Better Betas

Lisa R. Goldberg[*] Alex Papanicolaou[†] Alex Shkolnik[‡]

October 29, 2018[§]

Blume's 2/3-shrinkage rule for adjusting raw market betas is a staple of the CFA curriculum, and is ubiquitous in the finance industry. While Blume's rule is appealing for its simplicity and universality, it does not adapt to changing markets. Accurate portfolio construction and risk forecasting demand that betas be adjusted differently in calm and stressed periods. This is because betas themselves change in response to market stress. We describe a dynamic adjustment for estimates of betas that keeps up with their movement. We illustrate the power of our adjustment on long-only minimum variance portfolios simulated in calm and stressed markets.

1. Introduction

Finance practitioners have known for decades that estimates of equity market betas tend to be too dispersed. Influential articles by Blume (1971), Vasicek (1973) and Blume (1975) argue that market betas are routinely overestimated for riskier securities and underestimated for securities that are less risky. They consider adjustments of the form

$$\hat{\beta}^{\text{adj}} = c\hat{\beta}^{\text{raw}} + (1-c), \tag{1}$$

where $\hat{\beta}^{\text{raw}}$ is an empirical estimate and the parameter c lies between 0 and 1. Formula (1) shrinks raw betas toward the value 1, which is a reasonable prior because it is the cap-weighted average of the market betas.[1] Setting

$$c^{\text{Blume}} = \frac{2}{3} \tag{2}$$

in formula (1) gives the Blume adjustment, $\hat{\beta}^{\text{Blume}}$.[2] The Blume adjustment is motivated by two effects. The first is non-stationarity, the hypothesis that true market betas change over time. The second is statistical regression to the mean, the phenomenon that an extreme measurement will tend to revert toward an average on subsequent measurements. Blume refers to this phenomenon as *order bias*. Blume betas are a staple of the CFA curriculum and available on Bloomberg terminals. The wide use of the Blume adjustment reflects its simplicity and apparent universality.

The Vasicek adjustment, $\hat{\beta}^{\text{Vasicek}}$ relies on Bayes rule. For each security, Vasicek uses a Bayesian prior that is normal with mean one and standard deviation τ_β to derive a posterior distribution for the adjusted betas. The result is approximately normal with mean $\hat{\beta}^{\text{Vasicek}}$ given by for-

[*]Departments of Economics and Statistics and Consortium for Data Analytics in Risk, University of California, Berkeley, CA and Aperio Group. (lrg@berkeley.edu).

[†]Consortium for Data Analytics in Risk, University of California, Berkeley, CA. (apapanicolaou@berkeley.edu).

[‡]Consortium for Data Analytics in Risk, University of California, Berkeley, CA. Department of Statistics & Applied Probability, University of California, Santa Barbara, CA. (shkolnik@ucsb.edu).

[§]Acknowledgement. We are grateful to Simge Ulucam for providing us with an empirical analysis of how the distribution of US market betas evolves over time. We thank Torben Andersen, Ravi Jagannathan, Bob Korajczyk, Olivier Ledoit and Viktor Todorov for insightful dialogs about the material in this article.

[1]The seminal article on shrinkage is Stein et al. (1956). An accessible treatment with an extensive reference list is Efron (2010).

[2]Despite numerous attributions of the Blume adjustment to Blume (1971) and Blume (1975), the value $c = 2/3$ appears nowhere in those articles and the original source of the value $c = 2/3$ is a mystery to us. The many choices of c that do appear are based on historical regressions. An important conclusion from Blume (1975) is that betas are non-stationary, and that excess dispersion due to statistical error accounts for only a small part of the time variation in estimated betas.

mula (1) with

$$c^{\text{Vasicek}} = \frac{\tau_\beta^2}{\tau_\beta^2 + s^2_{\hat{\beta}^{\text{raw}}}}, \tag{3}$$

where $s_{\hat{\beta}^{\text{raw}}}$ is the ordinary least squares standard error of $\hat{\beta}^{\text{raw}}$.

The Blume and Vasicek corrections are typically applied to time series estimates of market betas. The central role of market betas in the investment process dates back to the introduction of the Capital Asset Pricing Model (CAPM), developed by Treynor (1962) and Sharpe (1964). The CAPM identifies the market factor as the main driver of security expected returns, volatilities and correlations.[3] Ross (1973) and Ross (1976) introduce the arbitrage pricing theory (APT), which is a more general factor-based approach to asset pricing and risk forecasting. Often, APT is implemented with unobserved or *latent* factors. This dynamic implementation can identify new drivers of correlation, like the internet factor in the late 1990s or a climate factor in the 2000s, as they develop.

Principal component analysis (PCA) of a security return covariance matrix is a standard way to identify latent factors.[4] In developed equity markets, the dominant PCA factor is *market-like*, meaning that it has relatively large variance and mostly positive exposures.[5] The exposures to this factor are the PCA-analogs of market betas, and we call them *PCA betas*. In this article, we examine shrinkage estimators analogous to the Blume and Vasicek adjustments for estimated PCA betas.

Like time series estimates of market betas, estimated PCA betas tend to be too dispersed. This tendency is identified in Goldberg, Papanicolaou & Shkolnik (2018), which develops an adaptive version of formula (1) to correct this bias. The GPS correction can be viewed as the analog of the Vasicek adjustment given in formula (3), where market betas are replaced with PCA betas and the squared standard error of observed betas, $s^2_{\hat{\beta}^{\text{raw}}}$, is replaced with $\mathscr{D}^2_{\text{bias}}$, a quantity chosen to mitigate the effect of estimation error on optimized portfolios.[6] Influential studies such as Green & Hollifield (1992) and Jagannathan & Ma (2003) demonstrate the important role that markets betas play in quantitative portfolio construction, and there is a large literature on how estimation error distorts portfolios weights and risk forecasts for optimized portfolios.[7] As demonstrated in Goldberg et al. (2018) and the analysis in this article, the same is true for PCA betas.

We extend Goldberg et al. (2018) in three ways. First, we provide a formula for the GPS correction that can be easily implemented by practitioners. Second, we test the GPS adjustment on long-only minimum variance portfolios, which are closer to practical investments than the long-short minimum variance portfolios considered previously. Finally, we benchmark the impact of the GPS adjustment on optimized portfolio to the impact of the Blume and Vasicek adjustments along with the the widely-used covariance matrix shrinkage adjustments in Ledoit & Wolf (2004*a*) and Ledoit & Wolf (2017).[8]

2. Dispersion bias in PCA betas

Principal component analysis of a sample covariance matrix of stocks, denoted $\mathbf{S}$, provides estimates of the PCA factors that drive equity return and risk. Of special interest are the PCA betas (nor-

[3] Considerations of the assumptions underlying the CAPM as well as empirical assessments of its predictions indicate that the CAPM is misspecified. A few of the many studies that take issue with the CAPM are Blume & Friend (1973) Fama & French (2004) and Markowitz (2005).

[4] The use of PCA to identify financial factors dates back, at least, to Chamberlain & Rothschild (1983), Connor & Korajczyk (1986), Connor & Korajczyk (1988).

[5] The market-like qualities of the dominant PCA factor in sample covariance matrices of public equity returns amount to to an empirical argument for the existence of a large common factor. The assertion that the best instance of this factor is the market, itself, requires additional argument.

[6] The emphasis in Goldberg et al. (2018) on the role estimated betas in the portfolio construction process distinguishes that paper from empirical studies of the Blume and Vasicek adjustments, which focus on the accuracy of next-period beta prediction. See, for example, Elton, Gruber & Urich (1978), Mantripragada (1980) and Gray, Hall, Diamomd & Brooks (2013).

[7] The literature on how estimation error corrupts optimized portfolios includes Klein & Bawa (1976), Brown (1979), Jobson & Korkie (1980), Britten-Jones (1999), Bender, Lee, Stefek & Yao (2009) and Bianchi, Goldberg & Rosenberg (2017) and references therein.

[8] The long history of applying shrinkage methods to improve portfolio construction includes Jorion (1986).

malized to have mean one), which are the exposures to the dominant factor. It is an empirical fact that estimated PCA betas are typically positive. Combined with the fact that PCA betas explain a disproportionate fraction of the risk of most portfolios, PCA betas determine a factor that is market-like. This factor is not, however, constrained to be the market.

We denote the vector of true PCA betas by β. For ease of comparison to the market betas we normalize the β to have unit[9] mean, i.e., $\mu_\beta = 1$.

$$\mu_\beta = \frac{1}{N}\sum_{n=1}^{N}\beta_n \tag{4}$$

Typically, PCA betas are dispersed about one, and we define the dispersion of the PCA betas, τ_β, to be

$$\tau_\beta \triangleq \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\beta_n - \mu_\beta)^2}, \tag{5}$$

the standard deviation of the betas about their mean. In practice, knowledge of the true PCA betas would require knowledge of the true covariance matrix $\mathbf{\Sigma}$, which we do not observe. Instead, we extract *raw betas* from the sample covariance matrix $\mathbf{S}$. We normalize the raw betas to have mean one and denote them by $\hat{\beta}^{\text{raw}}$. The dispersion of the raw betas, $\tau_{\hat{\beta}^{\text{raw}}}$, is obtained by replacing the β with $\hat{\beta}^{\text{raw}}$ in formula (5). It is an artifact of most statistical methods that estimated betas tend to be overly dispersed. In other words,

$$\tau^2_{\hat{\beta}^{\text{raw}}} = \tau^2_\beta + \mathscr{D}^2_{\text{bias}}$$

for some nonnegative dispersion bias $\mathscr{D}^2_{\text{bias}}$. PCA betas follow suit and for them, $\mathscr{D}^2_{\text{bias}} \geq 0$ with high probability for moderate and large N. Accordingly, we define the dispersion bias to be

$$\mathscr{D}^2_{\text{bias}} \triangleq \tau^2_{\hat{\beta}^{\text{raw}}} - \tau^2_\beta. \tag{6}$$

For PCA betas, this bias is, to a large extent, correctable. The correction we offer, which we call the *GPS adjustment*, substantially mitigates the effect of estimation error embedded in the $\hat{\beta}^{\text{raw}}$. This results in significantly greater accuracy in the estimated weights and risk forecasts of minimum variance portfolios, which are victims of the error-maximizing tendencies of optimizers.

3. Adjusting raw betas

Motivated by the fact[10] the dispersion bias $\mathscr{D}^2_{\text{bias}}$ is typically greater than zero, we wish to correct the raw betas so as to decrease their dispersion. It is not difficult to show that the beta adjustment recipe in (1) accomplishes just that since,

$$\tau^2_{\hat{\beta}^{\text{adj}}} = c^2\tau^2_{\hat{\beta}^{\text{raw}}} < \tau^2_{\hat{\beta}^{\text{raw}}}. \tag{7}$$

Note, the mean of the adjusted betas remains one. In what follows, we take $\hat{\beta}^{\text{raw}}$ as the vector of PCA betas extacted from the sample covariance $\mathbf{S}$ and normalized to have, $\mu_{\hat{\beta}^{\text{raw}}} = 1$. The estimated covariance $\hat{\mathbf{\Sigma}}$ is assembled from $\hat{\beta}^{\text{raw}}$, the other statistical factors and the specific risk estimates. For a full recipe of this PCA procedure, see Appendix A.

3.1. Blume's adjustment. As discussed in the introduction, Blume's method refers to taking the adjustment weight c as $c^{\text{Blume}} = 2/3$ in formula (1) to obtain the following adjusted betas ($\hat{\beta}^{\text{adj}} = \hat{\beta}^{\text{Blume}}$).

$$\hat{\beta}^{\text{Blume}} = \frac{2}{3}\hat{\beta}^{\text{raw}} + \frac{1}{3} \tag{8}$$

While here, $\hat{\beta}^{\text{raw}}$ denotes specifically the PCA betas, Blume's adjustment is "universal." It is not tailored to any particular method. This has some justification as most statistical methods produce estimates of beta that are overly dispersed (i.e. $\mathscr{D}^2_{\text{bias}} \geq 0$) But, its shortcoming is that it is not adaptive. Blume's adjustment shrinks the dispersion of the raw betas by a constant factor of $4/9$ths. Its correction towards one ("*the grand mean of all betas*") can overshoot or undershoot the goal depending upon the realized scenario (the $\hat{\beta}^{\text{raw}}$).

The estimated returns covariance matrix $\hat{\mathbf{\Sigma}}$ is assembled in a fashion identical to PCA (see Appendix A) but with $\hat{\beta}^{\text{Blume}}$ now replacing $\hat{\beta}^{\text{raw}}$.

[9]The (equally weighted) mean (4) can be also be cap-weighted, or weighted according to another heuristic; see Appendix B for a weighting we use for our method.

[10]This is proved by Goldberg et al. (2018) for PCA (statistical PCA betas) under some assumptions. Therein, the role of the dispersion τ_β is captured by $\gamma_{\beta,z}$ (the projection of β onto $z = (1,\ldots,1)/\sqrt{N}$). The two are related by $\gamma^2_{\beta,z} = \frac{1}{1+\tau^2_\beta}$.

3.2. The GPS adjustment. Goldberg et al. (2018) devise a statistical correction (GPS) in the spirit of Vasicek (see (3)) but adapted for PCA. One difference between the GPS and Vasicek adjustments is that the former corrects each raw beta separately, depending upon the standard error, $s_{\hat{\beta}^{\text{raw}}}$, for that particular security. In the GPS adjustment, we apply the same correction to each individual raw beta because we do not observe the (latent) factor returns. In this more general (APT) setting, only the security returns are known. The other deviation from Vasicek's adjustment is that is that an estimate for τ_β may actually be assembled from the observed data. No educated guess for the dispersion τ_β is required.

Here, we briefly discuss two GPS adjustments, deferring the details to Appendix B. The first GPS adjustment (akin to Vasicek's) presumes that the true beta dispersion τ_β is known. Accordingly, we refer to this adjustment as the GPS oracle. We denote its adjustment weight c of formula (1) by c_*^{GPS}. The more practical adjustment does not rely on knowledge of τ_β. We refer to it simply as the GPS adjustment and we denote its weight as c^{GPS}.

The GPS oracle adjustment takes its adjustment weight as

$$c_*^{\text{GPS}} = \frac{\tau_\beta^2}{\tau_\beta^2 + \mathscr{D}_{\text{bias}}^2} \quad \left(= \frac{\tau_\beta^2}{\tau_{\hat{\beta}^{\text{raw}}}^2} \right). \tag{9}$$

Here, Vasicek's squared standard error, $s_{\hat{\beta}^{\text{raw}}}^2$ in (3) is replaced with the dispersion bias, $\mathscr{D}_{\text{bias}}^2$ in (6). This leads to adjusted betas $\hat{\beta}^{\text{adj}}$ of the form

$$\hat{\beta}_*^{\text{GPS}} = \left(\frac{\tau_\beta^2}{\tau_{\hat{\beta}^{\text{raw}}}^2} \right) \hat{\beta}^{\text{raw}} + \left(1 - \frac{\tau_\beta^2}{\tau_{\hat{\beta}^{\text{raw}}}^2} \right), \tag{10}$$

upon substitution of the adjustment weights into formula (1).

When the oracle is not there to reveal the true beta dispersion τ_β, the GPS adjustment works to estimate τ_β from the observed security returns. In particular, under some distibutional assumptions on the data, we can infer a good estimate of θ_{raw}, the angle between the PCA betas β and their estimates $\hat{\beta}^{\text{raw}}$. This estimate is based on the eigenvalues of the sample covariance matrix $\mathbf{S}$ (see Appendix B for detail). Here, for brevity, we only state the adjustment weight c in terms of the angle θ_{raw}.

$$c^{\text{GPS}} = \cos^2(\theta_{\text{raw}}) - \frac{\sin^2(\theta_{\text{raw}})}{\tau_{\hat{\beta}^{\text{raw}}}^2} \tag{11}$$

The adjusted betas, denoted by $\hat{\beta}^{\text{GPS}}$, are formed from the usual formula (1) with c as in (11).

Again, since the GPS method adjustments only the raw PCA betas, the resulting estimated covariance matrix $\hat{\mathbf{\Sigma}}$ is assembled identically to PCA but with either $\hat{\beta}_*^{\text{GPS}}$ or $\hat{\beta}^{\text{GPS}}$ replacing $\hat{\beta}^{\text{raw}}$.

3.3. Ledoit-Wolf adjustments. A popular adjustment described in Ledoit & Wolf (2004*a*) does not modify PCA betas directly as in formula (1). Instead, it directly forms the estimated covariance matrix $\hat{\mathbf{\Sigma}} = z\mathbf{S} + (1-z)\mathbf{C}$ for $\mathbf{C}$, a "*constant correlation*" matrix, and some z between 0 and 1. We refer to this method as LWCC. This adjustment shrinks the extreme entries of the sample covariance $\mathbf{S}$ towards their center since $\mathbf{C}$ is formed form the average of the sample correlations. The covariance $\mathbf{C}$ corresponds to a model in which all pairwise correlations between pairs of security returns are identical. The constant z is selected to approximately minimize the expected error of the estimator $\hat{\mathbf{\Sigma}}$ (see Appendix C.1 for details). While this approach does not adjust the raw betas directly, the PCA beta estimates implied by $\hat{\mathbf{\Sigma}}$ are different from the raw estimate $\hat{\beta}^{\text{raw}}$ based on $\mathbf{S}$. In other words, there is some implied c determined as

> c^{LWCC} – *the value of c that minimizes distance between the adjusted betas $\hat{\beta}^{\text{adj}}$ in* (1) *and the PCA betas of the matrix $\hat{\mathbf{\Sigma}}$ above.*

However, this recipe adjusts more than just the raw betas. For example, it also adjusts the variance of the factors, in particular the market volatility that is associated with the PCA betas.

More recently, Ledoit & Wolf (2017) proposed an estimate $\hat{\mathbf{\Sigma}}$ that adjusts only the PCA factor variances leaving the estimated PCA betas and other factors intact. We include this method in our numerical results (Section 5), since many current state-of-the-art methods focus on adjusting (i.e. shrinking) the factor variance in some way. For this procedure, the shrinkage of the estimated (raw) variances is non-linear. We refer to this method as LWNL, the details of which may be found in Appendix C.2. The important point, in relation to the approaches above, is that the c of formula (1) for this method, denoted by c^{LWNL}, is one. That is the PCA betas are not adjusted in any way.

4. Portfolio construction and accuracy

Beta adjustments affect estimated covariance matrices, which, in turn, affect weights and risk forecasts of optimized portfolios. In the numerical experiments discussed in Section 5, we look at the impact of beta adjustments and other types of covariance matrix adjustments on minimum variance portfolios. In this section, we document our optimization process and evaluation metrics.

4.1. Construction. We consider the long-only global minimum variance portfolio

$$\begin{aligned} &\min_{w \in \mathbb{R}^N} \; w^\top \widehat{\mathbf{\Sigma}} w \\ &w^\top 1_N = 1 \\ &w \geq 0 . \end{aligned} \tag{12}$$

where $\widehat{\mathbf{\Sigma}}$ is a covariance matrix estimated from T observations of returns to N securities and 1_N is the N-vector of all ones. We denote by $\widehat{w}$ the solution to (12), the estimated minimum variance portfolio. It is well-known that the portfolio weights, $\widehat{w}$, are extremely sensitive to errors in the estimated model, and risk forecasts for the optimized portfolio tend to be too low.[11]

4.2. Accuracy. Estimation error causes two types of difficulties in optimized portfolios. It distorts portfolio weights, and it biases the risk of optimized portfolios downward. Both effects are present for the minimum variance portfolio $\widehat{w}$, constructed as the solution to (12) using some estimate $\widehat{\mathbf{\Sigma}}$ of the returns covariance $\mathbf{\Sigma}$. We now define the metrics for assessing the magnitude of these two errors.

We denote by w_* the optimal portfolio, i.e., the solution of (12) with $\widehat{\mathbf{\Sigma}}$ replaced by $\mathbf{\Sigma}$. Since the latter is positive definite, the optimal portfolio weights w_* may be given explicitly. We define,

$$\mathscr{T}^2 = (w_* - \widehat{w})^\top \, \mathbf{\Sigma} \, (w_* - \widehat{w}) \, , \tag{13}$$

the (squared) tracking error of $\widehat{w}$. Here, $\mathscr{T}^2_{\widehat{w}}$ measures the distance between the optimal and estimated portfolios, w_* and $\widehat{w}$. Specifically, it is the square of the width of the distribution of return differences $w_* - \widehat{w}$.

The variance of portfolio $\widehat{w}$ is given by $\widehat{w}^\top \widehat{\mathbf{\Sigma}} \widehat{w}$ and its true variance is $\widehat{w}^\top \mathbf{\Sigma} \widehat{w}$. We define,

$$\mathscr{R} = \frac{\widehat{w}^\top \widehat{\mathbf{\Sigma}} \widehat{w}}{\widehat{w}^\top \mathbf{\Sigma} \widehat{w}} , \tag{14}$$

the variance forecast ratio. Ratio (14) is less than one when the risk of the portfolio $\widehat{w}$ is underforecast.[12]

Metrics (13) and (14) quantify the errors in portfolio weights and risk forecasts induced by estimation error.[13]

Ledoit & Wolf (2017) rely on a third metric, which they all "loss function." It is the true variance of $\widehat{w}$, given by

$$\mathscr{L} = \widehat{w}^\top \mathbf{\Sigma} \widehat{w} . \tag{15}$$

The loss function is the denominator of the variance forecasting ratio, and it can be converted to an out-of-sample metric that can be applied to empirical data.

5. Numerical results

We use tracking error, the variance forecast ratio and the loss function to evaluate the accuracy of estimated long-only minimum variance portfolios composed of securities whose returns follow a 4-factor model. The estimates are generated with 6 methods. Our baseline is PCA. We implement three PCA beta adjustment methods, where the raw PCA betas are corrected by formula (1) with weight $c \neq 1$. The adjustment weights are $c^{\mathrm{Blume}} =$

[11] Extreme sensitivity of portfolio weights to estimation error and the downward bias of risk forecasts are also found in the optimized portfolios constructed by asset managers. Portfolio specific corrections of the dispersion bias discussed in this article are useful in addressing these practical problems. The focus on the global minimum variance portfolio in this article highlights the essential logic of our analysis in a simple setting of practical importance.

[12] With respect to the equally weighted portfolio, w_e, (the tracking error of which is zero) we only consider the variance forecast ratio.

[13] For a relationship to more standard error norms see Wang & Fan (2017).

2/3 for Blume's method, and our proposed (GPS) adjustment weights c_*^{GPS} and c^{GPS} described in Section 3. We also implement the two adjustments by Olivier Ledoit and Michael Wolf (LW) summarized in Section 3. Their first method, which we refer to as LWCC, shrinks the sample covariance matrix to a constant correlation model, and has an implied adjustment weight c^{LWCC} that corresponds to formula (1). Their second method, which we refer to as LWNL, applies a non-linear shrinkage to the raw PCA factor variances but has no PCA beta adjustment (i.e., $c^{\mathrm{LWNL}} = 1$).

5.1. Return generating process and covariance matrix. We simulate returns to $N \in \mathbb{N}$ securities that follow a K-factor model with heterogeneous specific risk,

$$R = B\phi + \epsilon, \tag{16}$$

where ϕ is the K-vector of factor returns, B is the matrix of factor exposures, and $\epsilon = (\epsilon^1, \epsilon^2, \ldots, \epsilon^N)$ is the vector of diversifiable specific returns. While the returns $(\phi, \epsilon) \in \mathbb{R}^K \times \mathbb{R}^N$ are random, we treat each exposure $\beta_{n,k} \in \mathbb{R}$ as a constant to be estimated. Assuming ϕ and the $\{\epsilon_n\}$ are mean zero and pairwise uncorrelated, the $N \times N$ covariance matrix of R can be expressed as

$$\mathbf{\Sigma} = BFB^\top + \mathbf{\Delta}\,. \tag{17}$$

Here, F is the covariance matrix of ϕ and $\mathbf{\Delta}$ is a diagonal matrix with nth entry $\mathbf{\Delta}_{nn} = \delta_n^2$, the variance of ϵ_n.

5.2. Calibration. Our numerical examples are based on a $K = 4$-factors instance of (16). The first factor is the market, so the first column of B is the vector market exposures, which we denote by β. We construct the vector β by drawing its entries from a normal distribution and standardize the vector to have mean equal to 1 and standard deviation equal to τ, the dispersion of the "market beta."

The three remaining factors are fashioned from equity styles such as volatility, earnings yield and size.[14] We draw the exposure of each security from a normal distribution and standarize to a Z-score with mean 0, variance 1.

We tune our model to calm and stressed regimes, and our empirical motivation comes from a number of sources. Clarke, De Silva & Thorley (2011) and Goldberg, Leshem & Geddes (2014) show levels of the volatility of minimum variance portfolios in the US through history. Green & Hollifield (1992) discuss the presence of a market-like factor, one that has high volatility and positive exposures, and its impact on quantitative portfolio construction. Sefton, Jessop, De Rossi, Jones & Zhang (2011) show that market betas tend to be more concentrated in calmer markets. Morozov, Wang, Borda & Menchero (2012, Table 4.3) and Bayraktar, Mashtaler, Meng & Radchenko (2014) provide guidance on the volatility of non-market factors. And as everyone knows, security correlations tend toward 1 in a crisis.

The regime-specific parameters we use in our simulation are in Table 1. The parameters contribute to outcomes in interesting ways. All else equal, an increase in market risk increases pairwise correlations and drives up risk in the long-only minimum variance portfolio. It has little effect on the global minimum variance portfolio, whose risk is largely driven by the specific risk, since the portfolio hedges away market risk. The increase in dispersion of betas drives down pairwise correlations, and thus increases the capacity to use short positions to hedge market risk. In the case of the global minimum variance portfolio, shorting frictions diminish the feasibility of constructing the portfolio in practice. Finally, the increase in specific risk drives up the risk of the global and long-only minimum variance portfolios. In aggregate, these characteristics represent a justified facsimile of the primary changes that occur between calm and stressed markets. The model parameters we use in our simulations are in Table 1.

In each experiment, we simulate a year's worth of independent daily returns, $T = 250$, to $N = 500$ securities. From this data set, we construct a sample covariance matrix, $\mathbf{S}$, from which we construct estimators $\widehat{\mathbf{\Sigma}}$.

5.3. Results. We begin by looking at *adjustment weights c* that the tested methods apply to the raw PCA betas via formula (1). There is no adjustment

[14] References on the importance of volatility (or beta), earnings yield and size in explaining cross-sectional correlations are Black, Jensen & Scholes (1972), Basu (1983) and Banz (1981).

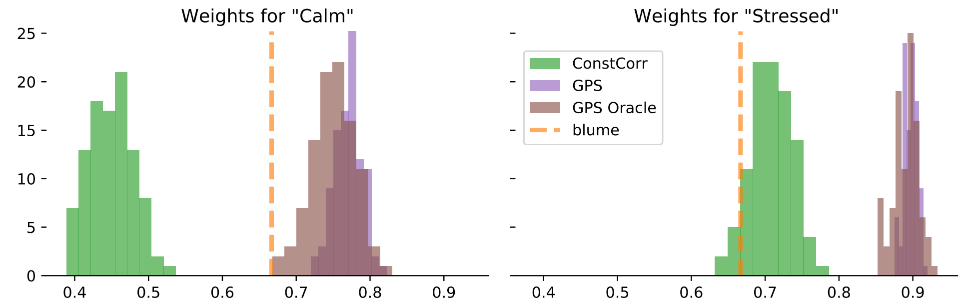

Figure 1: A comparison of the Blume shrinkage weights, the implied Constant Correlation shrinkage weight, and the weights derived from oracle and data-driven GPS models. Each histogram is based on $N = 500$ securities, $T = 250$ observations and P = 100 simulation paths.

Parameter	Calm	Stressed
market volatility	16%	32%
style volatility	[4%, 8%]	[4%, 8%]
specific volatility	[32%, 64%]	[48%, 96%]
beta dispersion	0.33	0.40

Table 1: Simulation parameters for two regimes.

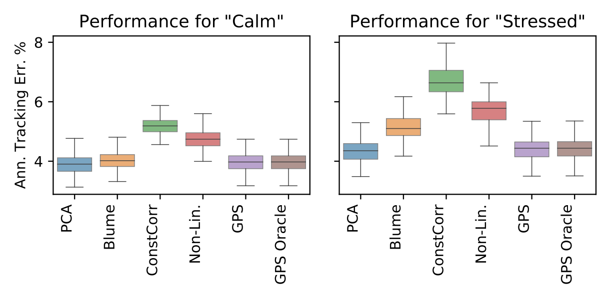

Figure 3: Tracking error for the long-only minimum variance portfolio with $N = 500$ securities estimated from $T = 250$ observations. Each boxplot is based on P = 100 simulation paths.

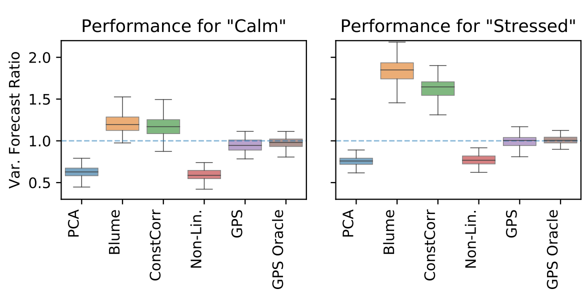

Figure 2: Variance forecast ratios for the long-only minimum variance portfolio with $N = 500$ securities estimated from $T = 250$ observations. Each boxplot is based on P = 100 simulation paths.

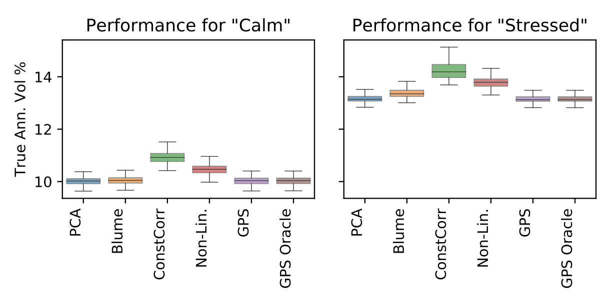

Figure 4: Out-of-sample volatility for the long-only minimum variance portfolio with $N = 500$ securities estimated from $T = 250$ observations. Each boxplot is based on P = 100 simulation paths.

for PCA nor for LWNL (both have $c = 1$). Figure 1 shows the histograms of the adjustment weights for four of the six methods that actually adjusts the raw PCA betas. The left and right panels show the adjustments for the calm and stressed regimes corresponding to Table 1. The Blume and LWCC methods put far less weight on raw betas than the GPS adjustments and so are more conservative relative to the optimal correction. The histograms of the GPS oracle and (data-driven) GPS weights were relatively similar to one another. This demonstrates that the true value of beta dispersion is not required to make a near optimal correction. The dynamic methods placed more weight on raw PCA betas in a calm regime than in a stressed regime. This is a consequence of the fact that true betas are more dispersed in the stress scenario leaving less dispersion bias to correct for.

We turn next to the performance performance metrics. Figure 2 reports the accuracy of variance forecasts for estimated long-only minimum variance portfolios. In both calm and stressed regimes, the GPS models gave the most accurate risk forecasts. PCA and the non-linear models underforecast, while Blume and the Constant Correlation models overforecast.

Figure 3 reports the accuracy of weights for estimated long-only minimum variance portfolios. The PCA and GPS models had the most accurate portfolios, as measured by tracking error, in both regimes. The accuracy of the Blume model was on par with PCA and GPS in the calm regime, but not in the stressed regime.

Figure 4 reports out-of-sample (true) volatility for estimated long-only minimum variance portfolios. The results for this metric are qualitatively simililar to the tracking error results.

6. Summary

The importance of a market-like factor driving developed market equity returns has been understood since the 1960s. The impact of factor estimation error on quantitative portfolio construction and risk forecasting, however, remain obscure.

In this article, we provide simple, implementable formulas for the GPS adjustment, an optimal method of shrinking estimated PCA betas. The GPS adjustment is analogous to the well-known Blume and Vasicek adjustments of historical betas.

The power of the GPS adjustment is illustrated in simulations tuned to calm and stressed markets. We consider performance metrics that measure the accuracy of portfolio weights and risk forecasts for long-only minimum variance portfolios, which are lightening rods for estimation error. Our numerical experiments compare GPS-adjusted PCA models to vanilla and Blume-adjusted PCA models, as well as two Ledoit and Wolf adjusted models that directly shrink an estimated covariance matrix. The efficacy of our easy-to-implement adjustment is a step toward making PCA-based risk models useful to financial practitioners.

A. Principal component analysis

PCA takes the statistical factors as the eigenvectors of the sample covariance matrix $\mathbf{S}$ and their variances as the eigenvalues. Specific returns are residuals to regression of returns on the factors. Precisely, given a $N \times T$ matrix of returns $\mathbf{R}$, PCA entails a spectral decomposition of the sample covariance matrix $\mathbf{S} = \mathbf{R}\mathbf{R}^\top/T$, given by

$$\mathbf{S} = \sum_{k=1}^{T} \ell_k^2 b_k b_k^\top . \tag{18}$$

where ℓ_k^2 is the kth largest eigenvalue of $\mathbf{S}$ and b_k is the corresponding eigenvector (exposures of securities to the kth factor). The PCA estimate of the dominant factor β is taken as

$$\hat{\beta}^{\text{raw}} = \frac{b_1}{\mu_{b_1}} \tag{19}$$

where μ_{b_1} is the mean of b_1, ensuring that $\hat{\beta}^{\text{raw}}$ has unit mean The remaining factors are identified as $b_2, \ldots, b_K$. We may then compute the estimates of the specific returns as

$$\mathbf{E} = \mathbf{R} - \sum_{k=1}^{K} \mathbf{R}\, b_k b_k^\top \tag{20}$$

where entry $\mathbf{E}_{ij}$ is the estimate of the specific return to security i in the jth observation. Consequently,

$$\hat{\mathbf{\Delta}} = \mathbf{diag}(\mathbf{E}\mathbf{E}^\top)/T \tag{21}$$

This may be turned into an iterative procedure by applying PCA to $\mathbf{S} - \widehat{\boldsymbol{\Delta}}_m$ where $\widehat{\boldsymbol{\Delta}}_0$ is the zero matrix (as above), then with $\widehat{\boldsymbol{\Delta}}_1$ as in (21) and so on. The iterations are carried out till convergence and improve the estimates in the presence of heteroskedasticity.

B. Full GPS recipe

In order to deal with heteroskedasticity we require an adjustment to our definition of dispersion. A security with a large specific risk, thus a large δ_n^2, is less likely to be included in a variance optimized portfolio. While we have no knowledge of the matrix $\boldsymbol{\Delta}$ of the $\{\delta_n^2\}$, the diagonal component of the true covariance matrix (17), it turns out that the main source of error in the minimum variance portfolio it suffices to use only the estimates $\{\hat{\delta}_n^2\}$. We define the (generalized) dispersion squared as

$$(22) \qquad \tau_\beta^2 = \frac{1}{N_\delta} \sum_{n=1}^{N} \frac{(\beta_n - \mu_\beta)^2}{\hat{\delta}_n^2}$$

where $N_\delta = \sum_{n=1}^{N} 1/\hat{\delta}_n^2$ and the variablae

$$(23) \qquad \mu_\beta = \frac{1}{N_\delta} \sum_{n=1}^{N} \beta_n / \hat{\delta}_n^2$$

is the risk weighted mean of the true PCA betas. The estimates $\{\hat{\delta}_n^2\}$ above are obtained by regression on the estimated PCA betas as in Appendix A.

Since (22) is in still an unknown (since the true PCA betas are unknown) it may only be used as part of the GPS oracle. For this we use the same adjustment weight c as in (9), that is

$$(24) \qquad c_*^{\mathrm{GPS}} = \left(\frac{\tau_\beta^2}{\tau_{\hat{\beta}^{\mathrm{raw}}}^2} \right)$$

where τ_β^2 is as in (22) and $\tau_{\hat{\beta}^{\mathrm{raw}}}^2$ takes (22) but with $\hat{\beta}^{\mathrm{raw}}$ (the estimated PCA betas) replacing the true PCA betas, β.

To obtain the (data-driven) GPS adjustment weight c^{GPS} we must estimated the (generalized) beta dispersion from the returns data. To this end, define the weighted dot product

$$(25) \qquad \hat{\pi}_{\beta,\hat{\beta}} = \frac{1}{N_\delta} \sum_{n=1}^{N} \left(\frac{\beta_n \hat{\beta}_n}{\hat{\delta}_n^2} \right)$$

which reduces to $\cos(\theta_{\mathrm{raw}})$ introduced in Section 3 when the specific risk matrix $\boldsymbol{\Delta}$ is not too sketeroskedastic. We can estimate this by the recipe,

$$(26) \qquad \hat{\pi}_{\beta,\hat{\beta}} \approx 1 - (\bar{\delta}/\ell_1)^2 (N/T)\,,$$

where the pair $(\ell_1^2, \bar{\delta}^2)$ is defined by

ℓ_k^2 – the kth largest eigenvalue of the scaled covariance matrix $\widehat{\boldsymbol{\Delta}}^{-1/2}\mathbf{S}\widehat{\boldsymbol{\Delta}}^{-1/2}$.

$\bar{\delta}^2$ – the estimate of the average specific risk which is computed by the formula,[15]

$$(27) \qquad \bar{\delta}^2 = \frac{\sum_{k>K}^{N} \ell_k^2}{N - K(N/T + 1)}$$

for K denoting the total number of statistical factors extracted.

Finally we take

$$(28) \qquad c^{\mathrm{GPS}} = \hat{\pi}_{\beta,\hat{\beta}}^2 - \frac{1 - \hat{\pi}_{\beta,\hat{\beta}}^2}{\tau_{\hat{\beta}^{\mathrm{raw}}}^2}$$

in formula (1).

C. Ledoit-Wolf shrinkage

C.1. Constant correlation shrinkage. Since the sample covariance matrix $\mathbf{S}$ tends to have large entries that amplify the estimation errors, (Ledoit & Wolf 2004*b*) propose a shrinkage of these extremes toward their center. This is accomplished by taking a convex combination of $\mathbf{S}$ and a covariance model for which pairwise correlations are identical. Taking $z \in (0,1)$, this estimator takes the form

$$(29) \qquad \widehat{\boldsymbol{\Sigma}} = z\mathbf{S} + (1-z)\mathbf{C}$$

where $\mathbf{C}_{ii} = \mathbf{S}_{ii}$ and off the diagonal ($i \neq j$),

$$(30) \qquad \mathbf{C}_{ij} = \frac{2}{(N-1)N} \sum_{l=1}^{N-1} \sum_{w=l+1}^{N} \frac{\mathbf{S}_{lw}}{\sqrt{\mathbf{S}_{ll}\mathbf{S}_{ww}}}.$$

The parameter z is selected to approximately minimize the estimator expected error, $\mathbf{E}\,|\widehat{\boldsymbol{\Sigma}}^{\mathrm{adj}} - \boldsymbol{\Sigma}|^2$. A

[15] The expression $\sum_{k>K}^{N} \ell_k^2$ is more efficiently computed by the trace of $\mathbf{S}$ minus $\sum_{k=1}^{K} \ell_k^2$.

formula for the approximate minimizer $\hat{z}$ may be found in Ledoit & Wolf (2004*a*, Appendix B). A theoretical treatment is in Ledoit & Wolf (2004*b*).

We compute the the implied weight c^{LWCC} of Section 3 for Figure 1 by projecting the new top factor component of shrinkage estimator onto the line between the vector of PCA-betas and the dispersionless vector of all ones, $(1, \ldots, 1)$.

C.2. Non-linear eigenvalue shrinkage. Recently, Ledoit & Wolf (2017) proposed an estimator that shrinks the eigenvalues of $\mathbf{S}$ toward their center. However, each eigenvalue is adjusted individually by a non-linear transformation.

Writing $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ for the eigenvalue decomposition of the sample covariance $\mathbf{S}$ the non-linear shrinkage estimate of $\boldsymbol{\Sigma}$ takes the form

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}} &= \mathbf{U}\mathbf{Q}\mathbf{U}^\top \\ \mathbf{Q}_{ij} &= \begin{cases} \frac{\boldsymbol{\Lambda}_{ii}^{-1}}{\hat{s}^2(\boldsymbol{\Lambda}_{ii})} & \boldsymbol{\Lambda}_{ii} > 0, \\ \frac{(N/T-1)^{-1}}{\hat{s}(\boldsymbol{\Lambda}_{ii})} & \text{otherwise.} \end{cases} \end{aligned} \tag{31}$$

The function $\hat{s}$ is an empirical approximation to the Stietjes transform of the eigenvalues. It may be computed via an implicit formula supplied in Ledoit & Wolf (2017, Appendix C). Theoretically, the estimator $\widehat{\boldsymbol{\Sigma}}$ aims to minimize the loss function (15) in a certain asymptotic regime.

References

Banz, R. W. (1981), ‘The relationship between return and market value of common stock’, *Journal of Financial Economics* **9**, 3–18.

Basu, S. (1983), ‘The relationship between earnings’ yield, market value and return for nyse common stocks’, *Journal of Financial Economics* **12**, 129–156.

Bayraktar, M. K., Mashtaler, I., Meng, N. & Radchenko, S. (2014), Barra us total market equity model for long-term investors: Empirical notes.

Bender, J., Lee, J.-H., Stefek, D. & Yao, J. (2009), ‘Forecast risk bias in optimized portfolios’.

Bianchi, S. W., Goldberg, L. R. & Rosenberg, A. (2017), ‘The impact of estimation error on latent factor models of portfolio risk’, *The Journal of Portfolio Management* **43**(5), 145–156.

Black, F., Jensen, M. C. & Scholes, M. (1972), The capital asset pricing model: some empirical tests, *in* M. C. Jensen, ed., ‘Studies in the Theory of Capital Markets’, Praeger.

Blume, M. E. (1971), ‘On the assessment of risk’, *The Journal of Finance* **26**(1), 1–10.

Blume, M. E. (1975), ‘Betas and their regression tendencies’, *The Journal of Finance* **30**(3), 785–795.

Blume, M. E. & Friend, I. (1973), ‘A new look at the capital asset pricing model’, *The Journal of Finance* **28**(1), 19–33.

Britten-Jones, M. (1999), ‘The sampling error in estimates of mean-variance efficient portfolio weights’, *The Journal of Finance* **54**(2), 655–671.

Brown, S. (1979), ‘The effecct of estimation risk on capital market equilibrium’, *Journal of Financial and Quantitative Analysis* **14**(2), 215–220.

Chamberlain, G. & Rothschild, M. (1983), ‘Arbitrage, factor structure, and mean-variance analysis on large asset markets’, *Econometrica* **51**(5), 1281–1304.

Clarke, R., De Silva, R. & Thorley, S. (2011), ‘Minimum-variance portfolio composition’, *Journal of Portfolio Management* **2**(37), 31–45.

Connor, G. & Korajczyk, R. A. (1986), ‘Performance measurement with the arbitrage pricing theory: A new framework for analysis’, *Journal of financial economics* **15**, 373–394.

Connor, G. & Korajczyk, R. A. (1988), ‘Risk and return in equilibrium apt: Application of a new test methodology’, *Journal of Financial Economics* **21**, 255–289.

Efron, B. (2010), *Large-Scale Inference*, Institute of Mathematical Statistics Monographs, Cambridge University Press.

Elton, E. J., Gruber, M. J. & Urich, T. J. (1978), 'Are betas best?', *The Journal of Finance* **33**(5), 1375–1384.

Fama, E. F. & French, K. R. (2004), 'The capital asset pricing model: theory and evidence', *Journal of Economic Perspectives* **18**(3), 25–46.

Goldberg, L., Leshem, R. & Geddes, P. (2014), 'Restoring value to minimum variance', *Journal of Investment Management* **12**(2), 32–39.

Goldberg, L., Papanicolaou, A. & Shkolnik, A. (2018), The dispersion bias. CDAR working paper.

Gray, S., Hall, J., Diamomd, N. & Brooks, R. (2013), The vasicek adjustment to beta estimates in the capital asset pricing model. SFG Consulting and Monash University.

Green, R. C. & Hollifield, B. (1992), 'When will mean-variance efficient portfolios be well diversified?', *The Journal of Finance* **47**(5), 1785–1809.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *The Journal of Finance* **58**(4), 1651–1683.

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

Jorion, P. (1986), 'Bayes-stein estimation for portfolio analysis', *Journal of Financial and Quantitative Analysis* **21**(3), 279–292.

Klein, R. W. & Bawa, V. S. (1976), 'The effect of estimation risk on optimal portfolio choice', *Journal of Financial Economics* **3**, 215–231.

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

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

Ledoit, O. & Wolf, M. (2017), 'Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks', *The Review of Financial Studies* **30**(12), 4349–4388.

Mantripragada, K. G. (1980), 'Beta adjustment methods', *Journal of Business Research* **8**(3), 329–339.

Markowitz, H. M. (2005), 'Market efficiency: a theoretical distinction and so what?', *Financial Analysts Journal* **61**(5), 17–30.

Morozov, A., Wang, J., Borda, L. & Menchero, J. (2012), The barra global equity model (gem3). MSCI Barra model documentation.

Ross, S. A. (1973), Return, risk and arbitrage, *in* I. Friend & J. Bicksler, eds, 'Return and Risk in Finance', Cambridge: Ballinger, pp. 189–217.

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of economic theory* **13**(3), 341–360.

Sefton, J., Jessop, D., De Rossi, G., Jones, C. & Zhang, H. (2011), Low-risk investing, Technical report, UBS.

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance* **19**(3), 425–442.

Stein, C. et al. (1956), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *in* 'Proceedings of the Third Berkeley symposium on mathematical statistics and probability', Vol. 1, University of California Press, Berkeley, California, pp. 197–206.

Treynor, J. L. (1962), Toward a theory of market value of risky assets. Presented to the MIT Finance Faculty Seminar.

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

Wang, W. & Fan, J. (2017), 'Asymptotics of empirical eigenstructure for high dimensional spiked covariance', *The Annals of Statistics* **45**(3), 1342–1374.