

Multiple Anchor Point Shrinkage for the Sample Covariance Matrix*

Hubeyb Gurdogan[†] and Alec Kercheval[‡]

This version: June 7, 2022
Forthcoming in SIAM J. Financial Math.

Abstract.

Covariance estimation for high-dimensional returns is well-known to be impeded by the lack of long data history. We extend the work of Goldberg, Papanicolaou, and Shkolnik (GPS) [14] on shrinkage estimates for the leading eigenvector of a covariance matrix in the high dimensional, low sample-size regime, which has immediate application to estimating minimum variance portfolios. We introduce a more general framework of eigenvector shrinkage targets – multiple anchor point shrinkage – that allows the practitioner to incorporate additional information, such as rank ordering or sector separation of equity betas, or prior beta estimates from the recent past. We show that certain rank ordering information can be used to define a consistent estimator of the leading eigenvector. We prove some asymptotic statements and illustrate our results with some numerical experiments.

Key words. Covariance matrix estimation, shrinkage, minimum variance portfolio

AMS subject classifications. 91G60, 91G70, 62H25

1. Introduction. This paper is about the problem of estimating covariance matrices for large random vectors, when the data for estimation is a relatively small sample. We discuss a shrinkage approach to reducing the estimation error asymptotically in the high dimensional, bounded sample size regime, denoted HL. We note at the outset that this context differs from that of the more well-known random matrix theory of the asymptotic “HH regime” in which the sample size grows in proportion to the dimension (e.g. [8]). See [19] for earlier discussion of the HL regime, and [9] for a discussion of the estimation problem for factor models in high dimension.

Our interest in the HL asymptotic regime comes from the problem of portfolio optimization in financial markets. There, a portfolio manager is likely to confront a large number of assets, like stocks, in a universe of hundreds or thousands of individual issues. However, typical return periods of days, weeks, or months, combined with the irrelevance of the distant past, mean that the useful length of data time series is usually much shorter than the dimension of the returns vectors being estimated.

In this paper we extend the successful shrinkage approach introduced in [14] (GPS) to a framework that allows the user to incorporate additional information into the shrinkage target and improve results. Our “multiple anchor point shrinkage” (MAPS) approach includes the

*Submitted to the editors September 14, 2021. The authors thank Lisa Goldberg, Alex Shkolnik, and two anonymous referees for many helpful comments. Any errors are our own.

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

[‡]Department of Mathematics, Florida State University, Tallahassee, FL (akercheval@fsu.edu, <http://www.math.fsu.edu/~kercheva/>)

Funding: AK is grateful for a Simons Foundation Collaboration Grant for Mathematicians that partially supported this work.

35 GPS method as a special case.

36 The problem of sampling error for portfolio optimization has been widely studied ever
 37 since Markowitz [25] introduced the approach of mean-variance optimization. That paper
 38 immediately gave rise to the importance of estimating the covariance matrix Σ of asset returns,
 39 as the risk, measured by variance of returns, is given by $w^T \Sigma w$, where w is the vector of weights
 40 defining the portfolio.

41 For a survey of various approaches over the years, see [14] and references therein. Reducing
 42 the number of parameters via factor models has long been standard; see for example [26]
 43 and [27]. The applicability of factor models in a very general HL setting is justified by [3].
 44 Discussion of consistent estimation of factors in the HL and HH regimes is contained in [5]
 45 and [6]. There, the HH regime in which both p and n tend to infinity is required for exact
 46 consistency. In comparison, Theorem 2.3 below attains a consistent estimator of a single factor
 47 in the HL setting for a bounded number of observations.

48 [30] and [12] initiated a Bayesian approach to portfolio estimation and the efficient frontier.
 49 Practitioners are frequently interested in estimating the sensitivity (called “beta”) of asset
 50 returns to the overall market return. Vasicek used a prior cross-sectional distribution for
 51 betas to produce an empirical Bayes estimator for beta that amounts to shrinking the least-
 52 squares estimator toward the prior in an optimal way. This is one of a number of “shrinkage”
 53 approaches in which initial sample estimates of the covariance matrix are “shrunk” toward
 54 a prior e.g. [21], [2], [22], [23], [10]. [24] describes a nonlinear shrinkage estimator of the
 55 covariance matrix focused on correcting the eigenvalues, set in the HH asymptotic regime.
 56 A number of results in the HL and HH regimes related to correcting biases in the spiked
 57 covariance setting of factor models are described in [31].

58 The key insight of [14] was to identify the PCA leading eigenvector of the sample covari-
 59 ance matrix as the primary culprit contributing to sampling error for the minimum variance
 60 portfolio problem in the HL asymptotic regime. Their approach to *eigenvector* shrinkage is
 61 not explicitly Bayesian, but can be viewed in that spirit (see section 2.5). This is the starting
 62 point for the present work.

63 It is worth pointing out that shrinkage approaches to estimation are far broader than
 64 estimating covariance matrices. The books [11] and [16] discusses an array of shrinkage esti-
 65 mators, mainly centered on the famous James-Stein (JS) estimator [20], [7]. The JS estimator
 66 as a prototype is not merely incidental to this work: it turns out that there are close structural
 67 parallels between JS and GPS/MAPS, as described in the recent works [29] and [13].

68 **1.1. Mathematical setting and background.** Next we describe the mathematical setting,
 69 motivation, and results in more detail. We restrict attention to a familiar and well-studied
 70 (e.g. [28]) baseline model for financial returns: the one-factor, “single-index” or “market”,
 71 model

$$72 \quad (1.1) \quad \mathbf{r} = \beta x + \mathbf{z},$$

73 where $\mathbf{r} \in \mathbb{R}^p$ is a p -dimensional random vector of asset (excess) returns in a universe of p
 74 assets, $\beta \in \mathbb{R}^p$ is an unobserved non-zero vector of parameters to be estimated, $x \in \mathbb{R}$ is
 75 an unobserved random variable representing the common factor return, and $\mathbf{z} \in \mathbb{R}^p$ is an
 76 unobserved random vector of residual returns specific to the individual assets.

77 With the assumption that the components of \mathbf{z} are uncorrelated with x and each other, the
 78 returns of different assets are correlated only through β , and therefore the covariance matrix
 79 of \mathbf{r} is

$$80 \quad \Sigma = \sigma^2 \beta \beta^T + \Delta,$$

81 where σ^2 denotes the variance of x , and Δ is the diagonal covariance matrix of \mathbf{z} . Typical
 82 models in practice use multiple drivers of correlation, so this model represents a base case in
 83 which to set our results. However, to the extent that we will measure success below by the
 84 performance of the estimated minimum variance portfolio, to a good approximation only a
 85 single market factor is relevant ([4], [15]).

86 Under the further simplifying model assumption¹ that each component of \mathbf{z} has a common
 87 variance δ^2 (also not observed), we obtain the covariance matrix of returns

$$88 \quad (1.2) \quad \Sigma = \sigma^2 \beta \beta^T + \delta^2 \mathbf{I},$$

89 where \mathbf{I} denotes the $p \times p$ identity matrix.

90 This means that β , or its normalization $b = \beta / \|\beta\|$, is the leading eigenvector of Σ ,
 91 corresponding to the largest eigenvalue $\sigma^2 \|\beta\|^2 + \delta^2$. As estimating b becomes the most
 92 significant part of the estimation problem for Σ , a natural approach is to take as an estimate
 93 the first principal component (leading unit eigenvector) h_{PCA} of the sample covariance of
 94 returns data generated by the model. This principal component analysis (PCA) estimate is
 95 our starting point.

96 Consider the optimization problem

$$97 \quad \min_{w \in \mathbb{R}^p} w^T \Sigma w$$

$$98 \quad e^T w = 1$$

99 where $e = (1, 1, \dots, 1)$, the vector of all ones.

100 The solution, the “minimum variance portfolio”, is the unique fully invested portfolio
 101 minimizing the variance of returns. Of course the true covariance matrix Σ is not observable
 102 and must be estimated from data. Denote an estimate by

$$103 \quad (1.3) \quad \hat{\Sigma} = \hat{\sigma}^2 \hat{\beta} \hat{\beta}^T + \hat{\delta}^2 \mathbf{I}$$

104 corresponding to estimated parameters $\hat{\sigma}$, $\hat{\beta}$, and $\hat{\delta}$.

105 Let \hat{w} denote the solution of the optimization problem

$$106 \quad \min_{w \in \mathbb{R}^p} w^T \hat{\Sigma} w$$

$$107 \quad e^T w = 1.$$

¹The assumption of homogeneous residual variance δ^2 is a mathematical convenience. If the diagonal covariance matrix Δ of residual returns can be reasonably estimated, then the problem can be rescaled as $\Delta^{-1/2} \mathbf{r} = \Delta^{-1/2} \beta x + \Delta^{-1/2} \mathbf{z}$, which has covariance matrix $\sigma^2 \beta_{\Delta} \beta_{\Delta}^T + I$, where $\beta_{\Delta} = \Delta^{-1/2} \beta$.

108 It is interesting to compare the estimated minimum variance

$$109 \quad \hat{V}^2 = \hat{w}^T \hat{\Sigma} \hat{w}$$

110 with the actual variance of \hat{w} :

$$111 \quad V^2 = \hat{w}^T \Sigma \hat{w},$$

112 and consider the variance forecast ratio V^2/\hat{V}^2 as one measure of the error made in the
113 estimation of minimum variance, hence of the covariance matrix Σ .

114 The remarkable fact proved in [14] is that, asymptotically as p tends to infinity with n
115 fixed, the true variance of the estimated portfolio doesn't depend on $\hat{\sigma}$, $\hat{\delta}$, or $\|\hat{\beta}\|$, but only
116 on the unit eigenvector $\hat{\beta}/\|\hat{\beta}\|$. Under some mild assumptions stated later, they show the
117 following.

118 **Definition 1.1.** For a p -vector $\beta = (\beta(1), \dots, \beta(p))$, define the mean $\mu(\beta)$ and dispersion
119 $d^2(\beta)$ of β by

$$120 \quad (1.4) \quad \mu(\beta) = \frac{1}{p} \sum_{i=1}^p \beta(i) \quad \text{and} \quad d^2(\beta) = \frac{1}{p} \sum_{i=1}^p \left(\frac{\beta(i)}{\mu(\beta)} - 1 \right)^2.$$

121 We use the notation for normalized vectors

$$122 \quad b = \frac{\beta}{\|\beta\|}, \quad q = \frac{e}{\sqrt{p}}, \quad \text{and} \quad h = \frac{\hat{\beta}}{\|\hat{\beta}\|}.$$

123

124 **Proposition 1.1 ([14]).** The true variance of the estimated portfolio \hat{w} is given by

$$125 \quad V^2 = \hat{w}^T \Sigma \hat{w} = \sigma^2 \mu^2(\beta) (1 + d^2(\beta)) \mathcal{E}^2(h) + o_p$$

126 where $\mathcal{E}(h)$ is defined by

$$127 \quad \mathcal{E}(h) = \frac{(b, q) - (b, h)(h, q)}{1 - (h, q)^2},$$

128 and where the remainder o_p is such that for some constants c, C , $c/p \leq o_p \leq C/p$ for all p .

129 In addition, the variance forecast ratio V^2/\hat{V}^2 is asymptotically equal to $p\mathcal{E}^2(h)$.

130 Goldberg, Papanicolaou and Shkolnik call the quantity $\mathcal{E}(h)$ the *optimization bias* associated
131 to an estimate h of the true vector b . They note that the optimization bias $\mathcal{E}(h_{PCA})$ is asymp-
132 totically bounded above zero almost surely, and hence the variance forecast ratio explodes as
133 $p \rightarrow \infty$.

134 With this background, the estimation problem becomes focused on finding a better esti-
135 mate h of b from an observed time series of returns. GPS [14] introduces a shrinkage estimate
136 for b – the GPS estimator h_{GPS} – obtained by “shrinking” the PCA eigenvector h_{PCA} along
137 the unit sphere toward q , to reduce excess dispersion. That is, h_{GPS} is obtained by moving a
138 specified distance (computed only from observed data) toward q along the spherical geodesic
139 connecting h_{PCA} and q . “Shrinkage” refers to the reduced geodesic distance to the “shrinkage
140 target” q .

141 The GPS estimator h_{GPS} is a significant improvement on h_{PCA} . First, $\mathcal{E}(h_{GPS})$ tends
 142 to zero with p , and in fact $p\mathcal{E}^2(h_{GPS})/\log\log(p)$ is bounded (proved in [17]). In [14] it
 143 is conjectured, with numerical support, that $E[p\mathcal{E}^2(h_{GPS})]$ is bounded in p , and hence the
 144 expected variance forecast ratio remains bounded. Moreover, asymptotically h_{GPS} is closer
 145 than h_{PCA} to the true value b in the ℓ_2 norm, and it yields a portfolio with better tracking
 146 error against the true minimum variance portfolio.

147 **1.2. Our contributions.** The purpose of this paper is to generalize the GPS estimator by
 148 introducing a way to use additional information about beta to adjust the shrinkage target q
 149 in order to improve the estimate.

150 We can consider the space of all possible shrinkage targets τ as determined by the family
 151 of all nontrivial proper linear subspaces L of \mathbb{R}^p as follows. Given L (assumed not orthogonal
 152 to h), let the unit vector $\tau(L)$ be the normalized orthogonal projection of h onto L . $\tau(L)$ is
 153 then a shrinkage target for h determined by L (and h). We will describe such a subspace L as
 154 the linear span of a set of unit vectors called “anchor points”. In the case of a single anchor
 155 point q , note that $\tau(\text{span}\{q\}) = q$, so this case corresponds to the GPS shrinkage target.

156 The “MAPS” estimator is a shrinkage estimator with a shrinkage target defined by an
 157 arbitrary collection of anchor points, usually including q . When q is the only anchor point,
 158 the MAPS estimator reduces to the GPS estimator. We can therefore think of the MAPS
 159 approach as allowing for the incorporation of additional anchor points when this provides
 160 additional information.

161 In Theorem 2.2, we show that expanding $\text{span}\{q\}$ by adding additional anchor points at
 162 random asymptotically does no harm, but makes no improvement.

163 In Theorem 2.3, we show that if the user has certain mild *a priori* rank ordering infor-
 164 mation about groups of components of β , even with no information about magnitudes, an
 165 appropriately constructed MAPS estimator is a consistent estimator in the sense that it con-
 166 verges exactly to the true vector b in the asymptotic limit, even though the sample size is held
 167 fixed.

168 Theorem 2.4 shows that if the betas have positive serial correlation over recent history, then
 169 adding the prior PCA estimator h as an anchor point improves the ℓ_2 error in comparison
 170 with the GPS estimator, even if the GPS estimator is computed with the same total data
 171 history.

172 The benefit of improving the ℓ_2 error in addition to the optimization bias is that it also al-
 173 lows us to reduce the tracking error of the estimated minimum variance fully invested portfolio,
 174 discussed in Section 3 and Theorem 3.1.

175 In the next sections we present the main results. The framework, assumptions, and state-
 176 ments of the main theorems are presented in Sections 2 and 3. Some simulation experiments
 177 are presented in Section 4 to illustrate the impact of the main results for some specific situ-
 178 ations. Proofs of the theorems of Section 2 are organized in Section 5, followed by Section 6
 179 describing some open questions for further work.

180 To limit the length of this article, the proofs of some of the needed technical propositions
 181 and lemmas appear in a separate document [18], available online. Additional details and
 182 computations may be found in [17].

183 2. Main Theorems.

184 **2.1. Assumptions and Definitions.** We consider a simple random sample history gener-
 185 ated from the basic model (1.1). The sample data can be summarized as

$$186 \quad (2.1) \quad R = \beta X^T + Z$$

187 where $R \in \mathbb{R}^{p \times n}$ holds the observed individual (excess) returns of p assets for a time window
 188 that is set by $n \geq 2$ consecutive observations. We may consider the observables R to be
 189 generated by non-observable random variables $\beta \in \mathbb{R}^p$, $X \in \mathbb{R}^n$ and $Z \in \mathbb{R}^{p \times n}$.

190 The entries of X are the market factor returns for each observation time; the entries
 191 of Z are the specific returns for each asset at each time; the entries of β are the exposure
 192 of each asset to the market factor, and we interpret β as random but fixed at the start of
 193 the observation window of times $1, 2, 3, \dots, n$ and remaining constant throughout the window.
 194 Only R is observable.

195 In this paper we are interested in asymptotic results as p tends to infinity with n fixed.
 196 Therefore we consider equation (2.1) as defining an infinite sequence of models, one for each
 197 p .

198 To specify the relationship between models with different values of p , we need a more
 199 precise notation. We'll let β refer to an infinite sequence $(\beta(1), \beta(2), \dots) \in \mathbb{R}^\infty$, and $\beta^p =$
 200 $(\beta(1), \dots, \beta(p)) \in \mathbb{R}^p$ the vector obtained by truncation after p entries. When the value p is
 201 understood or implied, we will frequently drop the superscript and write β for β^p .

202 Similarly, $Z \in \mathbb{R}^{\infty \times n}$ is a vector of n sequences (the columns), and $Z^p \in \mathbb{R}^{p \times n}$ is obtained
 203 by truncating the sequences at p .

204 With this setup, passing from p to $p + 1$ amounts to simply adding an additional asset to
 205 the model without changing the existing p assets. The p th model is denoted

$$206 \quad R^p = \beta^p X^T + Z^p,$$

207 but for convenience we will often drop the superscript p in our notation when there is no
 208 ambiguity, in favor of equation (2.1).

209 Let $\mu_p(\beta)$ and $d_p(\beta) \geq 0$ denote the mean and dispersion of β^p , given by

$$210 \quad (2.2) \quad \mu_p(\beta) = \frac{1}{p} \sum_{i=1}^p \beta(i) \quad \text{and} \quad d_p(\beta)^2 = \frac{1}{p} \sum_{i=1}^p \left(\frac{\beta(i) - \mu_p(\beta)}{\mu_p(\beta)} \right)^2.$$

211 We make the following assumptions regarding β , X and Z :

- 212 A1. (Regularity of beta) The entries $\beta(i)$ of β are uniformly bounded, independent random
 213 variables, fixed prior to time 1. The mean $\mu_p(\beta)$ and dispersion $d_p(\beta)$ converge to limits
 214 $\mu_\infty(\beta) \in (0, \infty)$ and $d_\infty(\beta) \in (0, \infty)$.
- 215 A2. (Independence of beta, X, Z) β , X and Z are jointly independent.
- 216 A3. (Regularity of X) The entries X_i of X are iid random variables with mean zero, variance
 217 σ^2 .
- 218 A4. (Regularity of Z) The entries Z_{ij} of Z have mean zero, finite variance δ^2 , and uniformly
 219 bounded fourth moment. In addition, the n -dimensional rows of Z are mutually
 220 independent, and within each row the entries are pairwise uncorrelated.²

²Note we do not assume β , X , or Z are Normal or belong to any specific family of distributions.

221 The assumptions above are for the sake of convenience and to simplify the statements
 222 of results, but in practice are non-binding or can be partly relaxed. In assumption A1,
 223 boundedness is automatic in a finite market, and the betas can be viewed as constants as a
 224 special case if desired (until section 2.4). Once β is determined, it is held fixed during the
 225 observation window of length n . In contrast, X and the columns of Z are drawn independently
 226 at each of the n observations times. The existence of the limits $\mu_\infty(\beta)$ and $d_\infty(\beta)$ could be
 227 relaxed by considering the limit superior and inferior of the sequence at the cost of more
 228 complicated theorem statements, so long as $\liminf \mu_p(\beta) \neq 0$, with a change of sign if needed
 229 to make it positive.

230 Assumptions A2 and A3 are conveniences that simplify the analysis and statements of
 231 results. In [14] X and Z are only assumed uncorrelated, so the stronger independence as-
 232 sumption, used in our proofs, is not necessary in all cases. Assumption A4 is one of a few
 233 alternatives that serve the proofs. The fourth moment condition can be dropped in favor
 234 of the additional assumption that the rows of Z are identically distributed, but we prefer
 235 boundedness conditions as they are always satisfied in finite markets.

236 With the given assumptions the covariance matrix Σ_β of R , conditional on β , is

$$237 \quad (2.3) \quad \Sigma_\beta = \sigma^2 \beta \beta^T + \delta^2 I.$$

238 Since β stays constant over the n observations, the sample covariance matrix $\frac{1}{n} R R^T$ converges
 239 to Σ_β almost surely if n is taken to ∞ , and is the maximum likelihood estimator of Σ_β .

240 We will work with normalized vectors on the unit sphere $\mathbb{S}^{p-1} \subset \mathbb{R}^p$. To that end we
 241 define

$$242 \quad (2.4) \quad b = \frac{\beta}{\|\beta\|}, \quad q = \frac{e}{\sqrt{p}},$$

243 where $e = e^p = (1, 1, \dots, 1) \in \mathbb{R}^p$, and $\|\cdot\|$ denotes the usual Euclidean norm.

244 The vector b is the leading eigenvector of Σ_β (corresponding to the largest eigenvalue). We
 245 denote by h the PCA estimator of b , i.e. h is the first principal component, or the unit leading
 246 eigenvector, of the sample covariance matrix $\frac{1}{n} R R^T$. For convenience we always select the
 247 sign of the unit eigenvector h such that the inner product $(h, q) > 0$, ignoring the probability
 248 zero case $(h, q) = 0$.

249 Since β and X appear in the model $R = \beta X + Z$ only as a product, there is a scale
 250 ambiguity that we can resolve by combining their scales into a single parameter η :

$$251 \quad \eta^p = \frac{1}{p} |\beta^p|^2 \sigma^2.$$

252 It is easy to verify that

$$253 \quad \eta^p = \mu_p(\beta)^2 (d_p(\beta)^2 + 1) \sigma^2,$$

254 and therefore by our assumptions η^p tends to a positive, finite limit η^∞ as $p \rightarrow \infty$.

255 Our covariance matrix becomes

$$256 \quad (2.5) \quad \Sigma_\beta \equiv \Sigma_b = p \eta b b^T + \delta^2 I,$$

257 where we drop the superscript p when convenient. The scalars η, δ and the unit vector b are
 258 to be estimated by $\hat{\eta}, \hat{\delta}$, and h . As described above, asymptotically only the estimate h of b
 259 will be significant. Improving this estimate is the main technical goal of this paper.

260 In [14] the PCA estimate h is replaced by an estimate h_{GPS} that is “data driven”, meaning
 261 that it is computable solely from the observed data R . We henceforth use the notation
 262 $h_{GPS} = \hat{h}_q$, for a reason that will be clear shortly. As an intermediate step we also consider a
 263 non-observable “oracle” version h_q , defined as the point on the short \mathbb{S}^{p-1} -geodesic joining h
 264 to q that is closest to b . (Recall that both b and h are chosen to lie in the half-sphere centered
 265 at q .) The oracle version is not data driven because it requires knowledge of the unobserved
 266 vector b that we are trying to estimate, but it is a useful concept in the definition and analysis
 267 of the data driven version. Both the data driven estimate \hat{h}_q and the oracle estimate h_q can be
 268 thought of as obtained from the eigenvector h via “shrinkage” along the geodesic connecting
 269 h to the anchor point, q .

270 The GPS data-driven estimator \hat{h}_q is successful in improving the variance forecast ratio,
 271 and in arriving at a better estimate of the true variance of the minimum variance portfolio.
 272 In this paper we have the additional goal of reducing the ℓ_2 error of the estimator, which, for
 273 example, is helpful in reducing tracking error. To that end, we introduce the following new
 274 data driven estimator, denoted \hat{h}_L .

275 Let $L = L_p \subset \mathbb{R}^p$ denote a nontrivial proper linear subspace of \mathbb{R}^p . If v is any vector in
 276 \mathbb{R}^p , we write

$$277 \quad \text{proj}_L(v)$$

278 for the Euclidean orthogonal projection of v onto L . Denote by k_p the dimension of L_p , with
 279 $1 \leq k_p \leq p - 1$.

280 Let $h = h^p$ denote our normalized leading eigenvector of $\frac{1}{n}R^p(R^p)^T$, s_p^2 its largest eigen-
 281 value, and l_p^2 the average of the remaining non-zero eigenvalues. Then we define the data
 282 driven “MAPS” (Multiple Anchor Point Shrinkage) estimator by

$$283 \quad (2.6) \quad \hat{h}_L = \frac{\tau_p h + \text{proj}_L(h)}{\|\tau_p h + \text{proj}_L(h)\|}$$

284 where

$$285 \quad (2.7) \quad \tau_p = \frac{\psi_p^2 - \|\text{proj}_L(h)\|^2}{1 - \psi_p^2} \quad \text{and} \quad \psi_p = \sqrt{\frac{s_p^2 - l_p^2}{s_p^2}}.$$

286 Here ψ_p measures the relative gap between s_p^2 and l_p^2 . The MAPS estimator can be viewed
 287 as obtained by “shrinking” the PCA estimator h toward the target $\text{proj}_L(h)$ along the sphere

288 \mathbb{S}^{p-1} by a specified amount.

289 Recall that we sometimes use a superscript to emphasize the dimension of a vector, and
 290 the notation (\cdot, \cdot) for the Euclidean inner product of two vectors. The next lemma from [14]
 291 describes the asymptotic limit of ψ_p and inner products (h^p, b^p) , (h^p, q^p) , and (b^p, q^p) as the
 292 dimension p tends to infinity.

293 **Lemma 2.1** ([14]). *The limits $\psi_\infty = \lim_{p \rightarrow \infty} \psi_p$, $(h, b)_\infty = \lim_{p \rightarrow \infty} (h^p, b^p)$, $(h, q)_\infty =$
 294 $\lim_{p \rightarrow \infty} (h^p, q^p)$, and $(b, q)_\infty = \lim_{p \rightarrow \infty} (b^p, q^p)$ exist almost surely. Moreover,*

295
$$\psi_\infty = (h, b)_\infty \in (0, 1),$$

296 *and*

297
$$(h, q)_\infty = (h, b)_\infty (b, q)_\infty \in (0, 1).$$

298 When L is the one-dimensional subspace spanned by the vector q , then \hat{h}_L is precisely the
 299 GPS estimator \hat{h}_q , located along the short spherical geodesic connecting h to q . The phrase
 300 “multiple anchor point” comes from thinking of q as an “anchor point” shrinkage target in
 301 the GPS paper, and L as a subspace spanned one or more anchor points. The new shrinkage
 302 target determined by L is the normalized orthogonal projection of h onto L . When L is the
 303 one-dimensional subspace spanned by q , the normalized projection of h onto L is just q itself.
 304 In the event that L is orthogonal to h , the MAPS estimator \hat{h}_L reverts to h itself.

305 **2.2. The MAPS estimator with random extra anchor points.** Does adding anchor points
 306 to create a MAPS estimator from a higher-dimensional subspace improve the estimation? The
 307 answer depends on whether there is any relevant information about b in the added anchor
 308 points. In the case where there is no added information and we simply add new anchor points
 309 at random, the next theorem says this doesn’t help.

310 First some terminology. We say that L_p is a *random linear subspace* of \mathbb{R}^p if it is non-
 311 trivial, proper, and the span of a collection of random, linearly independent unit vectors. The
 312 random linear subspace H_p is a *uniform random subspace* of \mathbb{R}^p if, in addition, it has spanning
 313 vectors are uniformly distributed on the sphere \mathbb{S}^{p-1} .³ We say L_p is *independent of a random*
 314 *variable* Ψ if it has spanning vectors that are independent of Ψ .

315 **Definition 2.1.** *A non-decreasing sequence $\{k_p\}$ of positive integers is square root domi-*
 316 *nated if*

317
$$\sum_{p=1}^{\infty} \frac{k_p^2}{p^2} < \infty.$$

318 For example, any non-decreasing sequence satisfying $k_p \leq Cp^\alpha$ for some $C > 0$ and $\alpha < 1/2$
 319 is square root dominated. Roughly speaking, a square-root dominated sequence is one that
 320 grows more slowly than \sqrt{p} . In particular, any constant sequence qualifies.

321 **Theorem 2.2.** *Let the assumptions 1, 2, 3 and 4 hold. Suppose, for each p , L_p is a random*
 322 *linear subspace and H_p is a uniform random subspace of \mathbb{R}^p . Suppose also that L_p is inde-*
 323 *pendent of Z , and H_p is independent of both Z and β . Assume also the sequences $\dim L_p$ and*
 324 *$\dim H_p$ are square root dominated.*

325 *Let $L'_p = \text{span}\{L_p, q^p\}$ and $H'_p = \text{span}\{H_p, q^p\}$.*

326 *Then, almost surely,*

327 (2.8)
$$\limsup_{p \rightarrow \infty} \|\hat{h}_{L'} - b\| \leq \lim_{p \rightarrow \infty} \|\hat{h}_q - b\|,$$

³Uniform random subspaces are called Haar random subspaces in [18] because they can be defined alterna-
 tively in terms of the Haar (uniform) measure on the orthogonal group.

328

$$329 \quad (2.9) \quad \lim_{p \rightarrow \infty} \|\hat{h}_{H'} - b\| = \lim_{p \rightarrow \infty} \|\hat{h}_q - b\|,$$

330 *and*

$$331 \quad (2.10) \quad \lim_{p \rightarrow \infty} \|\hat{h}_H - b\| = \lim_{p \rightarrow \infty} \|h - b\|.$$

332 The limits on the right hand sides of (2.8), (2.9), and (2.10) exist by an easy application
 333 of Lemma 2.1. The need for some upper bounds, such as square root domination, for the
 334 dimensions of L and H can be understood by considering the extreme case of maximum
 335 dimension p . In that case, the MAPS estimators all reduce to h itself, so (2.8) and (2.9) fail.

336 Theorem 2.2 says adding random anchor points to form a MAPS estimator does no harm
 337 asymptotically, but also makes no improvement asymptotically. Inequality (2.8) says that
 338 adding anchor points to q that are independent of Z creates a MAPS estimator that is asymp-
 339 totically never worse, in the Euclidean distance, than the GPS estimator \hat{h}_q , though it might
 340 be better (intuitively, if the MAPS estimator incorporates some additional information about
 341 β).

342 Equation (2.9) says that the GPS estimator is asymptotically neither improved nor harmed
 343 by adding extra anchor points uniformly at random when they are independent of β and Z .
 344 Therefore the goal will be to find useful anchor points that take advantage of additional
 345 information about β that might be available. Necessarily those anchor points will not be
 346 independent of β , but can be thought of as creating choices of L'_p to create a strict inequality
 347 in (2.8).

348 Equation (2.10) confirms that the anchor point q used by the GPS estimator has value:
 349 without it, a random selection of anchor points independent of β and Z will define a MAPS
 350 estimator that is asymptotically no better than the PCA estimator h . While q is not random,
 351 it has an implicit relationship to β coming from Assumption A1, which is motivated by the fact
 352 that equity betas are empirically observed to cluster around 1. In this sense, the non-random
 353 anchor point q contains baseline information about β . This is one of the central intuitions
 354 behind the GPS estimator in [14].

355 As a final remark, notice that in Theorem 2.2 we do not require L or H to be independent
 356 of X (but X , Z , and β are mutually independent by Assumption A2). The asymptotic analysis
 357 in the proof requires independence from Z in order to apply a version of the strong law of
 358 large numbers as $p \rightarrow \infty$. In contrast, X does not depend on p and so its contribution can be
 359 controlled *a priori* uniformly in p .

360 **2.3. The MAPS estimator with rank order information about the entries of beta.** We
 361 now wish to consider what kind of information about β could be added in the form of anchor
 362 points to create an improved MAPS estimator.

363 In this section we consider rank order information. Use of estimated rank ordering of
 364 unknown quantities is not new in finance, but has mostly been applied to estimated ordering
 365 of returns rather than betas, such as in [1]. Here we consider order information about betas,
 366 used in connection with shrinkage estimation.

367 It so happens that if a well-informed observer somehow knows the rank-ordering of the
 368 components of β^p for each p – that is, which entry is the largest, which second largest, etc.,

369 then that information alone, without knowing the actual magnitudes, is sufficient to determine
 370 b asymptotically with zero error almost surely, using an appropriate MAPS estimator. The
 371 resulting consistent estimator is unexpected because the asymptotics are not with regard to
 372 sample size n tending to infinity, but rather dimension $p \rightarrow \infty$ with fixed n .

373 In fact, significantly less information than this is needed to create a consistent MAPS
 374 estimator in this sense. It suffices to be able to separate the components of beta into ordered
 375 groups, where the rank ordering of the groups is known, but not the ordering within groups.
 376 The meaning of ordered groups and the constraints on group sizes are explained below.

377 **Definition 2.2.** For any $p \in \mathbb{N}$, let $\mathcal{P} = \mathcal{P}(p)$ be a partition of the index set $\{1, 2, \dots, p\}$ (i.e. a
 378 collection of pairwise disjoint non-empty subsets, called atoms, whose union is $\{1, 2, \dots, p\}$).
 379 The number of atoms of \mathcal{P} is denoted by $|\mathcal{P}|$.

380 We say the sequence of partitions $\mathcal{P}(p)$ is **semi-uniform** if there exists $M > 0$ such that
 381 for all p ,

$$382 \quad (2.11) \quad \max_{I \in \mathcal{P}(p)} |I| \leq M \frac{p}{|\mathcal{P}(p)|}.$$

383 In other words, no atom is larger than a fixed multiple M of the average atom size.

384 Given $\beta \in \mathbb{R}^p$, we say \mathcal{P} is **β -ordered** if, for each distinct $I, J \in \mathcal{P}$, either $\max_{i \in I} \beta_i \leq \min_{j \in J} \beta_j$
 385 or $\max_{j \in J} \beta_j \leq \min_{i \in I} \beta_i$.

386 Intuitively, a semi-uniform β -ordered partition $\mathcal{P}(p)$ defines a way to organize the elements
 387 β_i^p of β^p into disjoint groups (atoms) that are of similar size, and such that for each group,
 388 no element outside the group lies strictly in between two elements of the group.

389 It is easy to see that many such semi-uniform β -ordered partitions always exist, and are
 390 easily constructed if a rank ordering of the betas is known. For example, for each p , first
 391 rank order the elements of β^p , then divide the elements into deciles by taking the largest ten
 392 percent, then the next ten percent, etc., rounding as needed. The result is ten atoms, and
 393 each atom is approximately $p/10$ in size. If in addition we want the number of atoms to
 394 tend to infinity with p , we can replace “ten percent” by a percentage that declines toward
 395 zero as $p \rightarrow \infty$. If instead of ten percent we choose $0 < \alpha < 1/2$ and let the atoms be of
 396 size approximately $p^{1-\alpha}$, there will be approximately p^α atoms in the resulting semi-uniform,
 397 β -ordered partition $\mathcal{P}(p)$, and the sequence $|\mathcal{P}(p)|$ will be square root dominated.

398 Once we have such a partition, each atom $A \subset \{1, 2, \dots, p\}$ defines an anchor point as
 399 follows.

400 **Definition 2.3.** For any $A \subset \{1, 2, \dots, p\}$ let $1_A \in \mathbb{R}^p$ denote the vector defined by the
 401 indicator function of A : $1_A(i) = 1$ if $i \in A$, and otherwise $1_A(i) = 0$. We may then define,
 402 for any partition $\mathcal{P} = \mathcal{P}(p)$, an induced linear subspace $L(\mathcal{P})$ of \mathbb{R}^p by

$$403 \quad (2.12) \quad L(\mathcal{P}) = \text{span}_p\{1_A | A \in \mathcal{P}\} \equiv \langle 1_A | A \in \mathcal{P} \rangle.$$

404 **Theorem 2.3.** Let the assumptions **1, 2, 3** and **4** hold. Consider a semi-uniform sequence
 405 $\{\mathcal{P}(p) : p = 1, 2, 3, \dots\}$ of β -ordered partitions such that the sequence $\{|\mathcal{P}(p)|\}$ tends to infinity
 406 and is square root dominated. Then

$$407 \quad (2.13) \quad \lim_{p \rightarrow \infty} \|\hat{h}_{L(\mathcal{P}(p))} - b\| = 0 \quad \text{almost surely.}$$

408 Theorem 2.3 says that if we have certain prior information about the ordering of the β
 409 elements in the sense of finding an ordered partition (but with no other prior information
 410 about the actual magnitudes of the elements or their ordering within partition atoms), then
 411 asymptotically we can estimate b exactly.

412 Having in hand a true β -ordered partition *a priori* will usually not be possible because
 413 even the ordering of the betas is not likely to be known in practice. However, Theorem 2.3
 414 suggests the hypothesis that partial grouped order information about the betas can still be
 415 helpful in improving our estimate of β .

416 We test this hypothesis in Section 4.2 by considering industry sectors as a proposed way
 417 to form a partition of asset betas. To the extent that betas for equities belonging to the
 418 same sector are similar, and separated from those of other sectors, the partition will be
 419 approximately β -ordered. The experiments of Section 4.2 illustrate, as least in that case, that
 420 these approximations can suffice to create a MAPS estimator that improves on the PCA and
 421 GPS versions.

422 **2.4. A data-driven dynamic MAPS estimator.** Theorem 2.4 of this section shows that
 423 even with no *a priori* information about betas beyond the observed time series of returns, we
 424 can still use the MAPS framework to improve the GPS estimator by making more efficient
 425 use of the data history.

426 In the analysis above we have treated β as a constant throughout the sampling period,
 427 but in reality we expect β to vary slowly over time. To capture this in a simple way, let's
 428 now assume that we have access to returns observations for p assets over a fixed number of
 429 $2n$ periods. The first n periods we call the first (or previous) time block, and the second n
 430 periods the second (or current) time block. We then have returns matrices $R_1, R_2 \in \mathbb{R}^{p \times n}$
 431 corresponding to the two time blocks, and $R = [R_1 R_2] \in \mathbb{R}^{p \times 2n}$ the full returns matrix over
 432 the full set of $2n$ observation times.

433 Define the sample covariance matrices S, S_1, S_2 as $\frac{1}{2n}RR^T$, $\frac{1}{n}R_1R_1^T$, and $\frac{1}{n}R_2R_2^T$, respec-
 434 tively. Let h, h_1, h_2 denote the respective (normalized) leading eigenvectors (PCA estimators)
 435 of S, S_1, S_2 . (Of the two choices of eigenvector, we always select the one having non-negative
 436 inner product with q .)

437 Instead of a single β for the entire observation period, we suppose there are random vectors
 438 β_1 and β_2 that enter the model during the first and second time blocks, respectively, and are
 439 fixed during their respective blocks. We assume both β_1 and β_2 satisfy assumptions (1) and
 440 (2) above, and denote by b_1 and b_2 the corresponding normalized vectors. The vectors β_1 and
 441 β_2 should not be too dissimilar in the mild sense that $(\beta_1, \beta_2) \geq 0$.

Definition 2.4. Define the co-dispersion $d_p(\beta_1, \beta_2)$ and pointwise correlation $\rho_p(\beta_1, \beta_2)$ of
 β_1 and β_2 by

$$d_p(\beta_1, \beta_2) = \frac{1}{p} \sum_{i=1}^p \left(\frac{\beta_1(i)}{\mu_p(\beta_1)} - 1 \right) \left(\frac{\beta_2(i)}{\mu_p(\beta_2)} - 1 \right)$$

and

$$\rho_p(\beta_1, \beta_2) = \frac{d_p(\beta_1, \beta_2)}{d_p(\beta_1)d_p(\beta_2)}.$$

442 The Cauchy-Schwartz inequality shows $-1 \leq \rho_p(\beta_1, \beta_2) \leq 1$. Furthermore, it is straight-

443 forward to verify that

$$444 \quad (2.14) \quad (b_1, b_2) - (b_1, q)(b_2, q) = \frac{d_p(\beta_1, \beta_2)}{\sqrt{1 + d_p(\beta_1)^2} \sqrt{1 + d_p(\beta_2)^2}}.$$

445 and hence $d_p(\beta_1, \beta_2)$, and $\rho_p(\beta_1, \beta_2)$ have limits $d_\infty(\beta_1, \beta_2)$, and $\rho_\infty(\beta_1, \beta_2)$ as $p \rightarrow \infty$.

446 The motivation for this model is our expectation that estimated betas are not fixed, but
 447 nevertheless recent betas still provide some useful information about current betas. To make
 448 this precise in support of the following theorem, we make the following additional assumptions.

449 A5. [Relation between β_1 and β_2] Almost surely, $(\beta_1, \beta_2) > 0$, $\mu_\infty(\beta_1) = \mu_\infty(\beta_2)$, $d_\infty(\beta_1) =$
 450 $d_\infty(\beta_2)$, and $\lim_{p \rightarrow \infty} d_p(\beta_1, \beta_2) = d_\infty(\beta_1, \beta_2)$ exists.

451 **Theorem 2.4.** Assume $\beta_1, \beta_2, R, X, Z$ satisfy assumptions 1-5. Denote by \hat{h}_q^s and \hat{h}_q^d the
 452 GPS estimators for R_2 and R , respectively, i.e. the current (single) and previous plus current
 453 (double) time blocks. Let h_1 and h_2 be the PCA estimators for R_1 and R_2 , respectively.

454 Let $L_p = \langle h_1, q \rangle$ and define a MAPS estimator for the current time block as

$$455 \quad (2.15) \quad \hat{h}_L = \frac{\tau_p h_2 + \text{proj}_L(h_2)}{\|\tau_p h_2 + \text{proj}_L(h_2)\|} \quad \text{where} \quad \tau_p = \frac{\psi_p^2 - \|\text{proj}_L(h_2)\|^2}{1 - \psi_p^2},$$

456 where ψ_p is computed from the eigenvalues of the sample covariance matrix corresponding to
 457 the current time block R_2 . Then, almost surely,

$$458 \quad (2.16) \quad \lim_{p \rightarrow \infty} (\|\hat{h}_L - b_2\| - \|\hat{h}_q^s - b_2\|) \leq 0 \quad \text{and} \quad \lim_{p \rightarrow \infty} (\|\hat{h}_L - b_2\| - \|\hat{h}_q^d - b_2\|) \leq 0,$$

459 and, if $0 < |\rho_\infty(\beta_1, \beta_2)| < 1$,

$$460 \quad (2.17) \quad \lim_{p \rightarrow \infty} (\|\hat{h}_L - b_2\| - \|\hat{h}_q^s - b_2\|) < 0 \quad \text{and} \quad \lim_{p \rightarrow \infty} (\|\hat{h}_L - b_2\| - \|\hat{h}_q^d - b_2\|) < 0.$$

461 Theorem 2.4 says that the MAPS estimator obtained by adding the PCA estimator h from
 462 the previous time block as a second anchor point outperforms the GPS estimator asymptoti-
 463 cally, as measured by ℓ_2 error, whether the latter is estimated with the most recent time block
 464 R_2 or with the full $2n$ (double) data set. This works when the previous time block carries some
 465 information about the current beta (non-zero correlation). In the case of perfect correlation
 466 $\rho_\infty(\beta_1, \beta_2) = 1$ the two betas are equal, and we then return to the GPS setting where beta is
 467 assumed constant across the entire $2n$ observations, so no improved performance is expected.

468 The cost of implementing this “dynamic MAPS” estimator is comparable to that of the
 469 GPS estimator, so should generally be preferred when no rank order information is available
 470 for beta.

471 In this analysis we have chosen to use two historical time blocks of equal length n for the
 472 sake of a definite statement and to illustrate the idea. It is likely that the idea also works
 473 when the time blocks have different lengths, or when there are multiple historical time blocks
 474 in use. Theoretical or experimental analysis could determine rules for making such choices,
 475 but we do not do so here.

476 **2.5. Remarks and connections.** The theorems above illustrate a general theme of the
 477 MAPS framework: the performance of a shrinkage estimator like GPS can be improved when
 478 additional information can be added in the form of additional anchor points. For Theorem 2.3,
 479 that means a certain amount of prior ordering information about the betas can be converted
 480 to anchor points that are good enough to make a bona fide consistent estimator of b . For
 481 Theorem 2.4, the use of a PCA estimator from a prior interval in time as an additional anchor
 482 point improves the estimator if betas are correlated across time. The general point is that
 483 when there is some prior information about the betas that is independent of the time interval
 484 used for the estimation, the investigator should formulate that information as one or more
 485 anchor points and use the MAPS technique.

486 This discussion has close connections to Bayesian decision theory (BDT), which makes
 487 use of a prior distribution of a parameter to be estimated. One could view the addition of an
 488 anchor point in the MAPS framework as an adjustment to a prior distribution for beta.

489 We think it likely that the MAPS approach can be reformulated in BDT terms, although
 490 our results in the current form don't conform to them. We don't formulate the prior informa-
 491 tion in terms of a prior distribution of the parameters. And since our setting is asymptotic as
 492 $p \rightarrow \infty$, our conclusions are almost sure statements, rather than statements about minimizing
 493 posterior expected loss. However, the structural connections between GPS/MAPS and the
 494 James-Stein estimator mentioned in the introduction provides a link. The JS estimator is a
 495 kind of empirical Bayes estimator, for example see [11]. Similarly, the GPS/MAPS estimator
 496 is an empirical version of an "oracle" estimator – see Section 5.

497 Another connection, especially for Theorem 2.4, is to the setting of machine learning.
 498 Although Theorem 2.4 itself is not about machine learning because there is no training process,
 499 one could imagine the use of prior time intervals as input to a training process that finds
 500 optimal anchor points as a function of the prior data. This is likely to improve on our default
 501 use of the PCA leading eigenvector as additional anchor point.

502 **3. Tracking Error.** Our task has been to estimate the covariance matrix of returns for a
 503 large number p of assets but a short time series of n returns observations.

504 Recall that for the returns model (1.1), under the given assumptions, we have the true
 505 covariance matrix

$$506 \quad \Sigma_b = p\eta bb^T + \delta^2 I,$$

507 where η and δ are positive constants and b is a unit p -vector, and we are interested in corre-
 508 sponding estimates $\hat{\eta}$, $\hat{\delta}$, and h that define an estimator

$$509 \quad \Sigma_h = p\hat{\eta} h h^T + \hat{\delta}^2 I.$$

510 Our focus on the estimator h and relative neglect of $\hat{\eta}$ and $\hat{\delta}$ is justified by Proposition
 511 1.1, showing that the true variance of the estimated minimum variance portfolio \hat{w} , and the
 512 variance forecast ratio, are asymptotically controlled by h alone through the optimization bias

$$513 \quad \mathcal{E}(h) = \frac{(b, q) - (b, h)(h, q)}{1 - (h, q)^2}.$$

514 The preceding theorems have focused on a particular measure of estimation error for h :
 515 the ℓ_2 error (Euclidean distance) $\|h - b\| = 2(1 - (h, b))$. By comparison, [14, 15] focus on

516 the variance forecast ratio of the minimum variance portfolio. This error measure has the
 517 benefit of demonstrating improvement of a quantity of direct interest to practitioners, with
 518 the drawback of focusing on a single type of portfolio. The ℓ_2 error is not a familiar financial
 519 quantity, but is an ingredient in the optimization bias above, and also in estimating tracking
 520 error, as we describe next.

521 We turn to a third important measure of covariance estimation quality: the tracking error
 522 for the minimum variance portfolio, which is controlled in part by the ℓ_2 error of h . Tracking
 523 error is a term conventionally used in the finance industry as a measure of the distance between
 524 a portfolio and its benchmark. Here, we adopt the same idea to measure the distance between
 525 an estimated minimum variance portfolio and the true portfolio, as follows.

526 Recall that w denotes the true minimum variance portfolio using Σ , and \hat{w} is the minimum
 527 variance portfolio using the estimated covariance matrix $\hat{\Sigma}$.

528 **Definition 3.1.** *The (true) tracking error $\mathcal{T}(h)$ associated to \hat{w} is defined by*

529 (3.1)
$$\mathcal{T}^2(h) = (\hat{w} - w)^T \Sigma (\hat{w} - w).$$

530 **Definition 3.2.** *Given the notation above, define the eigenvector bias $\mathcal{D}(h)$ associated to a*
 531 *unit leading eigenvector estimate h as*

532
$$\mathcal{D}(h) = \frac{(h, q)^2 (1 - (h, b)^2)}{(1 - (h, q)^2)(1 - (b, q)^2)} = \frac{(h, q)^2 \|h - b\|^2}{\|h - q\|^2 \|b - q\|^2}.$$

533 **Theorem 3.1.** *Let h be an estimator of b such that $\mathcal{E}(h) \rightarrow 0$ as $p \rightarrow \infty$ (such as a GPS or*
 534 *MAPS estimator). Then the tracking error of h is asymptotically (neglecting terms of higher*
 535 *order in $1/p$) given by*

536 (3.2)
$$\mathcal{T}^2(h) = \eta \mathcal{E}^2(h) + \frac{\delta^2}{p} \mathcal{D}(h) + \frac{C}{p} \mathcal{E}(h),$$

537 where

538
$$C = \frac{2}{\xi(1 + d_\infty^2(\beta))} (\delta^2 + \frac{\eta}{\hat{\eta}} \delta^2)$$

539 and $\xi > 0$ is a constant depending only on ψ_∞ , $\mu_\infty(\beta)$, and $d_\infty(\beta)$.

540 We consider what this theorem means for various estimators h . For the PCA estimate, it
 541 was already shown in [14] that $\mathcal{E}(h_{PCA})$ is asymptotically bounded below, and hence so is the
 542 tracking error.

543 On the other hand, $\mathcal{E}(h_{GPS})$ tends to zero as $p \rightarrow \infty$. In addition [14] shows that

544
$$\limsup_{p \rightarrow \infty} p \mathcal{E}^2(h_{GPS}) = \infty$$

545 almost surely, while [17] shows

546
$$\limsup_{p \rightarrow \infty} \frac{p \mathcal{E}^2(h_{GPS})}{\log \log p} < \infty,$$

547 and we conjecture the same is true for the more general estimator h_{MAPS} .

548 This implies the leading terms, asymptotically, are

$$549 \quad \mathcal{T}^2(h_{MAPS}) \leq \eta \mathcal{E}^2(h_{MAPS}) + (\delta^2/p) \mathcal{D}(h_{MAPS})$$

550 Note here the estimated parameters $\hat{\eta}$ and $\hat{\delta}$ have dropped out, with the tracking error
551 asymptotically controlled by the eigenvector estimate h alone.

552 Theorem 3.1 helps justify our interest in the ℓ_2 error results of Theorems 2.3 and 2.4.
553 Reducing the ℓ_2 error $\|h-b\|$ of the h estimate controls the second term $\mathcal{D}(h)$ of the asymptotic
554 estimate for tracking error. We therefore expect to see improved total tracking error when
555 we are able to make an informed choice of additional anchor points in forming the MAPS
556 estimator. This is borne out in our numerical experiments described in Section 4.

557 *Proof of Theorem 3.1*

558 **Lemma 3.2.** *There exists $\xi > 0$, depending only on ψ_∞ , $\mu_\infty(\beta)$, and $d_\infty(\beta)$, such that for*
559 *any p sufficiently large, and any linear subspace L of \mathbb{R}^p that contains q ,*

$$560 \quad \|h_L - q\|^2 > \xi > 0,$$

561 where h_L is the MAPS estimator determined by L .

562 The Lemma follows from the fact that $(h_L, q) \leq (h_{GPS}, q)$, and is proved for the case h_{GPS}
563 using the definitions and the known limits

$$564 \quad (3.3) \quad (h_{PCA}, q)_\infty = (b, q)_\infty (h_{PCA}, b)_\infty$$

$$565 \quad (3.4) \quad (b, q)_\infty^2 = \frac{1}{1 + d_\infty^2(\beta)} \in (0, 1)$$

$$566 \quad (3.5) \quad (h_{PCA}, b)_\infty = \psi_\infty > 0.$$

567 From the Lemma and equation (3.4), we may assume without loss of generality that $\xi > 0$
568 is an asymptotic lower bound for both $\|h_L - q\|^2 = 1 - (h_L, q)^2$ and $\|b - q\|^2 = 1 - (b, q)^2$.

569 Next, we recall it is straightforward to find explicit formulas for the minimum variance
570 portfolios w and \hat{w} :

$$571 \quad (3.6) \quad w = \frac{1}{\sqrt{p}} \frac{\rho q - b}{\rho - (b, q)}, \quad \text{where } \rho = \frac{1 + k^2}{(b, q)}, \quad k^2 = \frac{\delta^2}{p\eta}$$

572 and

$$573 \quad (3.7) \quad \hat{w} = \frac{1}{\sqrt{p}} \frac{\hat{\rho} q - h}{\hat{\rho} - (h, q)}, \quad \text{where } \hat{\rho} = \frac{1 + \hat{k}^2}{(h, q)}, \quad \hat{k}^2 = \frac{\hat{\delta}^2}{p\hat{\eta}}.$$

574 We may use these expressions to obtain an explicit formula for the tracking error:

$$575 \quad \mathcal{T}^2(h) = (\hat{w} - w)^T \Sigma (\hat{w} - w) = (\hat{w} - w)^T (p\eta b b^T + \delta^2 I) (\hat{w} - w)$$

$$576 \quad = p\eta (\hat{w} - w, b)^2 + \delta^2 \|\hat{w} - w\|^2.$$

577 We now estimate the two terms on the right hand side separately.

578 (1) For the first term $p\eta(\hat{w} - w, b)^2$, it is convenient to introduce the notation

$$579 \quad \Gamma = \frac{k^2}{1 + k^2 - (b, q)^2} \text{ and } \hat{\Gamma} = \frac{\hat{k}^2}{1 + \hat{k}^2 - (h, q)^2},$$

580 and since

$$581 \quad \Gamma \leq \frac{k^2}{\xi} \text{ and } \hat{\Gamma} \leq \frac{\hat{k}^2}{\xi}$$

582 both Γ and $\hat{\Gamma}$ are of order $1/p$.

583 A straightforward computation verifies that

$$584 \quad (3.8) \quad (w, b) = \frac{1}{\sqrt{p}}\Gamma(b, q)$$

$$585 \quad (3.9) \quad (\hat{w}, b) = \frac{1}{\sqrt{p}}\left(\mathcal{E}(h) + \hat{\Gamma}[(b, q) - \mathcal{E}(h)]\right).$$

586 We then obtain

$$587 \quad (3.10) \quad p(\hat{w} - w, b)^2 = p[(\hat{w}, b) - (w, b)]^2$$

$$588 \quad (3.11) \quad = \mathcal{E}(h)^2 + 2\mathcal{E}(h)G + G^2,$$

589 where $G = \hat{\Gamma}((b, q) - \mathcal{E}(h)) - \Gamma(b, q)$.

590 Since asymptotically (b, q) is bounded below and $\mathcal{E}(h) \rightarrow 0$, the third term G^2 is of order
591 $1/p^2$ and can be dropped. We thus obtain the asymptotic estimate

$$592 \quad p(\hat{w} - w, b)^2 \leq \mathcal{E}^2 + 2\mathcal{E}(h)(\hat{\Gamma} - \Gamma)(b, q).$$

593 Multiplying by η and using the bounds on $\Gamma, \hat{\Gamma}$ and the limit of (b, q) , we obtain

$$594 \quad p\eta(\hat{w} - w, b)^2 \leq \mathcal{E}^2 + \frac{C}{p}\mathcal{E}(h),$$

595 where C is the constant defined in the statement of the theorem.

596 (2) We now turn to the second term $\|\hat{w} - w\|^2 = \|\hat{w}\|^2 + \|w\|^2 - 2(\hat{w}, w)$.

597 Using the definitions of \hat{w} and w and the fact that k^2, \hat{k}^2 are of order $1/p$, after a calculation
598 we obtain, to lowest order in $1/p$,

$$599 \quad (3.12) \quad p\|\hat{w} - w\|^2 = \frac{(h, q)^2[1 - (h, b)^2]}{(1 - (h, q)^2)(1 - (b, q)^2)} + \frac{1 - (h, q)^2}{1 - (b, q)^2}\mathcal{E}^2(h).$$

600 Since $\mathcal{E}(h) \rightarrow 0$, we may neglect the second term, and putting (1) and (2) together yields

$$601 \quad \mathcal{T}^2(h) \leq \mathcal{E}^2 + \frac{C}{p}\mathcal{E}(h) + \frac{\delta^2}{p}\mathcal{D}(h).$$

602 **4. Simulation Experiments.** To illustrate the previous theorems and test whether the
 603 MAPS estimators can be successful for realistic finite values of p , we present the results of two
 604 numerical experiments. In section 4.1, we draw two correlated random vectors β_1 and β_2 in
 605 \mathbb{R}^p , $p = 500$, with a variable correlation that we control. Returns are generated using β_1 for a
 606 first block of observations, then using β_2 for a second block of equal length. These are used to
 607 test whether the dynamic MAPS estimator of Theorem 2.4 is successful against GPS (which
 608 assumes $\beta_1 = \beta_2$). In addition, since we know the exact ordering of the beta components,
 609 we can compare results with a MAPS estimator defined with a beta-ordered partition as in
 610 Theorem 2.3.

611 In section 4.2, we turn to the use of historical CAPM betas for stocks in the S&P500,
 612 rather than simulated betas. This allows us to test a MAPS estimator defined by a partition
 613 determined by the 11 sectors of the familiar Global Industry Classification Standard of MSCI
 614 and S&P. Under the hypothesis that betas for stocks in the same industry sector tend to
 615 have similar magnitudes, classification by sector represents a potential approximation to a
 616 true (but usually not observable) beta-ordered partition. We test this data-driven MAPS
 617 estimator against PCA, GPS, and the consistent MAPS estimator defined with a true beta-
 618 ordered partition.

619 These simple experiments are only proof-of-concept examples illustrating the potential
 620 for success. We have not attempted the worthwhile project of systematically studying the
 621 possible choices of history length or sector divisions in order to optimize outcomes in real
 622 markets.

623 The Python code used to run these experiments and create the figures is available at
 624 <https://github.com/hugurdog/MAPS.NumericalExperiments>.

625 **4.1. Simulated betas with correlation.** To model the possibility that the true betas may
 626 vary slowly during the time window used for estimation, and as a test for Theorems 2.3 and
 627 2.4, we create a simple two-block simulation model with $p = 500$ stocks in which the true
 628 betas are held constant with value $\beta_1 \in \mathbb{R}^p$ during one block of time, and then shift to a
 629 second but correlated value β_2 for a subsequent block of time.

630 Each block has $n = 25$ observations, so the total observation window is of size $2n = 50$
 631 for each of our $p = 500$ stocks. The $p \times n$ returns matrix for the first block is denoted R_1 and
 632 for the second R_2 , and

$$633 \quad (4.1) \quad R_t = \beta_t X_t + Z_t, \quad t = 1, 2,$$

634 where $X_t \in \mathbb{R}^n$ is a vector of the n unobserved common factor returns in block t , and $Z_t \in \mathbb{R}^{p \times n}$
 635 is the matrix of specific returns in block t .

636 We generate the $p \times n$ matrices R_1 and R_2 from Equation (4.1) by randomly generating
 637 β , X , and Z :

- 638 • the market returns $X_t(j)$, $j = 1, \dots, n$, are an iid random sample drawn from a normal
 639 distribution with mean 0 and variance $\sigma^2 = 0.16$,
- 640 • all components of the asset specific returns $\{Z_t(i, j), i = 1, \dots, p; j = 1, \dots, n\}$ are
 641 i.i.d. normal with mean 0 and variance $\delta^2 = (.5)^2$, and
- the p -vectors β_1 and β_2 are defined by drawing $\beta, \eta \in \mathbb{R}^p$ independently from a Normal

distribution with mean 1 and variance $(.5)^2 I_{p \times p}$, and setting

$$\beta_1 = \beta \text{ and } \beta_2 = \rho\beta + \sqrt{1 - \rho^2}\eta,$$

642 where the correlation ρ ranges through values in $\{0, 0.3, 0.6, 1.0\}$.

643 With this simulated returns data, we compare performance for the following four choices
644 of h :

- 645 1. the PCA estimator on the double block $R = [R_1, R_2]$ (PCA)
- 646 2. the GPS estimator on the double block $R = [R_1, R_2]$ (GPS)
- 647 3. the dynamic MAPS estimator defined on the double block $R = [R_1, R_2]$ by equation
648 (2.15) (Dynamic MAPS)
- 649 4. the MAPS estimator on the single block R_2 incorporating knowledge of a beta ordered
650 partition \mathcal{P} as in Theorem 2.3. The partition is constructed by rank ordering the betas
651 and then grouping them into 7 ordered groups of 71, and a small eighth group of the
652 lowest three. (Beta Ordered MAPS)

653 We report the performance of each of these estimators according to the following two
654 metrics:

- 655 • The ℓ_2 error $\|b - h\|$ between the true normalized beta $b = \frac{\beta}{|\beta|}$ of the current data
656 block R_2 and the estimated unit vector h .
- 657 • The tracking error between the true and estimated minimum variance portfolios w
658 and \hat{w} :

$$659 \quad (4.2) \quad \mathcal{T}^2(\hat{w}) = (\hat{w} - w)^T \Sigma (\hat{w} - w).$$

660 In our double-block context, this tracking error is specified as follows. Σ in (4.2) is the
661 true covariance matrix of the most recent data block R_2 :

$$662 \quad (4.3) \quad \Sigma = \sigma^2 \beta_2 \beta_2^T + \delta^2 I,$$

663 which then also determines the true fully invested minimum variance portfolio w . The esti-
664 mated minimum variance portfolio \hat{w} is determined by the estimated covariance matrix

$$665 \quad (4.4) \quad \hat{\Sigma} = \hat{\sigma}^2 \hat{\beta} \hat{\beta}^T + \hat{\delta}^2 I = (\hat{\sigma}^2 |\hat{\beta}|^2) h h^T + \hat{\delta}^2 I.$$

666 For our comparison, and following the lead of [14], we fix the asymptotically correct values

$$667 \quad (4.5) \quad \hat{\sigma}^2 |\hat{\beta}|^2 = s_p^2 - l_p^2 \text{ and } \hat{\delta}^2 = \frac{n}{p} l_p^2$$

668 (notation as in equation 2.7) across each of the four cases, and vary only the estimator
669 $h = \hat{\beta}/|\hat{\beta}|$ as described above. The motivation for this choice is that in our simulation
670 the parameters σ^2 and δ^2 remain constant across the double time window. Hence the best
671 data-driven estimates for $\hat{\sigma}^2$ and $\hat{\delta}^2$ will be obtained by using s_p^2 and l_p^2 computed from the
672 full double block of data R . This puts all the methods compared on the same footing and
673 isolates h as the sole variable in the experiment.

674 Results of the comparison are displayed below. For each choice of ρ , the experiment was
 675 run 100 times, resulting in 100 ℓ_2 error and tracking error values each. These values are
 676 summarized using standard box-and-whisker plots generated in Python using the package
 677 `matplotlib.pyplot.boxplot`.

678 Figure 1 shows the squared ℓ_2 error $\|h - b\|^2$ for different estimators h (in the same order,
 679 left to right, as listed above) for the cases $\rho = 0, 0.3, 0.6, 1.0$. Throughout the range, the
 680 dynamical MAPS estimator outperforms the other two data-driven estimators, but the beta-
 681 ordered MAPS estimator remains in the lead. The case $\rho = 0$ could be compared to the case
 682 of a Bayesian estimator where the additional anchor point is providing information only about
 683 the distribution of the components of β . As the correlation ρ tends toward one, the GPS and
 684 Dynamic MAPS errors become equal. At $\rho = 1$, $\beta_1 = \beta_2$ and the GPS assumption of constant
 685 β over the $2n$ period is satisfied.

686 Figure 2 displays the scaled tracking error $p\mathcal{T}^2(h)$ outcomes across a range of correlation
 687 values $\rho(\beta_1, \beta_2)$. Dynamic MAPS does best among all data-driven methods, and beta ordered
 688 MAPS is significantly better than all others. As before, the Dynamic MAPS lead disappears
 689 as ρ tends to 1, when $\beta_1 = \beta_2$.

690 **4.2. Simulations with historical betas.** In this section we use historical rather than ran-
 691 domly generated betas to test the quality of MAPS estimators defined using a sector partition
 692 and a beta-ordered partition. We use 24 historical monthly CAPM betas for each of the
 693 $p = 488$ S&P500 firms provided by WRDS⁴ between the dates 01/01/2018 and 11/30/2020.
 694 We denote these betas by $\beta_1, \dots, \beta_{24} \in \mathbb{R}^p$.

695 The WRDS beta suite estimates beta each month from the prior 12 monthly returns.
 696 Therefore in this experiment we set $n = 12$ months, and using these betas simulate 24 different
 697 sets of monthly asset returns $R_t \in \mathbb{R}^{p \times n}$, each for $n = 12$ months.

698 For each $t = 1, \dots, 24$, we generate the returns matrix R_t according to

$$699 \quad (4.6) \quad R_t = \beta_t X_t + Z_t,$$

700 where the unobserved market return $X_t \in \mathbb{R}^n$ and the asset specific return $Z_t \in \mathbb{R}^{p \times n}$ are
 701 generated using the same settings as in the previous section.

702 For each t we also form partitions \mathcal{P}_t^{true} and \mathcal{P}_t^{sector} of the beta indices $\{1, 2, \dots, p\}$. \mathcal{P}_t^{true}
 703 is a true beta-ordered partition with 11 atoms constructed from the true rank ordering of
 704 β_t . \mathcal{P}_t^{sector} is a partition defined by the 11 industry sectors⁵, which we adopt as a possible
 705 data-driven proxy for \mathcal{P}_t^{true} .

706 For each t , we then compute the following four estimators of $b_t = \beta_t / |\beta_t|$:

- 707 1. The PCA estimator. (PCA)
- 708 2. The GPS estimator. (GPS)
- 709 3. The MAPS estimator defined as in Theorem 2.3 using the partition \mathcal{P}_t^{sector} . (Sector
 710 Separated)
- 711 4. The MAPS estimator defined using \mathcal{P}_t^{true} . (Beta Ordered)

⁴Wharton Research Data Services, wrds-www.wharton.upenn.edu

⁵The 11 sectors of the Global Industry Classification Standard are: Information Technology, Health Care, Financials, Consumer Discretionary, Communication Services, Industrials, Consumer Staples, Energy, Utilities, Real Estate, and Materials.

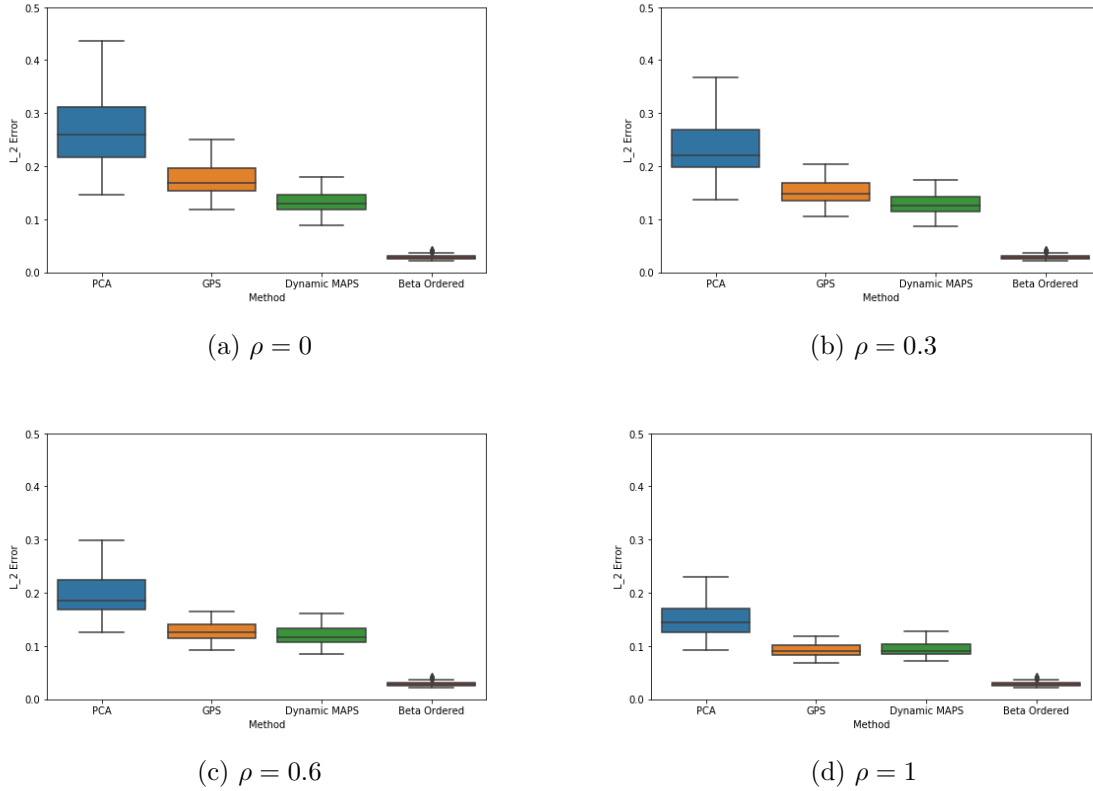


Figure 1: Results of simulation experiments measuring ℓ_2 error for different estimators: PCA, GPS, Dynamic MAPS, and Beta Ordered, and varying correlation ρ between betas in the two different time blocks. When beta correlation between time blocks is low, dynamic MAPS outperforms GPS. The non-empirical beta-ordered MAPS outperforms all others.

712 For each of these four choices of estimator h_t , we examine three different measures of
 713 error: the squared ℓ_2 error $\|h_t - b_t\|^2$, the scaled squared tracking error $p\mathcal{T}^2(h_t)$, and the
 714 scaled optimization bias $p\mathcal{E}_p^2(h_t)$.

715 Since we are interested in expected outcomes, we repeat the above experiment 100 times,
 716 and take the average of the errors as a monte carlo estimate of the expectations

717
$$\mathbb{E}[\|h_t - b_t\|^2], \quad \mathbb{E}[p\mathcal{T}^2(h_t)], \quad \mathbb{E}[p\mathcal{E}_p^2(h_t)],$$

718 once for each t . We then display box plots in Figure 3 for the resulting distribution of 24
 719 expected errors of each type, corresponding to the 24 historical betas. Outcomes are similar
 720 to the simulated beta experiments, where PCA has the poorest performance, Beta Ordered
 721 MAPS the best, and in between are the GPS and empirical MAPS.

722 Using sectors to partition the stocks evidently has some value, as the sector separated
 723 MAPS estimator outperforms GPS by a small but significant amount in both ℓ_2 and tracking

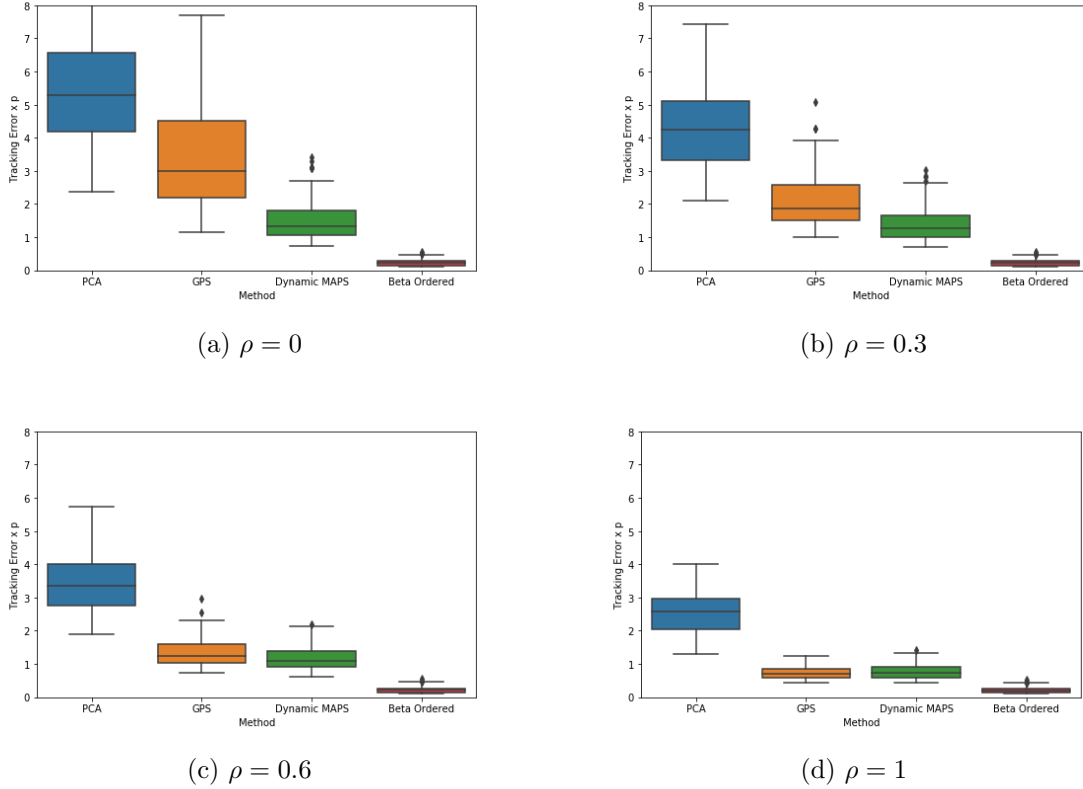


Figure 2: Tracking error results of simulation experiments for different estimators PCA, GPS, Dynamic MAPS, and Beta Ordered. The pointwise correlation ρ is the correlation between betas in the two different time blocks. Results are similar to the ℓ_2 error plots.

724 error. Its success is owed to the tendency for betas of stocks in a common sector to be closer
 725 to each other than to betas in other sectors. The Sector Separated MAPS estimator does not
 726 require any information not easily available to the practitioner, and so represents a costless
 727 improvement on the GPS estimation method.

728 We also note that further experiments are reported in [17] and [18], in which a dynamic
 729 double-block experiment using the historical betas is also carried out, with similar results.

730 **5. Proofs of the Main Theorems.** The proofs of the main theorems proceed by means
 731 of some intermediate results involving an “oracle estimator”, defined in terms of the unob-
 732 servable b but equal to the MAPS estimator in the asymptotic limit (Theorem 5.1 below).
 733 Several technical supporting propositions and lemmas are needed; to save space their proofs
 734 are collected in a separate document, [18], available online.

735 **5.1. Oracle Theorems.** A key tool in the proofs is the *oracle estimator* h_L , which is a
 736 version of \hat{h}_L but defined in terms of b , our estimation target.

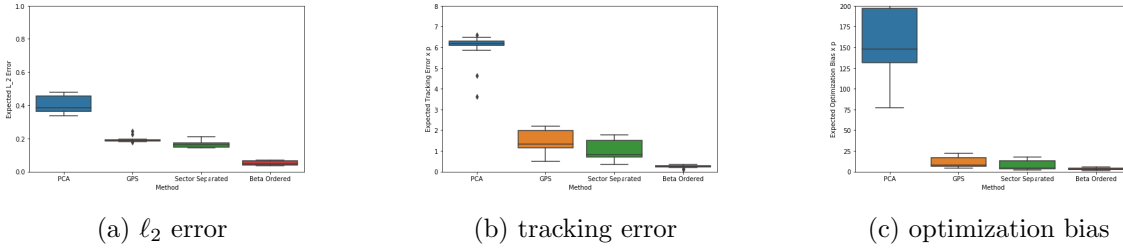


Figure 3: Box plots summarizing the distribution of 24 monte carlo-estimated expected errors for the PCA, GPS, Sector Separated, and Beta Ordered estimators (left to right in each figure). The experiment is conducted over 488 S&P 500 companies. This experiment reveals that the Sector Separated estimator is able to capture some of the ordering information and therefore outperforms the GPS estimator. The Beta Ordered estimator performs best.

737 Given a subspace $L = L_p$ of \mathbb{R}^p , we define

$$738 \quad (5.1) \quad h_L = \frac{\text{proj}(b)}{\frac{\langle h, L \rangle}{\|\text{proj}(b)\|}}.$$

739 Here $\langle h, L \rangle$ denotes the span of h and L , and note that if $L = \{0\}$ we get $h_L = h$,
 740 the PCA estimator. A nontrivial example for the selection would be $L_p = \langle q \rangle$, which
 741 generates h_q , the oracle version of the GPS estimator in [14]. The following theorem says that
 742 asymptotically the oracle estimator (5.1) converges to the MAPS estimator (2.6).

743 **Theorem 5.1.** *Let the assumptions 1,2,3 and 4 hold. Suppose $\{L_p\}$ be any sequence of*
 744 *random linear subspaces that is independent of the entries of Z , such that $\dim(L_p)$ is a square*
 745 *root dominated sequence. Then*

$$746 \quad (5.2) \quad \lim_{p \rightarrow \infty} \|\hat{h}_L - h_L\| = 0.$$

747 The proof of Theorem 5.1 requires the following proposition, proved in [18].

748 **Proposition 5.2.** *Under the assumptions of Theorem 5.1, let $h = h_{PCA}$ be the PCA esti-*
 749 *mator, equal to the unit leading eigenvector of the sample covariance matrix. Then, almost*
 750 *surely:*

$$751. \quad \lim_{p \rightarrow \infty} \left(\frac{\langle h, \text{proj}(h) \rangle}{\langle h, L \rangle} - \frac{\langle h, b \rangle^2}{\langle h, L \rangle \langle b, \text{proj}(b) \rangle} \right) = 0,$$

$$752. \quad \lim_{p \rightarrow \infty} \left(\frac{\langle b, \text{proj}(h) \rangle}{\langle b, L \rangle} - \frac{\langle h, b \rangle \langle b, \text{proj}(b) \rangle}{\langle h, L \rangle \langle b, L \rangle} \right) = 0, \quad \text{and}$$

$$753. \quad \lim_{p \rightarrow \infty} \left\| \frac{\text{proj}(h)}{\langle \text{proj}(h), L \rangle} - \frac{\langle h, b \rangle \text{proj}(b)}{\langle h, L \rangle \langle b, L \rangle} \right\| = 0.$$

754 *In particular, $\frac{\text{proj}(h)}{\langle \text{proj}(h), L \rangle}$ converges asymptotically to $\frac{\text{proj}(b)}{\langle \text{proj}(b), L \rangle}$.*

755 *Proof of the Theorem 5.1:* Recall from (2.6) that,

$$756 \quad \hat{h}_L = \frac{\tau_p h + \text{proj}(h)}{\|\tau_p h + \text{proj}(h)\|_L} \quad \text{where} \quad \tau_p = \frac{\psi_p^2 - \|\text{proj}(h)\|_L^2}{1 - \psi_p^2}.$$

757 By Lemma 2.1, ψ_p has an almost sure limit $\psi_\infty = (h, b)_\infty \in (0, 1)$, and hence τ_p is bounded
758 in p almost surely.

759 Let $\Omega_1 \subset \Omega$ be the almost sure set for which the conclusions of Proposition 5.2 hold.
760 Define the notation

$$761 \quad a_p(\omega) = \|\hat{h}_{L_p} - h_{L_p}\|$$

762 and

$$763 \quad \gamma_p = \frac{(h, b) - (b, \text{proj}(h))_L}{1 - \|\text{proj}(h)\|_L^2}.$$

764 The proof will follow steps 1-4 below:

1. For every $\omega \in \Omega_1$ and sub-sequence $\{p_k\}_{k=1}^\infty \subset \{p\}_1^\infty$ satisfying

$$\limsup_{k \rightarrow \infty} \|\text{proj}(b)\|_{L_{p_k}}(\omega) < 1$$

765 we prove

$$766 \quad 0 < \liminf_{k \rightarrow \infty} \gamma_{p_k}(\omega) \leq \limsup_{k \rightarrow \infty} \gamma_{p_k}(\omega) < \infty$$

767 and

$$768 \quad 0 < \liminf_{k \rightarrow \infty} \tau_{p_k}(\omega) \leq \limsup_{k \rightarrow \infty} \tau_{p_k}(\omega) < \infty.$$

2. For every $\omega \in \Omega_1$ and sub-sequence $\{p_k\}_{k=1}^\infty \subset \{p\}_1^\infty$ satisfying

$$\limsup_{k \rightarrow \infty} \|\text{proj}(b)\|_{L_{p_k}}(\omega) < 1$$

769 we use step 1 to prove $\lim_{k \rightarrow \infty} a_{p_k}(\omega) = 0$

770 Set $\Omega_0 = \{\omega \in \Omega \mid \limsup_{p \rightarrow \infty} \|\text{proj}(b)\|_{L_p} = 1\}$. Fix $\omega \in \Omega_0 \cap \Omega_1$ and prove using step 2 that

$$771 \quad \lim_{p \rightarrow \infty} a_p(\omega) = 0$$

772 Finish the proof by applying step 2 for all $\omega \in \Omega_0^c \cap \Omega_1$ when $\{p_k\}$ is set to $\{p\}$.

773 **Step 1:** Since $\omega \in \Omega_1$ we have the following immediate implications of Proposition 5.2,

$$774 \quad (5.3) \quad \limsup_{k \rightarrow \infty} \|\text{proj}(h)\|_{L_{p_k}}^2 = (h, b)_\infty^2 \limsup_{k \rightarrow \infty} \|\text{proj}(b)\|_{L_{p_k}}^2.$$

775

$$776 \quad (5.4) \quad \limsup_{k \rightarrow \infty} (b, \text{proj}(h))_{L_{p_k}} = (h, b)_\infty \limsup_{k \rightarrow \infty} \|\text{proj}(b)\|_{L_{p_k}}^2.$$

777 Using the assumption $\limsup_{k \rightarrow \infty} \|\text{proj}(b)\|_{L_{p_k}}^2 < 1$, we update (5.3) and (5.4) as,

$$778 \quad (5.5) \quad \limsup_{k \rightarrow \infty} \|\text{proj}(h)\|_{L_{p_k}}^2 < (h, b)_\infty^2 < 1$$

779

$$780 \quad (5.6) \quad \limsup_{k \rightarrow \infty} (b, \text{proj}(h))_{L_{p_k}} < (h, b)_\infty$$

781 for the given $\omega \in \Omega_1$. We can use (5.5) on the numerator of τ_{p_k} to show,

$$782 \quad \liminf_{k \rightarrow \infty} (\psi_{p_k}^2 - \|\text{proj}(h)\|_{L_{p_k}}) \geq \liminf_{k \rightarrow \infty} \psi_{p_k}^2 - \limsup_{k \rightarrow \infty} \|\text{proj}(h)\|_{L_{p_k}}^2$$

$$783 \quad = (h, b)_\infty^2 - \limsup_{k \rightarrow \infty} \|\text{proj}(h)\|_{L_{p_k}}^2 > 0.$$

784

785 That together with the fact that the denominator of τ_{p_k} has a limit in $(0, \infty)$ implies,

$$786 \quad (5.7) \quad 0 < \liminf_{k \rightarrow \infty} \tau_{p_k}(\omega) \leq \limsup_{k \rightarrow \infty} \tau_{p_k}(\omega) < \infty$$

787 Similarly we can use (5.6) on the numerator of γ_{p_k} as,

$$788 \quad (5.8) \quad \liminf_{k \rightarrow \infty} ((h, b) - (b, \text{proj}(h))_{L_{p_k}}) \geq (h, b)_\infty - \limsup_{k \rightarrow \infty} (b, \text{proj}(h))_{L_{p_k}} > 0.$$

789 Also (5.5) can be used on the denominator of γ_{p_k} as,

$$790 \quad (5.9) \quad \liminf_{k \rightarrow \infty} 1 - \|\text{proj}(h)\|_{L_{p_k}}^2 > 1 - \limsup_{k \rightarrow \infty} \|\text{proj}(h)\|_{L_{p_k}}^2 > 0$$

791 Using (5.8) and (5.9) we get,

$$792 \quad (5.10) \quad 0 < \liminf_{k \rightarrow \infty} \gamma_{p_k}(\omega) \leq \limsup_{k \rightarrow \infty} \gamma_{p_k}(\omega) < \infty$$

793 for the given $\omega \in \Omega_1$. This completes the step 1.

794

795 **Step 2:** We have the following initial observation,

$$796 \quad (5.11) \quad 1 \geq \|\text{proj}_{\langle h, L_{p_k} \rangle}(b)\| \geq \|\text{proj}_{\langle h \rangle}(b)\| = (h, b)$$

and using that we get

$$1 \geq \limsup_{p \rightarrow} \|\text{proj}_{\langle h, L_{p_k} \rangle}(b)\| \geq \liminf_{p \rightarrow} \|\text{proj}_{\langle h, L_{p_k} \rangle}(b)\| \geq (h, b)_\infty > 0.$$

797 Given that, in order to show $\lim_{k \rightarrow \infty} a_{p_k}(\omega) = 0$, it suffices to show $\tau_{p_k} h + \text{proj}_{L_{p_k}}(h)$ converges
 798 to a scalar multiple of $\text{proj}_{\langle h, L_{p_k} \rangle}(b)$ since that scalar clears after normalizing the vectors. To
 799 motivate that lets re-write $\text{proj}_{\langle h, L_{p_k} \rangle}(b)$ as,

$$\begin{aligned} 800 \quad \text{proj}_{\langle h, L_{p_k} \rangle}(b) &= \text{proj}_{\langle h - \text{proj}_{L_{p_k}}(h), L_{p_k} \rangle}(b) \\ 801 \quad &= \text{proj}_{L_{p_k}}(b) + \left(\frac{h - \text{proj}_{L_{p_k}}(h)}{\|h - \text{proj}_{L_{p_k}}(h)\|}, b \right) \frac{h - \text{proj}_{L_{p_k}}(h)}{\|h - \text{proj}_{L_{p_k}}(h)\|} \\ 802 \quad (5.12) \quad &= \text{proj}_{L_{p_k}}(b) + \gamma_{p_k}(h - \text{proj}_{L_{p_k}}(h)) \end{aligned}$$

$$803 \quad (5.13) \quad = \gamma_{p_k} \left(h + \frac{1}{\gamma_{p_k}} \text{proj}_{L_{p_k}}(b) - \text{proj}_{L_{p_k}}(h) \right).$$

805 We also have,

$$806 \quad (5.14) \quad \tau_{p_k} h + \text{proj}_{L_{p_k}}(h) = \tau_{p_k} \left(h + \frac{1}{\tau_{p_k}} \text{proj}_{L_{p_k}}(h) \right).$$

807 Since we have τ_{p_k} and γ_{p_k} satisfying (5.7) and (5.10) respectively, we have the equations (5.13)
 808 and (5.14) well defined asymptotically, which is sufficient for our purpose. Hence, from the
 809 above argument it is sufficient to show the convergence of $h + \frac{1}{\tau_{p_k}} \text{proj}_{L_{p_k}}(h)$ to $h + \frac{1}{\gamma_{p_k}} \text{proj}_{L_{p_k}}(b) -$
 810 $\text{proj}_{L_{p_k}}(h)$. That is equivalent to showing $\frac{1}{\tau_{p_k}} \text{proj}_{L_{p_k}}(h)$ converges to $\frac{1}{\gamma_{p_k}} \text{proj}_{L_{p_k}}(b) - \text{proj}_{L_{p_k}}(h)$. We can
 811 re-write the associated quantity as,

$$812 \quad (5.15) \quad \left| \frac{1}{\tau_{p_k}} \text{proj}_{L_{p_k}}(h) - \left(\frac{1}{\gamma_{p_k}} \text{proj}_{L_{p_k}}(b) - \text{proj}_{L_{p_k}}(h) \right) \right| = \left| \left(1 + \frac{1}{\tau_{p_k}} \right) \text{proj}_{L_{p_k}}(h) - \frac{1}{\gamma_{p_k}} \text{proj}_{L_{p_k}}(b) \right|$$

813 Using Proposition 5.2 part 3 in (5.15), it is equivalent to prove

814 $\left| \left(1 + \frac{1}{\tau_{p_k}} \right) (h, b) - \frac{1}{\gamma_{p_k}} \right|$ converges to 0. We re-write it as

$$\begin{aligned} 815 \quad \left| \left(\frac{1}{\tau_{p_k}} + 1 \right) (h, b) - \frac{1}{\gamma_{p_k}} \right| &= \left| \frac{(h, b)(1 - \|\text{proj}_{L_{p_k}}(h)\|^2)}{\psi_{p_k}^2 - \|\text{proj}_{L_{p_k}}(h)\|^2} - \frac{1 - \|\text{proj}_{L_{p_k}}(h)\|^2}{(h, b) - (\text{proj}_{L_{p_k}}(h), b)} \right| \\ 816 \quad (5.16) \quad &= \left| 1 - \|\text{proj}_{L_{p_k}}(h)\|^2 \right| \left| \frac{(h, b)}{\psi_{p_k}^2 - \|\text{proj}_{L_{p_k}}(h)\|^2} - \frac{1}{(h, b) - (\text{proj}_{L_{p_k}}(h), b)} \right| \\ 817 \end{aligned}$$

818 Using parts (1) and (2) of Proposition 5.2 and the fact that $\psi_{p_k}^2$ converges to $(h, b)_\infty^2$ shows
 819 that (5.16) converges to 0 for the given $\omega \in \Omega_1$. This completes step 2.

820 **Step 3:** Fix $\omega \in \Omega_0 \cap \Omega_1$. To show that $\lim_{p \rightarrow \infty} a_p(\omega) = 0$, it suffices to show that for any sub-
 821 sequence $\{p_k\}_{k=1}^\infty \subset \{p\}_1^\infty$ there exist a further sub-sequence $\{s_t\}_{t=1}^\infty$ such that $\lim_{t \rightarrow \infty} a_{s_t}(\omega) = 0$.
 822 Let $\{p_k\}_{k=1}^\infty$ be a subsequence. We have one of the following cases,

$$823 \quad \limsup_{k \rightarrow \infty} \|\text{proj}(b)\|_{L_{p_k}}(\omega)^2 < 1$$

824 or

$$825 \quad \limsup_{k \rightarrow \infty} \|\text{proj}(b)\|_{L_{p_k}}(\omega)^2 = 1$$

826 If it is strictly less than 1, then we get from the step 2 that $\lim_{k \rightarrow \infty} a_{p_k}(\omega) = 0$. In that case
 827 we take the further sub-sequence of equal to $\{p_k\}$.

828 If it is equal to 1, then we get a further sub-sequence $\{s_t\}$ s.t
 829 $\lim_{t \rightarrow \infty} \|\text{proj}(b)\|_{L_{s_t}}^2 = 1$. Using this and Proposition 5.2 we get the following,

$$830 \quad \lim_{t \rightarrow \infty} \|\text{proj}(h)\|_{L_{s_t}}^2 = (h, b)_\infty^2 \quad \text{and} \quad \lim_{t \rightarrow \infty} (b, \text{proj}(h))_{L_{s_t}} = (h, b)_\infty$$

831 which implies $\lim_{t \rightarrow \infty} \tau_{s_t}(\omega) = \lim_{t \rightarrow \infty} \gamma_{s_t}(\omega) = 0$. Using this on the definition of \hat{h}_L and the
 832 equation (5.12) we get,

$$833 \quad (5.17) \quad \lim_{t \rightarrow \infty} \left\| \hat{h}_{L_{s_t}} - \frac{\text{proj}(h)}{\|\text{proj}(h)\|_{L_{s_t}}} \right\| = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} \left\| h_{L_{s_t}} - \frac{\text{proj}(b)}{\|\text{proj}(b)\|_{L_{s_t}}} \right\| = 0$$

834 We can now decompose $a_{s_t} = \|\hat{h}_{L_{s_t}} - h_{L_{s_t}}\|$ into familiar components via the triangle inequality
 835 as follows,

$$836 \quad a_{s_t} = \|\hat{h}_{L_{s_t}} - h_{L_{s_t}}\| \leq \left\| \hat{h}_{L_{s_t}} - \frac{\text{proj}(h)}{\|\text{proj}(h)\|_{L_{s_t}}} \right\| + \left\| h_{L_{s_t}} - \frac{\text{proj}(b)}{\|\text{proj}(b)\|_{L_{s_t}}} \right\|$$

$$837 \quad + \left\| \frac{\text{proj}(b)}{\|\text{proj}(b)\|_{L_{s_t}}} - \frac{\text{proj}(h)}{\|\text{proj}(h)\|_{L_{s_t}}} \right\|$$

839 Using (5.17), we know that the first and the second terms on the right hand side converge to
 840 0 for the given $\omega \in \Omega_0 \cap \Omega_1$. Since we have $\lim_{t \rightarrow \infty} \|\text{proj}(h)\|_{L_{s_t}}^2 = (h, b)_\infty^2$ and $\lim_{t \rightarrow \infty} \|\text{proj}(b)\|_{L_{s_t}}^2 = 1$,
 841 proving the third term on the right hand side converges to 0 is equivalent to proving

$$842 \quad \lim_{t \rightarrow \infty} \left\| \frac{\text{proj}(h)}{\|\text{proj}(h)\|_{L_{s_t}}} - \frac{(h, b)\text{proj}(b)}{\|\text{proj}(b)\|_{L_{s_t}}} \right\| = 0,$$

843 which is true by Proposition 5.2. This completes the step 3.

844

845 **Step 4:** In step 3 we proved the theorem for every $\omega \in \Omega_0 \cap \Omega_1$. Replacing $\{p_k\}$ in step
846 2 by the whole sequence of indices $\{p\}$, we get the theorem for every $\omega \in \Omega_0^c \cap \Omega_1$. These
847 together shows that we have,

$$848 \quad \lim_{p \rightarrow \infty} a_p(w) = 0 \quad \text{for all } \omega \in \Omega_1$$

849 which completes the proof of Theorem 5.1. ■

850 **5.2. Proof of Theorem 2.2.** The proof of the first part of Theorem 2.2 is an immediate
851 application of Theorem 5.1.

852 *Proof of the Theorem 2.2(2.8):* From the definitions of h_L and h_q , and as long as $q \in L_p$,
853 we have

$$854 \quad \|h_{L_p} - b\| \leq \|h_q - b\|$$

855 and therefore

$$\begin{aligned} 856 \quad \|\hat{h}_{L_p} - b\| &\leq \|\hat{h}_{L_p} - h_{L_p}\| + \|h_{L_p} - b\| \\ 857 \quad &\leq \|\hat{h}_{L_p} - h_{L_p}\| + \|h_q - b\| \\ 858 \quad &\leq \|\hat{h}_{L_p} - h_{L_p}\| + \|\hat{h}_q - b\| \end{aligned}$$

859 since $\|h_q - b\| \leq \|\hat{h}_q - b\|$ for all p . Applying Theorem 5.1 gives ■

$$860 \quad \limsup \|\hat{h}_{L_p} - b\| \leq \lim_{p \rightarrow \infty} \|\hat{h}_q - b\|.$$

861 To prove the remainder of Theorem 2.2 we need the following intermediate result concern-
862 ing uniform random subspaces, proved in [18].

863 **Proposition 5.3.** *Suppose, for each p , z_p is a (possibly random) point in \mathbb{S}^{p-1} and \mathcal{H}_p is a*
864 *uniform random subspace of \mathbb{R}^p that is independent of z_p . Assume the sequence $\{\dim \mathcal{H}_p\}$ is*
865 *square root dominated.*

866 *Then*

$$867 \quad \lim_{p \rightarrow \infty} \|\text{proj}_{\mathcal{H}_p}(z_p)\|^2 = 0 \quad \text{almost surely.}$$

868 *Proof of the Theorem 2.2 (2.9 and 2.10).* Theorem 5.1 is applicable. Hence, it suffices to
869 prove the results for the oracle version of the MAPS estimator.

870 Since the scalars clear after normalization, it suffices to prove the following assertions,

$$871 \quad (5.18) \quad \lim_{p \rightarrow \infty} \left\| \text{proj}_{\langle h, \mathcal{H} \rangle}(b) - \text{proj}_{\langle h \rangle}(b) \right\|_2 = 0$$

872 and

$$873 \quad (5.19) \quad \lim_{p \rightarrow \infty} \left\| \text{proj}_{\langle h, q, \mathcal{H} \rangle}(b) - \text{proj}_{\langle h, q \rangle}(b) \right\|_2 = 0.$$

874 We first consider (5.18), rewriting the left hand side as

$$\begin{aligned}
 875 \quad & \lim_{p \rightarrow \infty} \left\| \text{proj}_{\mathcal{H}}(b) + \text{proj}_{h - \text{proj}_{\mathcal{H}}(h)}(b) - \text{proj}_{\langle h \rangle}(b) \right\|_2 \\
 876 \quad (5.20) \quad & \leq \left\| \text{proj}_{\mathcal{H}}(b) \right\|_2 + \left\| \text{proj}_{h - \text{proj}_{\mathcal{H}}(h)}(b) - \text{proj}_{\langle h \rangle}(b) \right\|_2 \\
 877 \quad &
 \end{aligned}$$

878 The first term of (5.20) converges to 0 by setting $z = b$ in Proposition 5.3. Moreover, Propo-
 879 sitions 5.3 and 5.2 imply $\text{proj}_{\mathcal{H}}(h)$ converges to the origin in the ℓ_2 norm. Hence we have
 880 $h - \text{proj}_{\mathcal{H}}(h)$ is converging to h in ℓ_2 norm. That implies the second term in (5.20) converges
 881 to 0, which in turn proves (5.18).

882 Next, rewrite the expression in the assertion (5.19) as,

$$\begin{aligned}
 883 \quad & \left\| \text{proj}_{\mathcal{H}}(b) + \text{proj}_{\langle h - \text{proj}_{\mathcal{H}}(h), q - \text{proj}_{\mathcal{H}}(q) \rangle}(b) - \text{proj}_{\langle h, q \rangle}(b) \right\| \\
 884 \quad (5.21) \quad & \leq \left\| \text{proj}_{\mathcal{H}}(b) \right\| + \left\| \text{proj}_{\langle h - \text{proj}_{\mathcal{H}}(h), q - \text{proj}_{\mathcal{H}}(q) \rangle}(b) - \text{proj}_{\langle h, q \rangle}(b) \right\| \\
 885 \quad &
 \end{aligned}$$

886 Similarly the first term of (5.21) converges to 0 by Proposition 5.3. Note that 5.3 also applies
 887 when we set $z = q$, and hence $\text{proj}_{\mathcal{H}}(q)$ converges to the origin in the ℓ_2 norm. Hence the basis
 888 elements of $\langle h - \text{proj}_{\mathcal{H}}(h), q - \text{proj}_{\mathcal{H}}(q) \rangle$ converge to the basis elements of $\langle h, q \rangle$, which
 889 implies the second term of (5.21) converges to 0 as well. That completes the proof. ■

890 **5.3. Proof of Theorem 2.3.** We need the following lemma.

891 **Lemma 5.4.** *Let $\mathcal{P}(p)$ be a sequence of uniform β -ordered partitions such that $\lim_{p \rightarrow \infty} |\mathcal{P}(p)| =$
 892 ∞ . Then for $L_p = L(\mathcal{P}(p))$ we have,*

$$893 \quad (5.22) \quad \lim_{p \rightarrow \infty} \left\| \text{proj}_L(b) \right\| = 1$$

894 *almost surely.*

895 *Proof.* To be more precise about $L = L(\mathcal{P})$, set $\mathcal{P}(p) = \{I_1, I_2, \dots, I_{k_p}\}$ and denote the
 896 defining basis of the corresponding subspace $L_p = L(\mathcal{P})$ by the orthonormal set $\{v_1, v_2, \dots, v_{k_p}\}$.
 897 Then

$$\begin{aligned}
898 \quad 1 - \|\text{proj}_L(b)\|^2 &= 1 - \lim_{p \rightarrow \infty} \sum_{i=1}^{k_p} (b, v_i)^2 \\
899 \quad &= \sum_{i=1}^p b_i^2 - \lim_{p \rightarrow \infty} \sum_{i=1}^{k_p} (b, v_i)^2 \\
900 \quad &= \lim_{p \rightarrow \infty} \frac{1}{\|\beta\|^2} \sum_{i=1}^{k_p} \left(\sum_{j \in I_i} \beta_j^2 - \frac{1}{|I_i|} \left(\sum_{j \in I_i} \beta_j \right)^2 \right) \\
901 \quad (5.23) \quad &= \lim_{p \rightarrow \infty} \frac{1}{\|\beta\|^2} \sum_{i=1}^{k_p} \left(\sum_{j \in I_i} \left(\beta_j - \frac{1}{|I_i|} \left(\sum_{j \in I_i} \beta_j \right) \right)^2 \right) \\
902 \quad &
\end{aligned}$$

903 Now define the random variables $a_i = \max_{j \in I_i}(\beta_j)$, $c_i = \min_{j \in I_i}(\beta_j)$ for all $1 \leq i \leq k_p$. Without
904 loss of generality, $c_{k_p} \leq a_{k_p} \leq \dots \leq c_1 \leq a_1$. Since the sequence $\{\mathcal{P}(p)\}$ is uniform, there exists
905 $M > 0$ such that

$$906 \quad (5.24) \quad \max_{I \in \mathcal{P}(p)} |I| \leq \frac{Mp}{|\mathcal{P}(p)|}.$$

907 Then

$$\begin{aligned}
908 \quad \lim_{p \rightarrow \infty} \frac{1}{\|\beta\|^2} \sum_{i=1}^{k_p} \left(\sum_{j \in I_i} \left(\beta_j - \frac{1}{|I_i|} \left(\sum_{j \in I_i} \beta_j \right) \right)^2 \right) &\leq \lim_{p \rightarrow \infty} \frac{1}{\|\beta\|^2} \sum_{i=1}^{k_p} |I_i| (a_i - c_i)^2 \\
909 \quad (5.25) \quad &\leq \lim_{p \rightarrow \infty} \frac{\frac{Mp}{k_p}}{\|\beta\|^2} \sum_{i=1}^{k_p} (a_i - c_i)^2 \\
910 \quad (5.26) \quad &= \lim_{p \rightarrow \infty} \frac{M}{\frac{\|\beta\|^2}{p}} \frac{1}{k_p} (a_1 - c_{k_p})^2 \\
911 \quad &
\end{aligned}$$

912 The term $a_1 - c_{k_p}$ appearing in (5.26) is uniformly bounded since the β 's are uniformly
913 bounded. Also, $\frac{\|\beta\|^2}{p}$ is finite and away from zero asymptotically. Using those together with
914 the fact that $\lim_{p \rightarrow \infty} k_p = \infty$ we get the limit in (5.26) equal to 0 for any realization of the
915 random variables β . Note that this is stronger than almost sure convergence. ■

916 *Proof of the Theorem 2.3:* By an application of Theorem 5.1 it suffices to prove the the-
917 orem for the oracle version of the MAPS estimator. Now

$$918 \quad (5.27) \quad \left\| b - \underset{\langle h, L \rangle}{\text{proj}}(b) \right\|^2 \leq \left\| b - \underset{L}{\text{proj}}(b) \right\|^2 = 1 - \left\| \underset{L}{\text{proj}}(b) \right\|^2$$

919 and note that application of Lemma 5.4 shows that $\left\| \underset{L}{\text{proj}}(b) \right\|$ converges to 1 as p tends to
920 ∞ . ■

921 **5.4. Proof of Theorem 2.4.** The proof of Theorem 2.4 requires the following proposition,
 922 from which the first part (2.16) of the theorem easily follows. The proof of the proposition,
 923 along with the more difficult proof of the the strict inequality (2.17), appears in [18].

924 Recall that h_1, h_2 and h are the PCA leading eigenvectors of the sample covariance matrices
 925 of the returns R_1, R_2 and R , respectively.

926 **Proposition 5.5.** *For each p there is a vector \tilde{h} in the linear subspace $L \subset R^p$ generated by*
 927 *h_1 and h_2 such that $\lim_{p \rightarrow \infty} \|\tilde{h} - h\| = 0$ almost surely.*

928 *Proof of (2.16) of Theorem 2.4.* Since $\dim(L_p) = 2$ and $L_p = \text{span}(h_1, q)$ is independent
 929 of the asset specific portion Z_2 of the current block, Theorem 2.1 implies that \hat{h}_L converges
 930 to h_L almost surely in ℓ_2 norm. Hence it suffices to establish the result for the oracle versions
 931 of the MAPS and the GPS estimators.

932 Note

$$933 \quad (5.28) \quad (h_L, b) = \left\| \text{proj}_{\text{span}(q, h_1, h_2)}(b) \right\|$$

934

$$935 \quad (5.29) \quad (h_q^s, b) = \left\| \text{proj}_{\text{span}(q, h_2)}(b) \right\|$$

936

$$937 \quad (5.30) \quad (h_q^d, b) = \left\| \text{proj}_{\text{span}(q, h)}(b) \right\|$$

Using Proposition 5.5 we know there exist $\tilde{h} \in \text{span}(h_1, h_2)$ such that \tilde{h} converges to h in ℓ_2
 almost surely. Since $\text{span}(q, \tilde{h}) \subset \text{span}(q, h_1, h_2)$,

$$\left\| \text{proj}_{\text{span}(q, h_1, h_2)}(b) \right\| \geq \left\| \text{proj}_{\text{span}(q, \tilde{h})}(b) \right\|.$$

938 Taking the limits of both sides we get

$$939 \quad (5.31) \quad \lim_{p \rightarrow \infty} (h_L, b) = \lim_{p \rightarrow \infty} \left\| \text{proj}_{\text{span}(q, h_1, h_2)}(b) \right\| \geq \lim_{p \rightarrow \infty} \left\| \text{proj}_{\text{span}(q, h)}(b) \right\| = \lim_{p \rightarrow \infty} (h_q^d, b).$$

940 Similarly, since $\text{span}(q, h_1) \subset \text{span}(q, h_1, h_2)$,

$$941 \quad (5.32) \quad \lim_{p \rightarrow \infty} (h_L, b) = \lim_{p \rightarrow \infty} \left\| \text{proj}_{\text{span}(q, h_1, h_2)}(b) \right\| \geq \lim_{p \rightarrow \infty} \left\| \text{proj}_{\text{span}(q, h_1)}(b) \right\| = \lim_{p \rightarrow \infty} (h_q^s, b).$$

942 Inequalities (5.31) and (5.32) complete the proof of Theorem 2.4(a). ■

943 **6. Open Questions.** The MAPS approach to estimation of eigenvectors in a factor model
 944 setting is flexible because it allows for a general way to inject additional information, in
 945 the form of additional anchor points, to improve the estimate. Yet in this paper we have
 946 focused on a very simple setting in order to highlight the ideas: a one-factor model with
 947 homogeneous specific risk. Moreover, our error measures related to portfolio optimization –
 948 tracking error and variance forecast ratio – have focused on the performance of the minimum
 949 variance portfolio (motivated by [14]).

950 Here are a few directions for ongoing and future research.

- 951 • How effective can MAPS estimators be in the context of multifactor models, and with
952 variable specific risk? In that setting what are more general connections between ℓ_2
953 error of betas and tracking error of optimal portfolios?
- 954 • What is the general relationship between optimal MAPS shrinkage targets and the
955 linear constraints in a portfolio optimization problem?
- 956 • What appropriate systematic empirical tests would be most useful in evaluating MAPS
957 for practical implementation?
- 958 • The MAPS approach is general and does not depend on the specific choices of anchor
959 points analyzed here. Are there other useful sets of anchor points, for example possibly
960 excluding the vector q ? What other sources of observable information in the market
961 translate into useful anchor points for a successful MAPS estimation of beta? A simple
962 extension of Theorem 2.4 would involve the use of multiple past time blocks to create
963 multiple anchor points, for example.
- 964 • The experiments of Section 4.2 involving historical betas and partitions defined by
965 industry sectors had the advantage that sectors define an *a priori* partition that doesn't
966 require unobservable information. This is only one way that a β -ordered partition
967 might be approximated. Another possibility could be to use historical volatilities to
968 form a rank ordering and subsequent partition and anchor points. However, since
969 volatilities are correlated with historical betas, adding volatility anchor points and
970 then computing ℓ_2 error against historical betas would be an unfair test. Instead, a
971 different experiment could be designed using some out-of-sample measure of success
972 in place of the ℓ_2 error.
- 973 • The selection of a shrinkage target from observable data may be suited to a machine
974 learning approach to covariance estimation. One or more anchor points could be the
975 output of a trained neural network that could in principle be fed with a much larger
976 universe of observable data than simply the history of returns. This could potentially
977 take the eigenvector shrinkage approach into a much wider realm of applicability.

REFERENCES

- 978
- 979 [1] R. ALMGREN AND N. CHRISS, *Optimal portfolios from ordering information*, J. of Risk, 9 (2006), pp. 1–47.
- 980 [2] P. J. BICKEL AND E. LEVINA, *Covariance regularization by thresholding*, The Annals of Statistics, (2008),
981 pp. 2577–2604.
- 982 [3] G. CHAMBERLAIN AND M. ROTHSCHILD, *Arbitrage, factor structure, and mean-variance analysis on large*
983 *asset markets*, Econometrica, 51 (1983), pp. 1281–1304.
- 984 [4] R. CLARKE, H. DE SILVA, AND S. THORLEY, *Minimum-variance portfolio composition*, Journal of Port-
985 folio Management, 2 (2011), pp. 31–45.
- 986 [5] G. CONNOR AND R. A. KORAJCZYK, *Performance measurement with the arbitrage pricing theory: A new*
987 *framework for analysis*, Journal of financial economics, 15 (1986), pp. 373–394.
- 988 [6] G. CONNOR AND R. A. KORAJCZYK, *Risk and return in equilibrium apt: Application of a new test*
989 *methodology*, Journal of financial economics, 21 (1988), pp. 255–289.
- 990 [7] B. EFRON AND C. MORRIS, *Data analysis using stein's estimator and its generalizations*, J. of the Amer-
991 ican Statistical Assoc., 70 (1975), pp. 311–319.
- 992 [8] N. E. EL KAROUI, *Spectrum estimation for large dimensional covariance matrices using random matrix*
993 *theory*, The Annals of Statistics, (2008), pp. 2757–2790.
- 994 [9] J. FAN, Y. FAN, AND J. LV, *High dimensional covariance matrix estimation using a factor model*, Journal
995 of Econometrics, 147 (2008), pp. 186–197.

- 996 [10] J. FAN, Y. LIAO, AND M. MINCHEVA, *Large covariance estimation by thresholding principal orthogonal*
997 *complements*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75 (2013),
998 pp. 603–680.
- 999 [11] D. FOURDRINIER, W. STRAWDERMAN, AND M. WELLS, *Shrinkage Estimation*, Springer, 2018.
- 1000 [12] P. A. FROST AND J. E. SAVARINO, *An empirical bayes approach to efficient portfolio selection*, Journal
1001 of Financial and Quantitative Analysis, 21 (1986), pp. 293–305.
- 1002 [13] L. GOLDBERG AND A. KERCHEVAL, *James stein for eigenvectors*, preprint, (2022).
- 1003 [14] L. GOLDBERG, A. PAPANICOLAOU, AND A. SHKOLNIK, *The dispersion bias*, SIAM Journal on Financial
1004 Mathematics, (2022). to appear.
- 1005 [15] L. R. GOLDBERG, A. PAPANICOLAOU, A. SHKOLNIK, AND S. ULUCAM, *Better betas*, Journal of Portfolio
1006 Management, 47 (2020), pp. 119–136.
- 1007 [16] M. GRUBER, *Improving efficiency by shrinkage*, CRC Press, 1998.
- 1008 [17] H. GURDOGAN, *Eigenvector Shrinkage for Estimating Covariance Matrices*, PhD thesis, Florida State
1009 University, 2021.
- 1010 [18] H. GURDOGAN AND A. KERCHEVAL, *Multi anchor point shrinkage for the sample covariance matrix*
1011 *(extended version)*, arXiv 2109.00148, (2021).
- 1012 [19] P. HALL, J. S. MARRON, AND A. NEEMAN, *Geometric representation of high dimension, low sample*
1013 *size data*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67 (2005),
1014 pp. 427–444.
- 1015 [20] W. JAMES AND C. STEIN, *Estimation with quadratic loss*, Proceedings of the fourth Berkeley symposium
1016 on mathematical statistics and probability, (1961), pp. 361–379.
- 1017 [21] T. L. LAI AND H. XING, *Statistical models and methods for financial markets*, Springer, 2008.
- 1018 [22] O. LEDOIT AND M. WOLF, *Improved estimation of the covariance matrix of stock returns with an appli-*
1019 *cation to portfolio selection*, Journal of empirical finance, 10 (2003), pp. 603–621.
- 1020 [23] O. LEDOIT AND M. WOLF, *Honey, I shrunk the sample covariance matrix*, The Journal of Portfolio
1021 Management, 30 (2004), pp. 110–119.
- 1022 [24] O. LEDOIT AND M. WOLF, *Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz*
1023 *meets goldilocks*, The Review of Financial Studies, 30 (2017), pp. 4349–4388.
- 1024 [25] H. MARKOWITZ, *Portfolio selection*, The Journal of Finance, 7 (1952), pp. 77–91.
- 1025 [26] B. ROSENBERG, *Extra-market components of covariance in security returns*, Journal of Financial and
1026 Quantitative Analysis, 9 (1974), pp. 263–274.
- 1027 [27] S. A. ROSS, *The arbitrage theory of capital asset pricing*, Journal of economic theory, 13 (1976), pp. 341–
1028 360.
- 1029 [28] W. SHARPE, *A simplified model for portfolio analysis*, Management Science, 9 (1963), pp. 277–293.
- 1030 [29] A. SHKOLNIK, *James-stein estimation of the first principal component*, Stat, Wiley Online Library,
1031 <https://doi.org/10.1002/sta4.419> (2021).
- 1032 [30] O. A. VASICEK, *A note on using cross-sectional information in bayesian estimation of security betas*, The
1033 Journal of Finance, 28 (1973), pp. 1233–1239.
- 1034 [31] W. WANG AND J. FAN, *Asymptotics of empirical eigenstructure for high dimensional spiked covariance*,
1035 The Annals of Statistics, 45 (2017), pp. 1342–1374.