as typically found for individual stocks). For instance, in the ME-BTM set, 6 optimized MV portfolios outperform the best GMV portfolio in terms of Sharpe ratio. The evidence of the Galton regression in table 7 provides an explanation for this non-trivial result: for some of these sets there is significant return predictability. Consequently, GMV portfolios that ignore this predictability are at a significant disadvantage.

For ME-MOM, it is noticeable that 16 out of 17 optimized MV portfolios have a

[24] We do not report turnover statistics as these misleadingly ignore the turnover of the portfolios themselves on underlying individual assets. We also omit here, for brevity, the hit rates and RMSFE. A more complete set of results is available in the Internet Appendix.

Sharpe ratio greater or equal than naïve portfolios. This is a starking contrast with the results in table 2. The Galton regression results in table 7 help understand why MVs are so successful with this set: these assets have the strongest predictability in returns.

Columns 4 to 8 show the results for the risk accuracy of each method. Specifically, we compare the ratio of ex post risk to its ex ante estimate. This is a natural comparison where we examine the ability of each method to forecast the risk of its respective optimized portfolio. This tests if methods produce optimized portfolios that meet expectations.[25]

In terms of risk accuracy, all methods underestimate the risk of optimized portfolios out of sample. The problem is most severe for portfolios not shrinking $\boldsymbol{\mu}$. For instance, the Markowitz portfolio with $\boldsymbol{\mu}_{\mathsf{H}}$ and test assets of ME-BTM has an ex post volatility that is, on average, 13.63 times higher than its ex ante estimate.

For scaled and unscaled MVE portfolios, Galton has the lowest ratio of ex post volatility to ex ante estimates in three of the four sets (OP-INV, ME-BETA, and ME-MOM). This suggests the ability to design portfolios that (almost) meet expectations is a somewhat robust feature of the method. In the Internet Appendix, we show this accuracy results in more than proportional gains estimating extreme quantiles as VaR at the 1% and 5% level.

For GMV portfolios, the best performing method estimating risk is Jagannathan and Ma (2003). The ratio of ex post volatility to ex ante estimates is only between 1.03 and 1.06, very accurate estimates. Yet, the covariance matrix for this method is $\boldsymbol{\Sigma}_{\mathsf{H}}$. This same matrix performs poorly estimating the volatility of Markowitz MV portfolios, especially with $\boldsymbol{\mu}_{\mathsf{H}}$. This suggests the accuracy estimating the volatility of Jagannathan and Ma's portfolio is due to its less extreme nature when non-negativity constraints are imposed. Our results thus underscore an additional shrinkage argument for imposing weight constraints– after shrinkage, reliability of performance forecasts increases.

In conclusion, this robustness exercise shows that the benefits of the Galton method

[25] Yet, this comparison does not necessarily isolate the accuracy of the estimator for the covariance matrix as the optimized portfolio being compared is not the same across rows.

are not specific to portfolios of individual stocks. In fact, the Galton method shows interesting flexibility and for each set of test assets it seems to produce sensible corrections for both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This results in Sharpe ratios, volatility estimates and accurate hit rates in the left tail. All are at least at par with other methods in the comparison set.

8 Conclusion

Two important questions in portfolio optimization are: how to construct portfolios that deliver the expected performance? And how can we best forecast the expected performance of different portfolios?

Shrinkage has long been recognized as a method that can help improve estimates of portfolio optimization inputs. However, shrinkage estimators have usually struggled to perform better than even naive strategies such as the “1 over N” portfolio. Even when they do outperform, they face another often overlooked problem in portfolio management: estimated covariance matrices, even when successful at picking portfolios with interesting Sharpe ratios, systematically fail to capture the risk of those same portfolios. Optimized portfolios typically do not deliver average performance or risk realizations in line with the ex-ante expected values. The problem is especially acute for the classical plug-in approach that uses sample moments as population counterparts.

This paper presents a method that helps overcome these issues and construct portfolios that deliver robust performance which, on average, meets expectations. The method is intuitive: instead of estimating shrinkage parameters using prior assumptions, it estimates them by minimizing historical OOS forecast errors. We suggest using such shrinkage parameters to forecast portfolio optimization inputs.

The resulting “Galton” optimized portfolios perform quite well OOS, and compare favorably to more than 15 competing approaches. Moreover, correcting the sample covariance matrix for past OOS errors dramatically reduces the risk estimation problem. The ex post risk of optimized portfolios is close in magnitude to ex ante estimates. For most of the extreme quantiles, the hit rates are either not statistically different from the respective targets or even below them.

The central insights motivating this method, that we should learn from past OOS forecast errors, have more general applications that we do not fully explore in this work. Firstly, the unique agnostic character of the Galton makes it a potential complement to other methods with established useful results. For instance, it could prove worthwhile to examine the Galton's potential when correcting OOS errors of more successful estimation methods such as Ledoit and Wolf (2004a).

Secondly, the Galton implementation in this study does not impose asset pricing restrictions. But it should be quite feasible to use Galton-corrected versions of approaches that rely on an asset-pricing model for establishing a prior such as Pástor (2000) and Pástor and Stambaugh (2000).

Thirdly, we chose an agnostic setting that does not assume any factor structure in the data. This choice reflects the current state of the literature where the right factor model is a matter of particularly contentious debate. Nevertheless, refraining from using any knowledge at all on factors is likely an excessive level of caution as well. The expected returns used by the Galton could equally well come from informative factor models such as Fama and French (2015) or Hou et al. (2015). We leave such directions of research for future work.

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.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, 91(1):1–23.

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

Basak, G. K., Jagannathan, R., and Ma, T. (2005). Estimating the risk in sample efficient portfolios.

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.*

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.

Cochrane, J. H. (2001). Asset pricing.

Council of European Union (2012). Regulation (eu) no 648/2012 of the european parliament and of the council of 4 july 2012 on otc derivatives, central counterparties and trade repositories.
`https://eur-lex.europa.eu/eli/reg/2012/648/oj`.

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.

Efron, B. and Morris, C. (1972). Empirical bayes on vector observations: An extension of stein's method. *Biometrika*, 59(2):335–347.

Elton, E. J. and Gruber, M. J. (1973). Estimating the dependence structure of share prices—implications 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.

Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1):1–25.

Frost, P. A. and Savarino, J. E. (1986). An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(03):293–305.

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

Garlappi, L., Uppal, R., and Wang, T. (2006). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1):41–81.

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.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of finance*, 58(4):1651–1684.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 361–379.

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.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

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.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):294–419.

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

Mehra, R. and Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of monetary Economics*, 15(2):145–161.

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

Pástor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223.

Pástor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381.

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.

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(1):155–160.

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

Tu, J. and Zhou, G. (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.

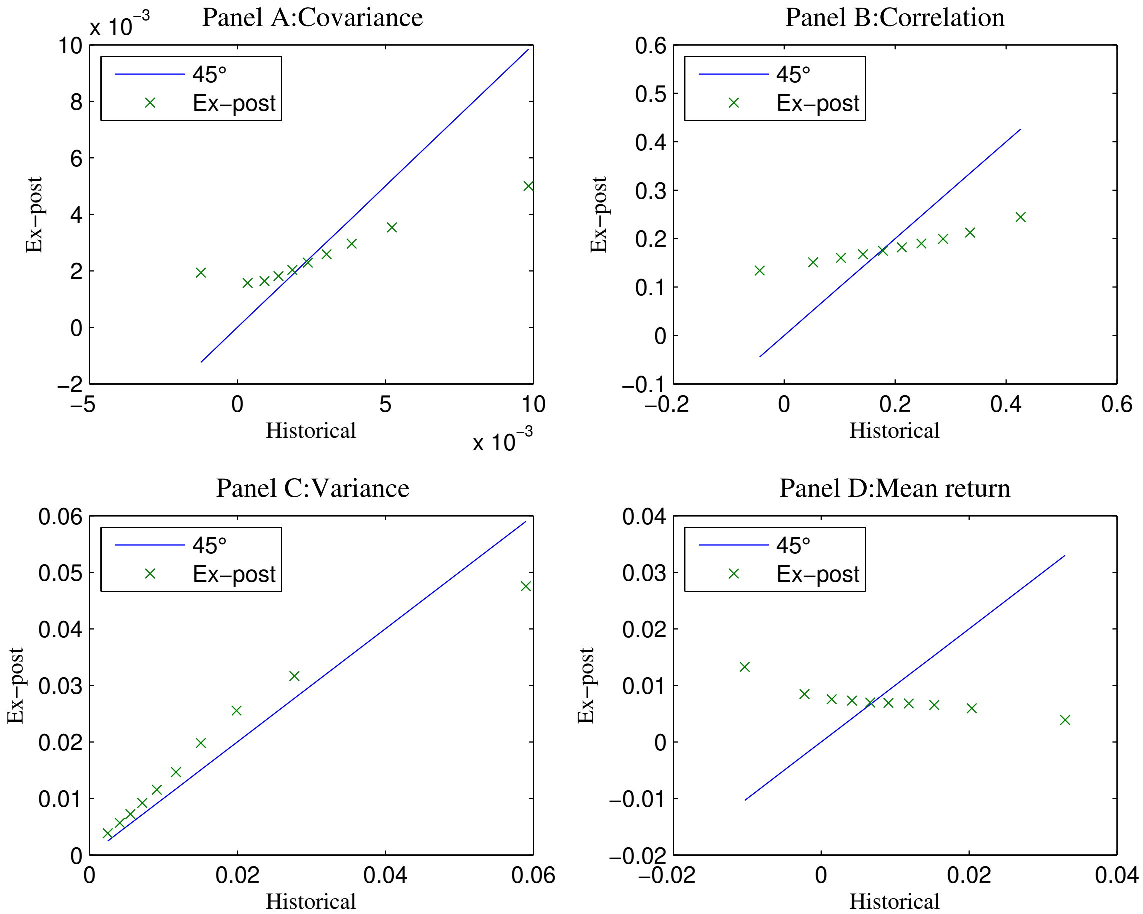

Fig. 1. Historical versus ex post realizations (excess returns). For each moment in time and observation (either a stock or a pair of stocks) we compute the value of a variable in the historical sample - the previous 60 months of observations -and its respective value in the future -the subsequent 12 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 excess returns of individual stocks. Panel B shows the same comparison for the pairwise correlations of excess returns. Panels C and D show the same comparison for, respectively, the variance of stocks and the mean total return. The data consists of monthly observations from 1952:01 to 2017:11.

Table 1

Regression to the mean

Regression of ex post values on historical estimates. For each month in the sample from 1952.01 to 2016.11, we 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	0.00	0.13	0.01	0.01
t-stat (=0)	8.63	11.20	8.63	5.87
Slope	0.36	0.28	0.55	-0.16
t-stat(=0)	10.30	17.48	11.43	-4.88
Greater than 0 (%)	1.00	1.00	0.98	0.21
t-stat(=1)	-17.94	-44.69	-9.46	-35.21
Smaller than 1 (%)	0.93	1.00	0.88	1.00
R-square (%)	4.18	3.11	12.17	1.94
Min	392941.00	392941.00	887.00	887.00
Average	5412795.63	5412795.63	2991.97	2991.97
Max	10720765.00	10720765.00	4631.00	4631.00

Table 2

OOS performance of the MVE portfolio.

At the end of every year, we 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, H=60, and the subsequent 12 months, E=12. The sample is kept fixed for the subsequent twelve months. The weights are re-balanced monthly, each month we use a rolling window of 60 months to estimate the covariance matrix and expected return vector, hence obtaining the mean-variance ('MV') portfolios for various methods The columns show descriptive statistics of the out-of-sample performance of each portfolio: sharpe ratio, active share, turnover, and various statistics calculated using the weight vector from each time period. In Panel B, the first two columns show the mean expected and realized returns. The next column reports the root mean squared forecast error (RMSFE) for returns on the MV portfolio. The next two columns report the mean expected volatility and the realized volatility. The columns r< (>)Qz(.) show how often the strategy delivered returns smaller (greater) than the 1st (99th), and 5th (95th) quantile, respectively, according to the ex-ante estimate of volatility and assuming a normal distribution. The evaluation sample to construct portfolios and compare performance of various strategies is from 1962.01 to 2016.12. The Galton strategy uses the additional parameters L=108 and E=12 and therefore requires an additional burn-in sample from 1952.01 to 1961.12. 60 months of data for each optimized strategy is required before the first horse races, so the horse race period is from 1967.01 to 2016.12.

Panel A: The mean variance portfolios - performance statistics

Strategies	Sharpe	Active Share	Turnover	Statistics for portfolio weights (w_i) and their stability				
				[Min w_i	Max w_i	Mean[$\sigma_t(w_i)$]	Std[$\sigma_t(w_i)$]	$\sum w_i I(w_i < 0)$]
Value weighted	0.37	0.00	2.18	0.69	9.17	1.66	0.51	0.00
Talmud 1/N	0.38	26.51	6.59	2.00	2.00	0.00	0.00	0.00
Galton	0.47	120.76	78.35	-13.42	15.63	6.14	1.66	-78.63
Markowitz	-0.27	6630.02	21427.91	-1001.13	927.99	364.13	2332.89	-6612.71
Jorion	-0.27	5594.37	19675.90	-839.44	782.75	306.75	2000.63	-5577.02
Kan Zhou	-0.26	1666.43	5697.77	-245.22	234.32	90.75	547.79	-1648.00
Tu Zhou (CKZ)	0.18	612.90	3632.82	-121.65	91.38	35.12	174.29	-574.67
Elton Gruber	-0.12	1858.39	8481.46	-212.27	223.71	94.42	786.50	-1834.89
Ledoit Wolf	0.12	3050.42	7763.35	-328.27	428.54	155.57	1753.93	-3016.85
			Strategies with shrunk means using James-Stein estimator					
Markowitz	-0.02	554.08	1432.85	-91.56	97.91	31.98	46.26	-508.30
Elton Gruber	-0.16	169.50	226.81	-13.29	31.82	9.19	36.89	-121.96
Ledoit Wolf	0.11	125.91	116.77	-9.97	28.61	7.28	5.08	-80.33
			Strategies with investment in the riskless rate					
Galton	0.46		104.48	-17.04	20.48	7.93	2.98	-99.65
Markowitz	-0.04		666410.46	-7722.72	5742.51	2196.04	2096.83	-35556.05
Jorion	-0.03		24191.00	-902.61	676.84	256.89	262.50	-4150.48
Kan Zhou	0.05		402.38	-52.33	43.82	15.60	23.52	-249.82
Tu Zhou (CKZ)	0.18		228.13	-27.86	24.80	8.63	11.34	-126.48
Elton Gruber	-0.06		5455.74	-169.67	194.58	76.81	22.66	-1397.89
Ledoit Wolf	-0.01		7657.48	-170.49	207.34	77.33	16.61	-1363.44

Panel B: The mean variance portfolios - expectation vs realization statistics									
	Excess Returns (r)			Standard Deviation (σ) and corresponding 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		5.51			14.80				
Talmud 1/N		5.92			15.37				
Galton	13.32	8.60	5.29	19.03	18.27	0.83	4.33	3.67	1.17
p-value		(0.07)				(0.68)	(0.45)	(0.13)	(0.68)
Markowitz	398.28	-253.57	431.91	54.08	933.51	49.33	54.67	26.17	22.33
p-value		(0.00)				(0.00)	(0.00)	(0.00)	(0.00)
Jorion	281.36	-214.58	346.61	122.78	791.95	23.17	30.67	16.17	11.50
p-value		(0.00)				(0.00)	(0.00)	(0.00)	(0.00)
Kan Zhou	89.49	-58.66	102.07	12.23	221.64	41.83	47.50	30.50	28.17
p-value		(0.00)				(0.00)	(0.00)	(0.00)	(0.00)
Tu Zhou (CKZ)	36.50	39.28	52.81	20.08	219.52	11.50	16.83	12.17	6.83
p-value		(0.91)				(0.00)	(0.00)	(0.00)	(0.00)
Elton Gruber	420.71	-39.53	315.40	138.81	318.65	19.17	31.17	6.17	3.33
p-value		(0.00)				(0.00)	(0.00)	(0.19)	(0.00)
Ledoit Wolf	651.56	970.18	1612.96	204.59	7892.81	15.50	26.33	4.17	1.33
p-value		(0.69)				(0.00)	(0.00)	(0.35)	(0.41)
				Strategies with shrunk means using James-Stein estimator					
Markowitz	14.74	-2.02	28.64	4.46	95.83	34.33	39.50	36.83	32.00
p-value		(0.23)				(0.00)	(0.00)	(0.00)	(0.00)
Elton Gruber	16.57	-26.24	50.24	12.76	162.61	10.83	16.00	12.33	5.67
p-value		(0.08)				(0.00)	(0.00)	(0.00)	(0.00)
Ledoit Wolf	13.64	2.39	6.60	10.17	21.05	8.00	13.83	9.33	4.17
p-value		(0.00)				(0.00)	(0.00)	(0.00)	(0.00)
				Strategies with investment in the riskless rate					
Galton	17.97	10.74	6.74	23.89	23.17	0.83	4.33	3.67	1.17
p-value		(0.03)				(0.68)	(0.45)	(0.13)	(0.68)
Markowitz	1992.88	-90.86	718.93	248.80	2402.49	49.33	54.67	26.17	22.33
p-value		(0.00)				(0.00)	(0.00)	(0.00)	(0.00)
Jorion	199.66	-7.55	87.91	77.10	296.92	23.17	30.67	16.17	11.50
p-value		(0.00)				(0.00)	(0.00)	(0.00)	(0.00)
Kan Zhou	11.35	1.19	7.46	1.58	25.71	41.83	47.50	30.50	28.17
p-value		(0.01)				(0.00)	(0.00)	(0.00)	(0.00)
Tu Zhou (CKZ)	9.08	2.90	4.90	7.32	16.31	11.50	16.83	12.17	6.83
p-value		(0.01)				(0.00)	(0.00)	(0.00)	(0.00)
Elton Gruber	343.92	-13.48	76.33	105.18	238.75	19.17	31.17	6.17	3.33
p-value		(0.00)				(0.00)	(0.00)	(0.19)	(0.00)
Ledoit Wolf	309.34	-2.07	54.74	100.49	165.00	15.50	26.33	4.17	1.33
p-value		(0.00)				(0.00)	(0.00)	(0.35)	(0.41)

Table 3

Robustness of the Galton MVE portfolio performance.

At the end of every year, we select the N stocks of the firms with the largest market capitalization for which there is a complete return history over the previous H months and the subsequent E months. The sample is kept fixed for the subsequent twelve months. The weights are re-balanced monthly, each month we use a rolling window of H months to estimate the covariance matrix and expected return vector, hence obtaining the mean-variance portfolios for various methods The columns show descriptive statistics of the out-of-sample performance of each portfolio. In Panel B, the columns r< (>)Qz(.) show how often the strategy delivered returns smaller (greater) than the 1st (99th), and 5th (95th) quantile, respectively, according to the ex-ante estimate of volatility and assuming a normal distribution. The sample returns are from 1952.01 to 2016.12. The first ten years (120 months) are used as a burn-in period by the Galton procedure. The learning period L is 120 - E months. The evaluation sample to compare performance of various strategies is from 1962.01 to 2016.12. Five years of data is required before the first horse races, so the horse race period is from 1967.01 to 2016.12.

Panel A: The mean variance portfolios - performance statistics

H	E	L	N	Sharpe	Active Share	Turnover	Statistics for portfolio weights (w_i) and their stability				
							[Min w_i	Max w_i	Mean[$\sigma_t(w_i)$]	Std[$\sigma_t(w_i)$]	$\sum w_i I(w_i < 0)$]
60	12	108	30	0.42	81	50.56	-12.12	17.78	6.82	2.00	-41
60	12	180	50	0.45	124	82.02	-14.03	15.80	6.29	1.68	-81.79
60	60	60	50	0.38	144	95.21	-17.06	18.23	7.39	2.13	-102
120	12	108	50	0.44	136	76.75	-14.92	18.37	6.99	1.89	-94
120	12	108	75	0.43	169	106.35	-13.52	16.15	5.81	1.37	-126
180	12	108	50	0.57	142	73.00	-14.45	19.54	7.27	2.52	-97.36
180	12	180	50	0.66	130	65.87	-13.08	17.05	6.56	1.60	-84.68
180	12	180	100	0.67	188	111.53	-10.29	15.08	4.75	0.95	-139.80

Panel B: The mean variance portfolios - expectation vs realization statistics

				Excess Returns (r)		Standard Deviation (σ) and corresponding Hit Rates					
H	E	L	N	Expected r	Realized r	Expected σ	Realized σ	r<Qz(1%)	r<Qz(5%)	r>Qz(95%)	r>Qz(99%)
60	12	108	30	11.71	7.01	18.93	16.74	1.17	3.67	3.50	0.83
60	12	180	50	13.22	8.56	19.76	18.92	0.95	4.17	3.60	1.33
60	60	60	50	14.39	7.92	19.82	20.68	1.81	4.89	4.71	1.63
120	12	108	50	13.70	7.92	18.39	17.84	1.30	4.26	2.41	0.93
120	12	108	75	14.81	8.20	17.90	19.23	1.67	5.93	3.70	1.67
180	12	108	50	11.44	8.75	17.72	15.31	1.46	2.50	2.92	0.62
180	12	180	50	10.80	9.94	16.94	15.15	1.72	2.70	3.19	0.49
180	12	180	100	12.21	10.62	15.70	15.88	1.96	4.41	4.17	1.47

Table 4

OOS performance of the MV portfolio constructed using Hybrid approaches.

At the end of every year, we 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 we use a rolling window of 60 months to estimate the covariance matrix and the expected return vector, hence obtaining the mean variance efficient portfolio for various methods. The 'Galton (Covariance)' method uses a covariance matrix in which covariances are shrunk instead of correlations. The 'Galton Mean Only' method uses Galton mean forecasts along with the sample covariance matrix. The 'Galton Cov Only' use the Galton Covariance matrix along with sample averages as mean forecasts. The 'Galton Mean LW (CC) Cov' method uses the Galton mean forecast along with the LW shrunk covariance matrix. The columns show descriptive statistics of the out-of-sample performance of each portfolio. The sample returns are from 1952.01 to 2016.12. The first ten years are used as a burn-in period by the Galton procedure (L=108, E=12). The evaluation period to compare performance of various strategies is from 1962.01 to 2016.12. Five years of data is required before the first horse races, so the horse race period is from 1967.01 to 2016.12.

Panel A: The mean variance portfolios - performance statistics

Strategies	Sharpe	Active Share	Turnover	Statistics for portfolio weights (w_i) and their stability				
				[Min w_i	Max w_i	Mean[$\sigma_t(w_i)$]	Std[$\sigma_t(w_i)$]	$\sum w_i I(w_i < 0)$]
Value weighted	0.37	0.00	2.18	0.69	9.17	1.66	0.51	0.00
Talmud 1/N	0.38	0.27	6.59	2.00	2.00	0.00	0.00	0.00
Galton (Correlation)	0.47	120.76	78.35	-13.42	15.63	6.14	1.66	-78.63
				Variations of the Galton strategy				
Galton (Covariance)	0.43	124.71	84.33	-14.29	15.82	6.30	2.77	-81.45
Galton Mean Only	0.24	542.71	587.80	-86.10	109.81	32.34	24.07	-492.64
Galton Cov Only	-0.21	1592.86	3336.55	-172.19	201.11	81.12	429.11	-1548.34
Galton LW (CC) Cov	0.53	133.94	79.88	-9.89	30.33	7.58	1.64	-85.16

Panel B: The mean variance portfolios - expectation vs realization statistics

	Excess Returns (r)			Standard Deviation (σ) and corresponding 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		5.51			14.80				
Talmud 1/N		5.92			15.37				
Galton (Correlation)	13.32	8.60	5.29	19.03	18.27	0.83	4.33	3.67	1.17
p-value		(0.07)				(0.68)	(0.45)	(0.13)	(0.68)
				Variations of the Galton strategy					
Galton (Covariance)	13.56	8.17	5.54	19.71	19.13	1.00	4.33	3.50	0.83
p-value		(0.05)				(1.00)	(0.45)	(0.09)	(0.68)
Galton Mean Only	12.54	7.75	9.35	4.31	32.61	35.50	40.33	37.00	32.83
p-value		(0.29)				(0.00)	(0.00)	(0.00)	(0.00)
Galton Cov Only	17.06	1.99	11.49	19.42	37.22	2.17	5.00	2.33	0.67
p-value		(0.01)				(0.00)	(1.00)	(0.00)	(0.41)
Galton LW (CC) Cov	12.02	8.27	4.50	9.85	15.55	5.67	12.67	10.50	4.83
p-value		(0.09)				(0.00)	(0.00)	(0.00)	(0.00)

Table 5

Results of 1000 Random Samples

At the start of every year, a sample is constructed by randomly selecting a set of stocks from the 100 largest stocks. This process is repeated from 1967 to 2016. Five years of data are used to construct optimal portfolios with information available up to each moment in time. Therefore the evaluation sample is from 1962.12 to 2016.12. The portfolios are rebalanced monthly, reflecting the update in information available to investors. So a simulated random sample comprises 50 stock universes and a total of 600 OOS monthly returns. The results are based on 1000 such simulated random samples, totaling 600,000 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 Galton 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 reports the proportion of samples where realized volatility is larger than expected volatility. Column 6 shows the root mean squared forecast error (RMSFE) of average expected volatility when predicting ex post volatility across 1000 random samples. Column 7 shows the average RMSFE of expected returns when predicting ex post returns across the random samples. Column 8 reports the average hit rate (for returns $\leq$ 5%) across the random samples. The standard deviation of these statistics across the random samples are reported in the rows below. The values in columns 3 to 8 are in percentage points. This table only reports the results for the better performing version of each method in Table 2. 'JS' denotes strategies where the means are shrunk using the positive part James-Stein estimator. 'unscaled' denotes strategies which invest in the riskless rate in addition to risky assets.

Sample of 50 stocks selected from universe of 100 largest stocks

Strategies	Sharpe	Sharpe $\geq$ Galton (p-value below*)	Expected σ	Realized σ	Realized $\sigma \geq$ Expected σ (p-value below*)	RMSFE (σ)	RMSFE (mean)	r<Qz(5%)
Value weighted	0.36	0.07		15.28				
std	(0.03)	(0.00)		(0.20)				
Talmud 1/N	0.36	0.08		15.92				
std	(0.02)	(0.00)		(0.16)				
Galton	0.44	1.00	19.04	18.99	0.46	0.51	5.50	4.53
std	(0.06)		(0.09)	(0.53)			(0.15)	(0.54)
Markowitz (JS)	0.04	0.00	4.99	198.31	1.00	962.70	59.75	42.91
std	(0.14)	(0.00)	(3.52)	(946.45)	(0.00)		(299.69)	(1.81)
Jorion	-0.02	0.00	77.09	277.37	1.00	203.46	82.88	33.16
std	(0.13)	(0.00)	(2.71)	(36.86)	(0.00)		(10.30)	(1.72)
Kan Zhou (unscaled)	0.03	0.00	1.63	21.19	1.00	20.07	6.29	48.90
std	(0.13)	(0.00)	(0.11)	(4.53)	(0.00)		(1.27)	(1.96)
Tu Zhou (CKZ) (unscaled)	0.10	0.01	172.73	2131.22	1.00	59350.97	1197.23	18.88
std	(0.14)	(0.00)	(5030.11)	(64378.30)	(0.00)		(36887.29)	(1.41)
Elton Gruber (JS)	-0.04	0.00	14.97	341.52	1.00	2103.64	105.54	16.78
std	(0.11)	(0.00)	(27.58)	(2101.37)	(0.00)		(622.57)	(1.01)
Ledoit Wolf (JS)	0.09	0.01	11.35	82.41	1.00	492.78	26.40	14.38
std	(0.16)	(0.00)	(6.74)	(494.01)	(0.00)		(159.51)	(0.93)

Table 6

OOS performance of the global minimum variance portfolio.

At the end of every year, we 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 we use a rolling window of 60 months to estimate the covariance matrix hence obtaining the global minimum variance ('GMV') for various methods. These covariance matrices are calculated using the historical sample matrix (Markowitz), the Elton-Gruber method, and the Ledoit and Wolf method. The 'Jagannathan Ma' rows report results for the 'GMV' calculated after imposing non-negativity constraints on the weights. The columns show descriptive statistics of the out-of-sample performance of each portfolio. These columns are described in table 2. The sample returns are from 1952.01 to 2016.12. The first ten years are used as a burn-in period by the Galton procedure (L=108, E=12). The evaluation period to compare performance of various strategies is from 1962.01 to 2016.12.

Panel A: The global minimum variance portfolio - performance statistics

Strategies	Sharpe	Active Share	Turnover	Statistics for portfolio weights (w_i) and their stability				
				[Min w_i	Max w_i	Mean[$\sigma_t(w_i)$]	Std[$\sigma_t(w_i)$]	$\sum w_i I(w_i < 0)$]
Value weighted	0.37	0.00	2.18	0.69	9.17	1.66	0.51	0.00
Talmud 1/N	0.38	26.51	6.59	2.00	2.00	0.00	0.00	0.00
Galton	0.60	60.49	31.83	-4.33	10.19	3.09	0.75	-22.62
Markowitz	0.19	441.59	470.99	-68.70	83.98	25.63	12.07	-392.93
Elton Gruber	0.52	105.80	48.59	-4.56	26.32	6.11	1.25	-55.65
Ledoit Wolf	0.57	106.67	58.31	-6.60	27.80	6.37	1.23	-59.06
Jagannathan Ma	0.51	56.11	29.80	0.15	24.67	4.23	1.17	0.00

Panel B: The global minimum variance portfolio - expectation vs realization statistics

	Excess Returns (r)			Standard Deviation (σ) and corresponding 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		5.51			14.80				
Talmud 1/N		5.92			15.37				
Galton	9.24	7.33	3.53	15.75	12.22	1.17	3.33	1.67	0.33
p-value		(0.27)				(0.68)	(0.06)	(0.00)	(0.10)
Markowitz	4.85	4.70	7.33	3.66	25.32	33.67	38.50	39.17	34.17
p-value		(0.97)				(0.00)	(0.00)	(0.00)	(0.00)
Elton Gruber	3.77	7.43	4.19	8.55	14.33	7.17	13.67	16.33	8.83
p-value		(0.07)				(0.00)	(0.00)	(0.00)	(0.00)
Ledoit Wolf	5.26	7.22	3.66	8.73	12.57	5.83	11.33	12.50	6.17
p-value		(0.28)				(0.00)	(0.00)	(0.00)	(0.00)
Jagannathan Ma	8.79	6.26	3.58	9.97	12.28	4.50	8.83	8.83	3.00
p-value		(0.15)				(0.00)	(0.00)	(0.00)	(0.00)

Table 7

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. The start of the sample in this library is 1926.07 for size / book-to-market (ME-BTM) and size / momentum (ME-MOM) portfolios and 1963.07 for operating profitability / investment (OP-IN) and size / beta (ME-BETA) double-sorted portfolios. The horse race period ends in 2016.12 for all panels. We 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. 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.

	Covariance	Correlation	Variance	Mean return
Panel A: Size and value				
Intercept	0.00	0.02	0.00	0.01
t-stat (=0)	0.96	0.27	1.20	3.63
Slope	0.83	0.95	0.83	0.37
t-stat(=0)	7.29	13.79	6.98	3.53
t-stat(=1)	-1.53	-0.78	-1.42	-6.09
R-square (%)	45.23	37.87	46.88	15.99
Panel B: Operating profitability and investment				
Intercept	0.00	0.35	0.00	0.01
t-stat (=0)	2.54	5.82	2.00	2.85
Slope	0.81	0.51	0.87	0.30
t-stat(=0)	6.87	8.71	6.75	3.42
t-stat(=1)	-1.66	-8.37	-1.01	-8.04
R-square (%)	25.40	5.21	32.61	8.44
Panel C: Size and beta				
Intercept	0.00	-0.05	0.00	0.01
t-stat (=0)	0.87	-0.52	0.80	4.42
Slope	0.97	1.03	0.98	0.17
t-stat(=0)	8.20	9.94	7.62	1.01
t-stat(=1)	-0.24	0.25	-0.13	-4.77
R-square (%)	70.34	46.41	71.28	16.99
Panel D: Size and momentum				
Intercept	0.00	0.12	0.00	0.00
t-stat (=0)	3.66	2.50	2.87	1.90
Slope	0.83	0.84	0.84	0.67
t-stat(=0)	7.96	18.31	7.74	8.42
t-stat(=1)	-1.67	-3.49	-1.46	-4.20
R-square (%)	35.18	38.43	37.45	29.97

Table 8

Portfolios of portfolios: OOS performance and risk summary

The out-of-sample performance of the optimized portfolios with different test assets. The test assets in each panel are 25 double-sorted portfolios are obtained from Kenneth French's online data library. The start of the sample in this library is 1926.07 for size / book-to-market (ME-BTM), 1927.01 for size / momentum (ME-MOM) portfolios and 1963.07 for operating profitability / investment (OP-IN) and size / beta (ME-BETA) double-sorted portfolios. The horse race period ends in 2016.12 for all panels. For each set of assets, we follow DeMiguel et al. (2009b) and use a rolling historical window of 120 months.

Strategies	Sharpe Ratio				Ratio of realized σ to average expected σ			
	ME-BTM	OP-IN	ME-BETA	ME-MOM	ME-BTM	OP-IN	ME-BETA	ME-MOM
Value weighted	0.53	0.56	0.57	0.53				
Talmud 1/N	0.57	0.59	0.59	0.54				
Galton	1.03	0.96	0.76	1.41	1.20	1.11	1.12	1.39
Markowitz	-0.11	0.71	-0.08	1.33	13.63	1.32	7.67	1.56
Jorion	-0.11	0.79	-0.06	1.40	12.05	1.19	6.36	1.39
Kan Zhou	-0.09	0.83	-0.03	1.44	13.33	1.40	5.94	1.64
Tu Zhou (CKZ)	0.97	0.84	0.79	1.52	1.17	1.20	1.17	1.35
Elton Gruber	0.35	0.78	0.28	0.19	4.62	1.69	3.56	41.77
Ledoit Wolf	0.14	0.77	0.10	0.84	15.32	1.58	2.62	1.80
Strategies with shrunk means using James-Stein estimator								
Markowitz	0.80	0.77	0.69	1.08	1.39	1.36	1.40	1.44
Elton Gruber	0.44	0.87	0.56	0.64	2.37	1.61	2.41	2.24
Ledoit Wolf	0.65	0.83	0.59	0.95	1.52	1.53	1.64	1.28
Strategies with investment in the riskless rate								
Galton	1.05	0.89	0.68	1.60	1.37	1.16	1.24	1.28
Markowitz	1.02	0.69	0.45	1.51	1.44	1.34	1.43	1.59
Jorion	1.07	0.76	0.53	1.51	1.34	1.22	1.29	1.45
Kan Zhou	1.07	0.77	0.59	1.49	1.65	1.44	1.54	1.71
Tu Zhou (CKZ)	0.54	0.82	0.75	1.49	1.76	1.20	1.23	1.43
Elton Gruber	0.67	0.79	0.34	1.05	1.97	1.55	2.20	2.64
Ledoit Wolf	0.79	0.78	0.27	1.42	1.49	1.51	1.94	1.43
Global minimum variance strategies								
Galton GMV	0.85	0.91	0.71	0.95	1.25	1.08	1.17	1.02
Markowitz GMV	0.87	0.77	0.69	0.97	1.38	1.36	1.40	1.40
Elton Gruber GMV	0.48	0.88	0.56	0.57	2.36	1.61	2.41	1.83
Ledoit Wolf GMV	0.70	0.85	0.59	0.79	1.51	1.53	1.64	1.27
Jagannathan Ma	0.60	0.78	0.68	0.56	1.06	1.06	1.03	1.05