Arbitrage Pricing Restrictions and the Predictability of Stock Returns by Statistical Factor Analysis*

Ronald Balvers
DeGroote School of Business
McMaster University
Email: balvers@mcmaster.ca

Adam Stivers
University of Wisconsin-La Crosse
Email: astivers@uwlax.edu

September 21, 2016

Abstract

In standard principal components estimation of the APT, the factors are obtained without employing the restrictions on mean returns implied by the APT. We modify the principal components methodology to allow mean returns to reflect the theoretical restrictions up to any level of accuracy and generate optimal constrained APT factors from the eigenvectors of a suitably modified covariance matrix. With the 30 industry portfolios as test assets, the resulting risk factors predict returns hedged for systematic risk better out of sample than the standard CAPM, Fama-French, and Carhart asset pricing models and better than conventional principal component factors.

JEL Classification: G12, G17, G11.

Keywords: asset pricing, arbitrage pricing, factor analysis, hedging

* We thank workshop participants at McMaster University for valuable comments.

Arbitrage Pricing Restrictions and the Predictability of Stock Returns by Statistical Factor Analysis

Introduction

The Arbitrage Pricing Theory (APT) of Ross (1976) provides the crucial insight that any set of factors covering the systematic risk of a particular group of assets should also explain the mean returns of these assets. Empirical implementation may be subdivided into three approaches. First, the macroeconomic sources approach as employed by Chen, Roll, and Ross (1986) in which a set of macro variables, thought on a priori theoretical grounds to contain all systematic risks, is used as the factors to explain asset returns. Second, the mimicking factors approach, as prominently employed by Fama and French (1993) and Fama and French (1996), for instance, in which a predetermined number of well-diversified portfolios of the assets to be evaluated are used as the risk factors.[1] The idea is that, as long as the factor portfolios are well diversified, sufficiently diverse, and equal in number to the true number of underlying factors, they will be linear combinations of the underlying factors and hence capture all systematic risk.

The third approach, the factor analysis approach by which factors are obtained statistically as linear combinations of the assets to be evaluated, is the focus of our paper.[2] We adapt the factor analytical approach here by including the theoretical APT restrictions on mean returns to improve estimation and forecast performance under the maintained hypothesis that the APT is (approximately) correct. Roll and Ross (1980) and others utilize factor analysis (FA) or principal components analysis (PCA) to identify the portfolios that best explain the variation in

[1] See also Hugerman, Kandel, and Stambaugh (1987) and Grinblatt and Titman (1987) for more on the mimicking portfolio approach.

[2] While there have been limited empirical evaluations of the APT recently, Lehmann and Modest (1988) provide a good overview of the empirical issues and estimation techniques of the APT. Grinblatt and Titman (1987) introduce the idea of local efficiency and show that the APT implies that the factors are locally efficient. This also helps the APT avoid the Roll (1977) critique.

realized returns of all assets under consideration. The number of such portfolios that have significant explanatory power for return covariances may be determined statistically from Bai and Ng (2002). Having established the factors that best explain covariances, subsequently the implication of the APT that exposures to this set of factors fully explain the differences in mean returns across assets may be tested.

Whereas FA focuses on finding the factors that explain systematic covariances and PCA focuses on finding the factors that explain total return covariances, Chamberlain and Rothschild (1983) show that asymptotically either approach will correctly identify the systematic risk factors because as the number of assets goes to infinity the systematic risk dominates any non-pervasive idiosyncratic risk. PCA allows the idiosyncratic risk matrix not to be diagonal so that returns may be correlated after correction for systematic risk as long as the idiosyncratic risk is not pervasive. PCA is also easy to implement and so we focus on (adapting) this approach.

Based on the third approach, principal component factors are used as systematic asset pricing factors to test the APT and can then be applied in practice if the APT is not rejected. However, for purposes of obtaining factors that are useful for generating improved decisions in practical applications, dealing e.g. with costs of capital[3], hedging strategies, forecasting asset returns, evaluation of portfolio managers, and event studies, it is advantageous to impose the

[3] For examples of asset pricing models used for cost of capital estimates, see Pratt and Grabowski (2008), Myers and Turnbull (1977), Goldenberg and Robin (1991), Bartholdy and Peare (2005), Fuller and Kerr (1981), and Cummins and Phillips (2005). Bruner et al. (1998), among others, provide survey evidence that practitioners typically use the CAPM in cost of capital estimation. Koedijk and van Dijk (2004) examine cost of capital from a global perspective with the CAPM, finding that the domestic market factor is sufficient for a firm to estimate cost of capital and that the CAPM and international CAPM provide little difference. Aside from estimates implied by asset pricing theory, previous research has used implied cost of capital measures (Li, Ng, and Swaminathan, 2013; Lee, Ng, and Swaminathan, 2011; Pástor, Sinha, and Swaminathan, 2008; Hou, Van Dijk, and Zhang, 2012; Easton, 2004) and accounting measures (Easton and Monahan, 2005) to estimate cost of capital.

APT restrictions in advance as part of the factor estimation so that the resulting factor weights are more reliably estimated.

The objective in this paper is to provide a new method for obtaining statistical asset pricing factors that imposes a condition limiting the pricing errors for mean returns jointly with maximizing the variance in returns explained by the factors. In the language of the standard two-pass asset pricing approach, it maximizes the first-pass fit averaged across all test assets subject to a guaranteed minimum fraction of second-pass mean pricing errors. A related work is that of Nardari and Scruggs (2007), who impose the APT pricing restriction (that of mean returns being linear combinations of systematic factors) into a latent factor model with a multivariate stochastic volatility (MSV) process under a Bayesian methodology. They find their best model in terms of its Bayes factor is the model with three latent factors with MSV and the APT restrictions. Nardari and Scruggs use a Bayesian methodology, whereas we use an approach similar to PCA (PCA while imposing APT restrictions). They assign a unit weight to certain test assets (based on correlation with a preliminary first principal component) to form their three factors while we use linear combinations of all test assets to form optimal APT factors. This allows for an investor/practitioner to find the portfolio weights and form our factors in real time. We argue that our approach provides true APT factors (rather than selecting one portfolio to serve as a factor): a linear combination of all assets that is designed to capture covariance risk while perfectly explaining mean returns. Another difference is that we look out of sample whereas Nardari and Scruggs only look at in-sample posterior distributions and cross-sectional averages. Other differences include the following: Nardari and Scruggs allow for mean errors to be nonzero, whereas our main focus is on restricting mean errors to be zero; we do not make any assumptions on volatility; we use industry portfolios with no strong factor structure while they use ten size-sorted portfolios.

Another related work is MacKinlay and Pástor (2000), which assumes that a factor is missing from a linear factor model. They show that, under the strong assumptions of homoscedastic errors and uncorrelated errors, accounting for an unobserved factor amounts to the alpha showing up in the covariance matrix. Linking means and covariances allows for improved estimation of expected returns/means and higher Sharpe ratios in out-of-sample portfolio selection/optimization. They point out that they essentially are combining two factor approaches: FA/PCA, which focuses on covariance only, and factor mimicking portfolios (Fama-French three-factor model), which focuses on means only.[4] Our approach does the same: we link means and covariances by explaining mean returns up to an allowable error while maximizing covariance simultaneously. The benefits of our approach are as follows: we do not make any strong assumptions on the error terms as does MacKinlay and Pástor, and their approach requires conditioning on the Sharpe ratio of the missing factor (which must be pre-specified by the practitioner/econometrician). Since we impose that the number of specified factors perfectly price all assets, the latter issue is avoided by our approach. We also compare to the traditional asset pricing models (means only and covariances only as mentioned previously), which MacKinlay and Pástor do not. Finally, while MacKinlay and Pástor examine maximum possible Sharpe ratios out-of-sample, we implement an out-of-sample hedging approach.

While we perform forecasting of a hedged position rather than the raw portfolio returns with our derived factors, Simin (2008) examines the out-of-sample forecast performance of the

[4] Pukthuanthong and Roll (2014) provide a way of identifying whether a factor is related to mean returns or risks (or both).

CAPM and Fama-French three-factor model, finding that they perform poorly in predicting future asset returns compared to historical averages or constants. Simin further points out that statistical APT factors and macroeconomic factors perform worse than either of these two models. This differs starkly from our results, where the APT factors perform better than the CAPM and Fama-French three-factor model. One reason may be that Simin uses a rolling 60-month estimation window, while we find that a larger estimation window allows PCA factors to become more reliable (at least in terms of bias). Simin also uses the 25 size and book-to-market portfolios as the test assets, while we use industry portfolios. The Fama-French three-factor model can explain size and book-to-market portfolios well, but not industry portfolios.

One simple approach to performing PCA and also imposing zero pricing errors would be to restrict time series constants, "alphas", to equal zero. However, in this case there is no requirement that mean pricing errors are zero so that an unknown level of pricing errors is allowed. Our approach, however, allows imposing in advance any desired mean return errors as a fraction of the squared sum of mean returns, including setting the fraction equal to zero.

An alternative way to view our method, if zero mean return errors are enforced (a 100% cross-sectional R-squared), is as an approach that generates a tangency portfolio as a portfolio of the factors and subject to this constraint finds the factors that also explain as much as possible of the variation in the time series of returns of all assets. Our approach differs from the Markowitz optimization approach, which has had limited success in out-of-sample application. For example, DeMiguel, Garlappi, and Uppal (2009) find that a naïve equal-weighted approach performs better out of sample in terms of its Sharpe ratio, turnover, and certainty equivalent return compared to the tangency portfolio. For an examination of investing in the tangency portfolio out of sample see also Best and Grauer (1991) and Michaud (1989). Kan and Zhou

(2007) show the pitfalls of using sample estimates rather than population mean and covariance when performing out-of-sample mean-variance optimization, while also finding that a sample tangency portfolio and riskless asset can be beaten by adding the sample minimum-variance portfolio. This shows that using a larger sample/estimation window (as close to population as possible) may work better. Thus, we look at a large expanding estimation window (as well as a rolling window which does not perform as well perhaps due to this reason) when performing our out-of-sample analysis. Kan and Smith (2008) show that sample minimum-variance frontiers can be highly biased. Kan, Wang, and Zhou (2016) show that accounting for estimation error allows for portfolios that outperform the equal-weighted approach of DeMiguel, Garlappi, and Uppal (2009).

A fundamental issue in this paper is that an asset pricing model generates return forecasts. To evaluate the performance of these forecasts, we may simply calculate the mean square errors (MSE) generated from realized returns relative to the forecast. As we confirm below, PCA identifies factors that minimize these mean square errors, at least in sample. And this is true, irrespective of whether the APT holds. However, a typical decision maker employing the asset pricing model may distinguish between components of the errors that are systematic and caused by a mean bias in the asset pricing model, and errors that are simply part of return risk. From this perspective, it is important to consider separately the bias and the noise in the pricing model, as does Simin (2008). Simin (2008) and Leitch and Tanner (1991) both point out that investors/practitioners may have asymmetric loss functions, which requires a need to examine both bias and variance. Others that examine the limitations of symmetric loss functions/MSE and the implications of asymmetric loss functions include Patton and Timmermann (2007), Basu and Markov (2004), Clatworthy, Peel, and Pope (2012), and Elliot, Komunjer, and Timmermann (2008). Although they do not examine the breakdown of MSE, MacKinlay and Pástor (2000) do report both bias and MSE separately. Giacomini and White (2006) and Diebold and Mariano (2002) examine the issue of separating bias and variance (and the need for examining more than just MSE) from a forecasting perspective.

To provide an indication of the usefulness of imposing APT restrictions we apply our approach to a set of test assets and compare the results both in-sample and out-of-sample against those for other basic models. In particular, we consider the Fama-French 3-factor model as representative of a model typically applied and we consider a factor model obtained from conventional PCA. The test assets we consider are the 30 Industry portfolios obtained from Fama and French (available from Kenneth French's website). These assets constitute a good target as they do not have a strong factor structure and their mean returns have traditionally been difficult to explain (see Lewellen, Nagel, and Shanken, 2010; Fama and French, 1997; Chou, Ho, and Ko, 2012).

Imposing the APT restrictions perfectly will obviously result in a zero alpha/bias in sample, and PCA by design will have the lowest possible MSE in sample.[5] We do indeed find that the MSE when imposing the APT restrictions ("constrained" PCA or C-PCA) is only slightly higher while completely eliminating bias. To judge the out-of-sample performance we consider a hedging approach where the in-sample portfolio weights and betas are used to hedge each portfolio over the next month. While this is not a true forecasting approach in that factor realizations are used, it is nevertheless an approach that is applicable in real time. The out-of-

[5] Comparing the in-sample cross-sectional R-squared of the models does not provide much insight either; Kan, Robotti, and Shanken (2013) investigate the significance of differences between cross-sectional R-squared values, rather than a simple comparison of the values, and they find that most models are not significantly different.

sample results show that again the MSEs are not significantly different between PCA and C-PCA. However, C-PCA reduces the bias out of sample by around 30% compared to PCA.

We argue that any investor or practitioner who has a long-term investment horizon in mind would prefer C-PCA due to the reduced bias, regardless of loss function.[6] In our out-of-sample application we re-estimate and change the hedged position each month, and then evaluate performance over the full out-of-sample window. While it is possible that an investor would still prefer reduced bias over this large sample period/window, it is also conceivable that the investor would implement a longer forecast horizon and avoid monthly re-estimation and rebalancing. As the forecast horizon lengthens, the target return will approach the mean return, so C-PCA should perform better regardless of loss function (its MSE should improve relative to all other models). This is also consistent with the findings of MacKinlay and Pástor (2000): as the sample size gets larger (240 and/or 360 months) the unbiased/unrestricted estimator they use performs better than their restricted estimator that links means and covariances but is biased. We also provide a theoretical and empirical discussion of the relationship between investment horizon and bias: lower bias at a smaller horizon is preferable as the horizon grows. Thus, lower bias will mean as the horizon is increased, the particular model will perform better in terms of MSE compared to a model with a higher bias.

Higher MSE due to higher variance is akin to higher month-to-month fluctuations around the target (albeit only slightly higher MSE/variance for C-PCA), while lower bias means being closer to the target on average. For example, for a long-term project, a firm would want a cost of capital estimate that is as accurate as possible, while the firm may not care much about short-

[6] Investment horizon also has a clear monotonic relationship with bias, whereas various loss functions may have contrasting implications. Also, asymmetric loss functions may place more weight on bias *or* variance depending on the specification/investor or analyst.

term variations. If the application emphasizes bias (average accuracy) even slightly in comparison to variance, then this particular practitioner prefers C-PCA. C-PCA also has a better second pass fit both in sample and out. We additionally find that PCA and C-PCA are both significantly better than the traditional asset pricing models for these industry portfolios.

While the application of the derived C-PCA pricing factors that we focus on is out-of-sample hedging, there are many potential applications of these factors. There has been extensive research applying APT/PCA factors to various settings.[7] Kozak, Nagel, and Santosh (2014) find that a stochastic discount factor based on PCA can explain away many anomalies out-of-sample. They also show it may not be possible to distinguish whether these factors are sentiment-based or risk-based. Bussière, Hoerova, and Klaus (2015) apply PCA to hedge fund returns in order to construct the main factor for a set of hedge fund assets and examine changes in the hedge funds' sensitivity to the factor over time. Herskovic et al. (2016) develop a common idiosyncratic volatility (CIV) factor, which they note can be seen as an APT factor. They find this factor to be priced and to help explain various pricing anomalies. One of the models used to find CIV (residual variance) is the first five principal components. Ludvigson and Ng (2009) conduct factor analysis on a panel of macroeconomic variables to construct macro factors which are used to forecast bond returns, and examine bond risk premia. Goyal, Pérignon, and Villa (2008) find that there are two common APT factors in both the NASDAQ and NYSE, but each one has another separate risk factor that does not extend to the other market. Chen, Hsieh, and Jordan (1997) applies the APT to the real estate market, by examining both macroeconomic risk factors and derived factors as they relate to REIT returns. They find that for some time periods the

[7] Many implementations of PCA, especially those for forecasting purposes, follow either the methodology of Stock and Watson (2002) or Forni et al. (2000). Stock and Watson use static principal components while Forni et al. use dynamic principal components, although both use dynamic factor models to forecast.

macro factors explain REIT returns better, while in other periods the two models perform equally. An example of an application of macroeconomic factors is Kavussanos, Marcoulis, and Arkoulis (2002), which applies macroeconomic factors to international industry returns and examine the investment implications of their findings.

In the following we first review in Section 2 the basic theory regarding the PCA approach and its optimality properties. In Section 3 we present the derivation of an approach that optimally imposes the APT pricing restrictions when extracting factors from a group of assets based on applying the Ky Fan maximum principle for Rayleigh quotient pencils. Section 4 applies the approach to explain the returns of the thirty Fama-French industry portfolios in and out of sample and contrasts the performance with alternative approaches. Section 5 concludes.

2. Basic Properties of Statistical Factor Analysis

Returns Specification with Factor Structure

Consider the returns of a group of assets as given by

$$(1) \qquad \mathbf{r} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{Q}'\mathbf{r} + \mathbf{e}\,,$$

where $\mathbf{r}, \boldsymbol{\alpha}, \mathbf{e}$ are *n x 1* vectors indicating, respectively, the set of returns, the mean pricing errors, and the non-systematic pricing errors for each of *n* risky assets. The matrix $\mathbf{Q}$ is an *n x k* matrix of full column rank that represents for each of the *k* factors the weights put on the *n* assets. Thus, $\mathbf{r_k} = \mathbf{Q}'\mathbf{r}$ represents the *k x 1* vector of factor returns. The factor loadings (betas) for all assets are given by the *n x k* matrix $\mathbf{B}$. We define the *n x n* covariance matrix of returns as $\boldsymbol{\Sigma} = \mathbf{E}[(\mathbf{r} - \boldsymbol{\mu})(\mathbf{r} - \boldsymbol{\mu})']$. The distribution of returns is constant over time, and, by extension, $\boldsymbol{\mu}$

and $\mathbf{\Sigma}$ are time invariant. Note that any set of returns may be represented by equation (1) without loss of generality such that $E(\mathbf{e}) = 0$ and such that $E(\mathbf{r_k e'}) = \mathbf{0}$.

Taking expectation in equation (1), the mean returns of all assets are given as

$$(2) \qquad \boldsymbol{\mu} = \boldsymbol{\alpha} + \mathbf{BQ}'\boldsymbol{\mu}\ .$$

Factors Chosen to Minimize Errors and PCA

We show first that selecting factors to minimize the sum of squared non-systematic pricing errors is equivalent to adopting the principal components as factors. While this result was noted previously (Jong and Kotz, 1999), it is not well known in the finance literature. We emphasize that PCA does not simply select the linear combinations of assets that best captures the communalities in returns, but that adopting these components as factors provides the lowest attainable quadratic variation in the return errors:

$$\underset{\mathbf{B,Q}}{Min} \quad \left[E(\mathbf{e}'\mathbf{e})\right] .$$

Since $E(\mathbf{e}'\mathbf{e}) = Tr[E(\mathbf{ee}')]$, where Tr represents the trace of the matrix, minimizing the sum of squared errors from equations (1) and (2) amounts to:

$$(3) \qquad \underset{\mathbf{B,Q}}{Min} \quad Tr\left[(\mathbf{I} - \mathbf{BQ}')\,\mathbf{\Sigma}\,(\mathbf{I} - \mathbf{QB}')\right].$$

First, for given factor weights, choose the factor loadings $\mathbf{B}$ that minimize the squared errors. Matrix differentiation in equation (3) and finding the stationary point implies $\mathbf{\Sigma Q} - \mathbf{B}(\mathbf{Q}'\mathbf{\Sigma Q}) = 0$ so that

$$(4) \qquad \mathbf{B} = \mathbf{\Sigma Q}\,(\mathbf{Q}'\mathbf{\Sigma Q})^{-1},$$

the standard formula for obtaining multivariate betas. Plug equation (4) into equation (3) to find the concentrated objective:

$$\text{(5)} \qquad \underset{\mathbf{Q}}{Min} \; Tr\left[\mathbf{\Sigma} - \mathbf{\Sigma Q}(\mathbf{Q}'\mathbf{\Sigma Q})^{-1}\mathbf{Q}'\mathbf{\Sigma}\right].$$

Since Tr ($\mathbf{\Sigma}$) is constant and using the property that Tr ($\mathbf{AB}$) = Tr ($\mathbf{BA}$) equation (5) becomes

$$\text{(6)} \qquad \underset{\mathbf{Q}}{Max} \; Tr\left[(\mathbf{Q}'\mathbf{\Sigma}^2\mathbf{Q})(\mathbf{Q}'\mathbf{\Sigma Q})^{-1}\right].$$

Taking the "square root" of the positive definite $\mathbf{\Sigma}$ and defining $\mathbf{Q}^* = \mathbf{\Sigma}^{½}\mathbf{Q}$ we obtain

$$\text{(7)} \qquad \underset{\mathbf{Q}^*}{Max} \; Tr\left[(\mathbf{Q}^{*}{}'\mathbf{\Sigma Q}^*)(\mathbf{Q}^{*}{}'\mathbf{Q}^*)^{-1}\right].$$

The trace in equation (7) is known as the Rayleigh Quotient which by the Ky Fan maximum principle (Fan, 1949. See also for instance Bhatia, 2013, p. 35, and Li, 2015) is maximized when the n x k matrix $\mathbf{Q}^*$ is orthonormal and consists of the k eigenvectors associated with the k largest eigenvalues of matrix $\mathbf{\Sigma}$ (and is minimized if $\mathbf{Q}^*$ contains the k eigenvectors corresponding to the k smallest eigenvalues of $\mathbf{\Sigma}$). Thus, we know that

(8) $\qquad \mathbf{\Sigma Q}^* = \mathbf{Q}^*\mathbf{\Delta}_k$, as well as $\mathbf{\Sigma Q} = \mathbf{Q\Delta}_k$

where $\mathbf{\Delta}_k$ is the k x k diagonal matrix containing the k largest eigenvalues of $\mathbf{\Sigma}$. Pre-multiplying both sides of equation (8) by $\mathbf{Q}^{*}{}'$ (or the second equation by $\mathbf{Q}'$) shows that the maximum trace of the Rayleigh Quotient in equation (7) is equal to Tr ($\mathbf{\Delta}_k$) which is the sum of the k largest eigenvalues. Also, we find that $Tr(\mathbf{Q}'\mathbf{\Sigma}^2\mathbf{Q})(\mathbf{Q}'\mathbf{\Sigma Q})^{-1} = Tr(\mathbf{Q}'\mathbf{\Sigma Q})(\mathbf{Q}'\mathbf{Q})^{-1}$. It follows that either the eigenvectors $\mathbf{Q}^*$ or $\mathbf{Q}$ may be used to generate the k factors $\mathbf{r}_k^* = \mathbf{Q}^{*\prime}\mathbf{r}$ or $\mathbf{r}^k = \mathbf{Q'r}$ that minimize the unexplained variation. Note that equation (8) satisfies the first-order conditions for a stationary point in $\mathbf{Q}^*$ or $\mathbf{Q}$ subject to a normalization constraint $\mathbf{Q'\Sigma Q} = \mathbf{I}_k$: if we define the Lagrangian $L = Max\ Tr[(\mathbf{Q}'\mathbf{\Sigma}^2\mathbf{Q}) - \Delta_k(\mathbf{Q}'\mathbf{\Sigma Q} - \mathbf{I}_k)]$ it generates equation (8). The second-order conditions do not hold, however, which is the reason that we need to appeal to the properties of Rayleigh Quotients to generate the result. We can view the fraction of return variation explained by the k factors as $Tr\ (\mathbf{\Delta}_k)\ /\ Tr\ (\mathbf{\Sigma})$, the sum of the k largest eigenvalues divided by the sum of all eigenvalues. The $\mathbf{Q}^*$ or $\mathbf{Q}$ matrix obtained via equation (8) consists of the first k principal components and, by Ky Fan's theorem, the optimal realized objective of equation (6) equals the sum of the k largest eigenvalues of the $\mathbf{\Sigma}$ matrix.

3. Incorporating Arbitrage Pricing Restriction

Adding APT Constraint

If the factor returns explain most of the variance in returns so that the remaining variance is non-pervasive and may be diversified (Chamberlain and Rothschild, 1983), then the APT of Ross (1976) implies that the $\boldsymbol{\alpha}$'s are (approximately) zero. To optimally generate the statistical set of factors characterized by $\mathbf{Q}$ and obtain factor returns $\mathbf{r}_k = \mathbf{Q}'\mathbf{r}$ we now consider the following optimization problem:

$$(9) \qquad \underset{\mathbf{B},\mathbf{Q},\lambda}{Min} \ \left[E(\mathbf{e}'\mathbf{e}) + \lambda(\boldsymbol{\alpha}'\boldsymbol{\alpha} - \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu})\right],$$

where $0 \le \gamma < 1$ represents the target level for the sum of squared alphas relative to the sum of squared mean returns. Thus, γ represents the targeted fraction of mean return errors allowed, i.e., the sum of squared mean pricing errors compared to the sum of squared means. Most often we will consider $\gamma = 0$ as the target which means that all mean pricing errors, the "alphas", are constrained to be zero.

Using equations (1) and (2), the properties of traces, and taking expectations we may replace equation (9) by

$$(10) \qquad \underset{\mathbf{B},\mathbf{Q},\lambda}{Min} \; Tr\left[(\mathbf{I}-\mathbf{B}\mathbf{Q}')\,\mathbf{V}(\lambda)\,(\mathbf{I}-\mathbf{Q}\mathbf{B}')\right] - \lambda\gamma\,\boldsymbol{\mu}'\boldsymbol{\mu},\; \mathbf{V}(\lambda) \equiv \boldsymbol{\Sigma} + \lambda\,\boldsymbol{\mu}\boldsymbol{\mu}'.$$

Analogously to the unconstrained case, we have

$$(11) \qquad \mathbf{B} = \mathbf{V}(\lambda)\mathbf{Q}[\mathbf{Q}'\mathbf{V}(\lambda)\mathbf{Q}]^{-1},$$

and in addition the constraint implies that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$. Note that the objective in equation (9) requires that the optimally estimated betas, $\mathbf{B}$, are not the standard OLS betas of equation (4). The relevant association between asset return and factors here also involves consideration of the mean returns. The concentrated objective becomes after eliminating $\mathbf{B}$:

$$(12) \qquad \underset{\mathbf{Q}}{Max} \; Tr\left([\mathbf{Q}'\mathbf{V}(\lambda)^2\mathbf{Q}][\mathbf{Q}'\mathbf{V}(\lambda)\mathbf{Q}]^{-1}\right),$$

where λ is set to force $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$. Similar to the unrestricted case the eigenvectors are found from

$$(13) \qquad \mathbf{V}(\lambda)\,\mathbf{Q} = (\boldsymbol{\Sigma} + \lambda\,\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{Q} = \mathbf{Q}\,\boldsymbol{\Delta}_k.$$

Pre-multiplying by $\mathbf{Q}'$ and subsequently pre- and post-multiplying both sides by $\boldsymbol{\mu}'\mathbf{Q}$ and its transpose, respectively, and using the normalization that $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_k$ implies for the optimal choice of $\mathbf{Q}$ that $\boldsymbol{\mu}_k'(\boldsymbol{\Delta}_k - \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})\boldsymbol{\mu}_k = \lambda(\boldsymbol{\mu}'\mathbf{Q}\mathbf{Q}'\boldsymbol{\mu})^2$. To impose $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$ eliminate $\boldsymbol{\alpha}$ from equation (2). Equations (11) and (13) imply $\boldsymbol{\mu}'\mathbf{Q}\mathbf{Q}'\boldsymbol{\mu} = (1-\gamma)\,\boldsymbol{\mu}'\boldsymbol{\mu}$. Thus, imposing $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$ for optimal $\mathbf{Q}$ into $\boldsymbol{\mu}_k'(\boldsymbol{\Delta}_k - \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})\boldsymbol{\mu}_k = \lambda(\boldsymbol{\mu}'\mathbf{Q}\mathbf{Q}'\boldsymbol{\mu})^2$ produces

(14) $\lambda = \boldsymbol{\mu}_k'(\boldsymbol{\Delta}_k - \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})\boldsymbol{\mu}_k / (1-\gamma)^2(\boldsymbol{\mu}'\boldsymbol{\mu})^2$.

Solving equations (13) and (14) jointly generates both the optimal $\mathbf{Q}$ and λ. Accordingly, the *k* factor returns are $\mathbf{r}^k = \mathbf{Q}'\mathbf{r}$ which guarantee the lowest non-systematic errors subject to the mean pricing errors corresponding to a target fraction $\gamma = \boldsymbol{\alpha}'\boldsymbol{\alpha}/\boldsymbol{\mu}'\boldsymbol{\mu} \geq 0$.

Appendix A provides a Matlab program for solving equations (13) and (14) simultaneously. The Lagrangian multiplier λ must be determined so that the optimization weight on reducing the alphas is sufficient to deviate enough from minimizing overall pricing errors to meet the alpha target. The resulting optimal $\mathbf{Q}$ must generate a mean square error in-sample that is higher than in the standard PCA case, but produces a larger cross-sectional R-squared that would rise to 100% for $\gamma = 0$. The approach comes down to calculating the eigenvalues and eigenvectors of matrix $\mathbf{V}(\lambda) = \boldsymbol{\Sigma} + \lambda\,\boldsymbol{\mu}\boldsymbol{\mu}'$ and using the eigenvectors of the *k* largest eigenvalues as follows from the Ky Fan maximum principle. As in standard PCA the factor loadings may be obtained from $\mathbf{B} = \mathbf{Q}$ so that for the case of $\gamma = 0$ (zero alphas) from equation (1) the predicted returns equal $\hat{\mathbf{r}} = \mathbf{Q}\mathbf{Q}'\mathbf{r} = \mathbf{Q}\mathbf{r}_{\mathbf{k}}$.

Imposing OLS Betas

The standard approach for determining the unexplained component of asset returns implied by a particular asset pricing model is to calculate alphas from “first-pass” time series regressions of the asset returns on the factor returns. In this approach the factor sensitivity estimates are usually obtained as OLS betas: The empirical results for the (non-optimal) case of OLS betas, $\mathbf{B} = \mathbf{\Sigma Q}(\mathbf{Q'\Sigma Q})^{-1}$, when the factors are specified as $\mathbf{Q}'(\mathbf{r}-\boldsymbol{\mu})$ are available upon request from the authors. While this does not provide the minimum MSE it is a reasonable case to consider as it corresponds to imposing a constraint on alphas as typically calculated. For reasons of tractability we focus here on the weighted alphas which relate to the difference in maximum squared Sharpe ratios of the assets compared to the factors as discussed by Gibbons, Ross, and Shanken (1989): $\boldsymbol{\alpha}'\mathbf{\Sigma}^{-1}\boldsymbol{\alpha} = \boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{Q}(\mathbf{Q}'\mathbf{\Sigma Q})^{-1}\mathbf{Q}'\boldsymbol{\mu}$. We set the weighted alphas equal to a target fraction of the weighted mean returns: $\boldsymbol{\alpha}'\mathbf{\Sigma}^{-1}\boldsymbol{\alpha} = \gamma(\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu})$. As before, if we set $\gamma = 0$ then this implies the constraint $\boldsymbol{\alpha} = 0$ (for all assets). Appendix B derives that for the case of

(15) $$\underset{\mathbf{Q},\lambda}{Min} \left[E(\mathbf{e}'\mathbf{e}) + \lambda(\boldsymbol{\alpha}'\mathbf{\Sigma}^{-1}\boldsymbol{\alpha} - \boldsymbol{\gamma}\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu})\right],$$

subject to equations (1) and (2) and $\mathbf{B} = \mathbf{\Sigma Q}(\mathbf{Q'\Sigma Q})^{-1}$, that the factors are found first from

(16) $$\mathbf{V}(\lambda)\mathbf{Q} = (\mathbf{\Sigma} + \lambda\,\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{\Sigma}^{-\frac{1}{2}})\mathbf{Q} = \mathbf{Q}\,\mathbf{\Delta}_k\,.$$

using the normalization that $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_k$ and, second defining $\boldsymbol{\mu}_\mathbf{k} = \mathbf{Q}'\boldsymbol{\mu}$, from

(17) $$\lambda = \boldsymbol{\mu}_k'(\mathbf{\Delta}_k - \mathbf{Q}'\mathbf{\Sigma Q})\boldsymbol{\mu}_k/(1-\gamma)^2(\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu})^2\,.$$

Then, subject to a target fit in the second pass, the factors maximize the first-pass fit.

Sub-Optimal Direct Restriction on the Alphas

An alternative way of reducing the mean pricing errors, as mentioned previously, is to omit the constant in estimating $\mathbf{B}$ and $\mathbf{Q}$. Thus, in equation (1) we set $\boldsymbol{\alpha}=\mathbf{0}$. Then $E(\mathbf{e}'\mathbf{e})$ is given as $Tr\left[(\mathbf{I}-\mathbf{BQ}')\,E(\mathbf{rr}')\,(\mathbf{I}-\mathbf{QB}')\right]$. Since $E(\mathbf{rr}')=\mathbf{\Sigma}+\boldsymbol{\mu\mu}'$, the estimation problem becomes similar to both the restricted and unrestricted cases but with $\lambda=1$ instead of the constrained on the mean pricing errors. Due to the omission of the constant, $E(\mathbf{e})\neq 0$ so that mean pricing errors are not avoided, as is the case in Nardari and Scruggs (2007). In fact, the fraction of squared mean pricing errors may be obtained from the constrained PCA case by searching for the particular target fraction of mean pricing errors, γ, that implies $\lambda=1$.

Correlation Matrix instead of Covariance Matrix

PCA analysis is often based on a correlation matrix rather than a covariance matrix. See, for instance, Jolliffe (2002). While working with returns implies there is no serious need for further normalization, there may still be cases in which a correlation matrix is preferable.

In effect, we now examine normalized returns for which $\mathbf{r}^*=\mathbf{S}^{-1}\mathbf{r}$, where $\mathbf{S}$ is an *n x n* diagonal matrix containing the *standard deviations* of the returns on the diagonal such that $\mathbf{S}^2=\mathbf{\Sigma}$. The covariance matrix of $\mathbf{r}^*$ is then the correlation matrix of $\mathbf{r}$. The only difference compared to the covariance analysis case is in the interpretation of the results. In particular, maximizing explained covariance (first-pass) now becomes maximizing explained correlation, which is in turn equivalent to maximizing the squared correlation. The latter amounts to maximizing the equal-weighted average of the R-squares for the *n* first-pass regressions (subject to a restriction on the squared mean pricing errors for the normalized returns).

4. Application to Industry Portfolios

Number of Factors and MSE Breakdown

The in-sample and out-of-sample performances of PCA and C-PCA will be compared, along with the traditional asset pricing models (CAPM, Fama-French three-factor model (FF3), and Carhart, from Carhart, 1997), using the thirty Fama-French industry-sorted monthly portfolios. The industry portfolios are chosen due to their weak factor structure. The sample starts in January 1927, when the momentum factor is first available. The sample ends in December 2015. Excess returns are used (with three-month T-bill used as the risk-free rate), so that each portfolio can be considered as a zero-investment portfolio.

First, we investigate the optimal number of factors to implement. While there have been many that investigate the appropriate number of factors in PCA/FA (see Connor and Korajczyk, 1993), a widely used test is that of Bai and Ng (2002).[8] Bai and Ng develop a test for factor models with large panels. The loss functions they implement penalize in both the time and cross-sectional dimension. They propose three variations of a panel information criterion (IC), which do not depend on the specified maximum number of factors, and three variations of a panel C_p[9] criterion (PC) which implement the Mallows (1973) C_p criterion in a panel setting. Table 1 shows that five of the six criteria suggest k=4 factors for the set of industry portfolios, with the last PC criterion suggesting k=5 factors. Therefore, we will use four factors as our base model (with PCA and C-PCA both having four factors unless otherwise specified).

[Table 1 goes here]

Before examining the results, the breakdown of MSE should be discussed further. MSE in this context can be broken down into two parts: the variance of the error term and the bias squared (average error squared). The simple summation of the two gives MSE; this implies putting an equal weight on each, however. We argue there are situations such as those related to

[8] See also Harvey, Liu, and Zhu (2015) for an examination of the statistical hurdle that a new factor must clear in order to be accepted.

[9] The C_p criterion stands for the Mallows (1973) information criterion for a model with P regressors.

cost of capital, hedging, long-term investments/projects, etc. where an investor would care more about bias. This equates to caring more about on average how far from the target (bias) an investor/practitioner/firm is rather than about the short-term fluctuations from the target (variance). In finance and forecasting especially, most researchers report MSE but do not consider the two components separately. The breakdown is as follows (using the definition of variance):

$$(18) \qquad MSE \equiv \frac{1}{T}\sum_{t=1}^{T}\left[(r_t - \hat{r}_t)^2\right] \equiv \left[\frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{r}_t)\right]^2 + \sigma_\varepsilon^2 \ ,$$

where r is the realized return, $\hat{r}$ is the predicted return, and σ_ε^2 is the variance of the error term $\varepsilon = (r - \hat{r})$. The first term on the right-hand side in (18) is the bias squared, and the last term is the variance term. We will mostly focus on the absolute level of the bias, the square root of the bias squared (in order to give an economic interpretation), and the MSE.

Expanding Window

Panel A of Table 2 shows the in-sample results of PCA, C-PCA, the CAPM, the FF3 model, and the Carhart model. Reported are the MSE (multiplied by 10,000) and absolute value of the bias multiplied by 100 to be in percent form (bias is the mean of the thirty portfolios is given for each test statistic), as well as the second pass adjusted R-squared (using mean returns). Note that this implementation of C-PCA does not use OLS alphas; therefore, the 2nd pass fit will not be 100%. We know two things will be true of the in-sample results: first, PCA will have the lowest possible MSE. Second, C-PCA with $\gamma = 0$ will have zero bias (with a much better second pass fit).[10] Kan, Robotti, and Shanken (2013) argue that simple comparisons of second-pass R-

[10] As discussed in Section 3, the alphas and betas we use in our main approach differ from the traditional OLS betas and alphas. Therefore, these C-PCA factors will not achieve a perfect second-pass fit, as they are designed to achieve zero error when applying the weights and betas (which are the same) directly from the approach, rather than using the weights as regressors in a second-pass regression. An alternative C-PCA methodology achieves zero OLS

squared values are not sufficient or may induce misleading conclusions.[11] The MSEs of C-PCA and PCA are not significantly different, however. The interpretation is that an investor would only have to sacrifice a bit of MSE (due to variance) in order to achieve zero bias. Also, the panel shows that the traditional asset pricing models perform much worse in all areas.

[Table 2 goes here]

Since we know that PCA will have the lowest MSE and variance while C-PCA will have zero bias by construction, we now turn to the out-of-sample results for a more meaningful comparison. Panel B shows the results when out-of-sample estimation begins in January 1977, with an anchored starting point for the estimation window of January 1927 so that the estimation window begins at 50 years. Recursive monthly estimation is performed, with an expanding estimation window. The out-of-sample analysis here can be thought of as a hedging approach. The portfolio weights and betas (for PCA and C-PCA, the weights and betas are the same) are applied one month ahead to the portfolio realizations. Thus, the portfolio betas are applied to the factor realizations to provide a predicted return for each portfolio. While realizations are used (rather than a pure forecasting approach), this strategy is still implementable by an investor in real time. The portfolio weights can be found prior to the upcoming month via the in-sample estimation. The investor would then hold the portfolios with the given weights in order to form the factors for the next month. Then, the in-sample betas are used as factor weights. If the model holds perfectly out-of-sample, then each portfolio would be hedged against systematic risk perfectly out-of-sample.

---

alphas and a perfect second-pass fit. The empirical results of this alternative approach are available from the authors upon request.

[11] Kan, Robotti, and Shanken (2013) examine R-squared differences when the models have beta misspecification. However, C-PCA would not apply here as a perfect fit is imposed.

Panel B shows the MSE and the absolute value of the bias for each model out of sample. Again, PCA has a slightly lower MSE compared to C-PCA, while the other models perform much worse. Whereas C-PCA obviously no longer has a zero bias, it is still much lower than PCA, with a roughly 30% reduction in bias. This translates to an average error of 1.8% for PCA annually, versus 1.3% for C-PCA. If an investor wished to eliminate the systematic risk of these industry portfolios for any given application, she would have a slightly higher MSE (and variance) with C-PCA. This would mean the predicted return has more fluctuation around the realized return for a given month compared to PCA. However, in the long run (from 1977 to 2015), the investor's average error is roughly 0.5% lower annually (which equates to 19% over the sample). We argue that for certain applications and for those more concerned with the long-run performance rather than month-to-month fluctuations, bias would be of more importance. In other words, imposing the APT restrictions allows an investor to be closer to the target *on average*. Those more concerned with less error on average rather than short-term fluctuations would prefer C-PCA to PCA.

Panel B of Table 2 also reports a second pass fit and an error ratio. In this instance, the second pass fit is the adjusted R-squared of a regression of mean returns on mean predicted returns for the thirty portfolios. This gives an idea how well the predicted returns explain returns on average. The C-PCA model again is superior to the competing models. The error ratio is the ratio of the sum of the squared errors (mean returns less mean predicted returns, squared) to squared mean returns.[12] Hence, a lower ratio means lower errors on average, and C-PCA has a ratio much lower than the other models.

Panel C shows the results when only an initial estimation window of 60 months is used, so that hedging starts in January 1932. This not only provides a larger sample, but also allows

[12] Note that this is similar to, but not exactly the same as, 1 – second pass fit.

for comparison to the results that follow where a rolling 60-month window is used. Again, imposing the APT restrictions reduces bias greatly (around 33% on average). C-PCA provides a better fit and also now a slightly lower MSE compared to PCA. However, the overall fit based on second pass fit and error ratio compared to Panel B in this case is worse. Again, the competing models perform much worse compared to PCA and C-PCA.

Bias and Random Error

The MSE may be written as

(19) $$MSE \equiv E(r_t - \hat{\mu}_t)^2 \equiv E[(r_t - \mu_t) + (\mu_t - \hat{\mu}_t)]^2 ,$$

where

(20) $$r_t = \mu_t + \varepsilon_t , \qquad \varepsilon_t \sim IID(0, \sigma_\varepsilon^2) .$$

(21) $$\hat{\mu}_t = \mu_t + \eta_t , \qquad \eta_t \sim IID(b, \sigma_\eta^2) .$$

The bias is given by *b*. Since ε_t is uncorrelated with anything we can write equation (19) as

(22) $$MSE = \sigma_\varepsilon^2 + \sigma_\eta^2 + b^2 .$$

The PCA estimation of $\hat{\mu}_t$ provides the lowest possible *MSE based on the perspective of the observation frequency*. I.e., it effectively minimizes the *MSE* for the one-period ahead forecast since the observation frequency is (by definition) one period, a month for our data. We may, however, be interested in a *T*-period forecast horizon. In that case, given the IID assumptions, the *MSE* becomes:

(23) $$MSE(T) = T(\sigma_\varepsilon^2 + \sigma_\eta^2) + (Tb)^2 .$$

Thus, if we define the fraction of MSE due to bias as $\lambda \equiv b^2 / (\sigma_\varepsilon^2 + \sigma_\eta^2)$, then the *MSE* per unit of the forecast horizon is:

$$(24) \qquad MSE(T)/T = (\sigma_\varepsilon^2 + \sigma_\eta^2)(1 + \lambda T).$$

It is constant if the forecast is not biased and otherwise, whether the bias is positive or negative, increases linearly with the forecast horizon *T*.

Thus, if we compare two forecasts, based on PCA and C-PCA, such that:

$MSE_{PCA}(1) < MSE_{C-PCA}(1)$ and $\lambda_{PCA} > \lambda_{C-PCA}$ (both conditions should hold in our set-up), then there is a critical $T=T^*$ such that for all $T<T^*$ we have $MSE_{PCA}(T) < MSE_{C-PCA}(T)$ and for all $T>T^*$ we have $MSE_{PCA}(T) > MSE_{C-PCA}(T)$.

Varying time horizon

While it is possible to have an asymmetric loss function that does not use MSE as the loss function and therefore prefer less bias to variance, the time horizon of the investment or forecast matters as well. Specifically, we argue (and show empirically) that as the out-of-sample time horizon increases, bias matters proportionately more compared to the variance term. This section performs the same out-of-sample risk hedging of the previous sections, but varies the out-of-sample time horizon. The results from a one-month horizon are shown again for comparison purposes, but now 3, 12, and 48-month horizons are added. The out-of-sample estimation starts after 60 months, so that an initial estimation window of 60 months is implemented, expanding to add the most recent month. Again, the sample is January 1927 to December 2015, with the out-of-sample window starting in January 1932 and ending in December 2015. Monthly returns are still used to find the loadings for risk hedging purposes. Table 3 shows the results for the C-PCA and PCA methods.

Table 3 provides the MSE, root MSE (RMSE), bias, bias squared, and variance and standard deviation of the error term. Bias, standard deviation, and RMSE are multiplied by 100 and reported in percent form. MSE, bias squared, and variance are multiplied by 10,000 to be consistent. The RMSE and standard deviation are provided in order to give units in percentage terms. Also, the proportion of bias to RMSE is provided as well in order to give an economic magnitude of how much bias contributes to the investor's loss (RMSE).[13] The first thing that can be seen is that extending the horizon for hedging purposes results in C-PCA performing relatively better compared to PCA in terms of MSE. At a one-month horizon, PCA has a MSE that is 0.51% higher than C-PCA. At three months this proportionate increase rises to 1.14%, and it is 3.01% higher at twelve months. However, it does drop to 0.76% at 48 months, perhaps due to lost forecast power. Nonetheless, the less biased C-PCA factors perform relatively better than PCA as the time horizon increases, even when using MSE as the loss function.

[Table 3 goes here]

The results in Table 3 also show that the bias increase for both models is relatively linear and proportionate to the horizon. Going from one month to three months results in a bias that is very close to three times larger (and twelve times larger for twelve months). The variance terms are also close to linear increases as the horizon increases, albeit with slightly larger increases compared to bias. However, bias squared is what matters for MSE, so the increase from one month to three months is actually times nine, and times 144 for twelve months (assuming exactly

[13] It should be noted that bias over RMSE and standard deviation over RMSE do not sum to one. The reason is that there are covariance terms that are ignored. However, the ratio of bias to RMSE does provide a better economic interpretation than bias squared over MSE, since the units (percent returns) are unchanged. It should also be pointed that while equations (17) – (24) assume that the covariance terms are zero this is not the case empirically, as the covariance terms are noticeable.

linear increases, which breaks down a bit as the horizon approaches 12 and 48 months). Therefore, bias has a larger impact empirically as the horizon increases.

In order to relate to equation (24), we can also examine MSE/T. For C-PCA the MSE/T is 11.2, 12.5, 16, and 29.4 at 1, 3, 12, and 48 months, respectively. For PCA it is 11.3, 12.7, 16.5, and 29.7. The pattern is clear in going from one to three to twelve months: MSE/T is getting progressively larger for PCA compared to C-PCA. Thus, it would seem that $\lambda_{PCA} > \lambda_{C-PCA}$, as expected from equations (19) to (24). At the longest horizon of 48 months, we can infer (roughly) the lambdas. For C-PCA the lambda is roughly 1.9% and for PCA it is roughly 3.4%. The difference in lambdas between the two models is approximately the same for the other horizons as well, and the lambdas are what should be expected based on the bias for each model. We know that the larger lambda for PCA will make it perform worse as the horizon grows.

Note that this effect would be amplified further if the investor's loss function was asymmetrically shifted towards bias. Table 3 shows how a slightly higher weight on bias squared can make a large difference. This can be seen by comparing the magnitudes of the bias squared and variance terms (which together compose MSE). Bias squared is of an extremely small magnitude compared to variance, so placing a weight even slightly above an equal weight would dramatically change the resulting loss. The ratio of bias to RMSE can be examined to obtain the economic magnitude of how much more bias matters with a longer horizon. At one month, bias only makes up around 2.4% to 3.6% of RMSE depending on the model (and ignoring covariance terms in RMSE). However, this increases to 4.1%-5.8% at three months and 8%-10.9% at twelve months. At the longer-term horizon of 48 months, bias makes up as much as 18.5% of RMSE. It is expected that this would increase further as the horizon grows beyond four years (which we find in unreported results).

Rolling Window

The same out-of-sample analysis is performed, but with a rolling 60-month window rather than an expanding window with an anchored starting point. This is done both for robustness and because many researchers use a rolling 60-month window. Table 4 shows that while both PCA and C-PCA easily beat the traditional asset pricing models, C-PCA has the lowest MSE. However, it is only slightly lower and PCA has a slightly lower bias. In comparing Table 4 to Panel C (so that the sample period is the same), it can be seen that using a rolling window results in lower MSE for all models. However, for C-PCA the bias is now slightly higher. It appears that a smaller, rolling window allows the model to capture the local variation, resulting in less variance and lower MSE. However, if one would rather have a lower bias, then a larger, expanding window would be the proper choice. Also, the benefits of imposing the APT restrictions are greater when using a larger estimation window. In both Panels B and C of Table 2, C-PCA is able to drastically reduce bias (while having either a slightly higher or lower MSE, depending on the sample). However, with the rolling window the bias is not reduced (although MSE is still slightly lower). Therefore, if an investor/practitioner is interested in obtaining a lower bias (perhaps due to their asymmetric loss function or time horizon), then she should use an expanding window.

[Table 4 goes here]

Using in-sample betas

Another test is performed to see how well the factors of each model explain future returns. Here the portfolio weights are applied to the next month's realized portfolio returns (so that the factor realization is stored each month based on information up to the previous month). Where the methodology differs is that now instead of applying the previously estimated betas, a full-sample regression (similar to following the Black, Jensen, and Scholes, 1972 approach) is run of portfolio returns on factor returns.[14] The sample here is the previous out-of-sample estimation period (1977 – 2015 for the expanding window). While this is not a true out-of-sample test in that it cannot be applied in real-time by a potential investor, it still gives an idea of how well these portfolio weights (and consequently factors) explain future returns. Additionally, the betas are now allowed to be constructed optimally based on the realized returns.

Table 5 shows the regression results for the 1977 – 2015 sample, where the expanding window is used. The results with and without a constant (alpha) are shown for both PCA and C-PCA. Note that these are averages of the thirty time-series regressions. Now that we allow each portfolio to optimize its betas based on the factor realizations, the C-PCA appears to better explain returns. The R-squared values are higher than those for PCA factors, and the MSE is lower. Further, when allowing for a non-zero alpha, the C-PCA approach produces an alpha that is 4 basis points lower per month. It can also be seen that only the first factor for each model is priced, although all four help explain covariance risk.

[Table 5 goes here]

The same analysis is repeated with the rolling window, where now the sample period is 1932 – 2015. Table 6 shows that while the MSE is slightly higher and the R-squared is slightly

[14] We also perform the traditional Fama and MacBeth (1973) test on all models, but the results did not vary much across models and are therefore omitted, but available upon request.

lower for C-PCA, it does produce an average alpha that is 10 basis points lower per month. Again, it appears that when imposing the APT restrictions, it is better to use a larger window. PCA is slightly better when implementing a rolling 60-month window, consistent with the findings in Table 4. The short estimation window may not allow for sufficient information in generating the weights, which should hold one-month-ahead out of sample if expected returns are efficiently estimated.

[Table 6 goes here]

5. Conclusion

In this paper we derive an approach in which the APT result of zero pricing error is imposed on principal components analysis (PCA). We also show that PCA can be represented as an approach that minimizes the sum of squared errors of a panel (or maximizing first pass time series fit). Thus, our approach imposes a perfect second pass fit (or up to a specified fraction of mispricing) while simultaneously maximizing first-pass fit. While PCA has the lowest possible mean squared error (MSE), we examine the breakdown of MSE into variance and bias (squared) terms. In sample, PCA will have the lowest MSE (and variance) of any model, while C-PCA (the model imposing APT restrictions) will have the lowest bias if mispricing is set to zero. We argue that for most applications of asset pricing models (cost of capital, forecasting, hedging, etc.) the practitioner may prefer lower bias rather than variance/MSE. Relatedly, any practitioner with a more long-term horizon would most likely prefer a model that is closer to the target on average (less bias), rather than being concerned with period-to-period fluctuations around the target (variance). We also show both analytically and empirically that a model with a lower bias will perform relatively better in terms of MSE as the investment horizon grows.

Due to the in-sample properties of the two models, an out-of-sample test is needed in order to properly compare the two. We implement a hedging approach based on portfolio realizations, where the weights and betas based on previous in-sample estimation are applied to the realized portfolio returns in the following month. We use the thirty Fama-French industry portfolios due to their weak factor structure in order to give a stronger test. With in-sample estimation starting (and anchored) in 1927 and out-of-sample estimation starting in 1977 and ending in 2015, we show that bias is reduced by almost one-third when imposing the APT restrictions while MSE is only slightly higher (not significantly different). Both in sample and out of sample, an investor or practitioner could adopt our approach and greatly reduce bias while only sacrificing a bit of MSE. We also show that allowing betas to be optimized in-sample (with factor weights applied out-of-sample) results in C-PCA better explaining returns in terms of both bias and MSE. For robustness purposes we also examine a rolling 60-month window. However, it appears that a larger sample is needed to properly estimate the factors. We also show that PCA and C-PCA perform significantly better than the traditional asset pricing models (CAPM, Fama-French three-factor model, and Carhart model).

It is argued that in order to properly implement the APT model, the model restrictions should be imposed a priori. Doing so results in lower bias out-of-sample, which would be desirable in certain applications. The methodology is derived herein, with possible extensions examined as well. This allows for future research to implement our approach for various applications. We implement the model to a portfolio hedged against systematic risk so that there is no need to estimate the risk premia, and show that bias is indeed reduced greatly. Future research in this area should also consider the breakdown of MSE and whether bias or variance is more important in the appropriate context.

References

Bai, J., and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica, 70*(1), 191-221.

Bartholdy, J., and Peare, P. (2005). Estimation of expected return: CAPM vs. Fama and French. *International Review of Financial Analysis, 14*(4), 407-427.

Basu, S., and Markov, S. (2004). Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics*, *38*, 171-203.

Best, M. J., and Grauer, R. R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *Rev. Financ. Stud., 4*(2), 315-342.

Bhatia, R. (2013). *Matrix analysis* (Vol. 169). Springer Science and Business Media.

Black, F., Jensen, M. C., and Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests.

Bruner, R. F., Eades, K. M., Harris, R. S., and Higgins, R. C. (1998). Best practices in estimating the cost of capital: survey and synthesis. *Financial Practice and Education, 8*, 13-28.

Bussière, M., Hoerova, M., and Klaus, B. (2015). Commonality in hedge fund returns: Driving factors and implications. *Journal of Banking and Finance, 54*, 266-280.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance, 52*(1), 57-82.

Chen, N.-F., Roll, R., and Ross, S. A. (1986). Economic Forces and the Stock Market. *The Journal of Business, 59*(3), 383-403.

Chen, S.-J., Hsieh, C.-H., and Jordan, B. D. (1997). Real Estate and the Arbitrage Pricing Theory: Macrovariables vs. Derived Factors. *Real Estate Economics, 25*(3), 506-523.

Chou, P.-H., Ho, P.-H., and Ko, K.-C. (2012). Do industries matter in explaining stock returns and asset-pricing anomalies? *Journal of Banking and Finance, 36*(2), 355-370.

Clatworthy, M. A., Peel, D. A., and Pope, P. F. (2012). Are analysts' loss functions asymmetric? *Journal of Forecasting*, *31*(8), 736-756.

Connor, G., and Korajczyk, R. A. (1993). A Test for the Number of Factors in an Approximate Factor Model. *The Journal of Finance, 48*(4), 1263-1291.

Cummins, J. D., and Phillips, R. D. (2005). Estimating the Cost of Equity Capital for Property-Liability Insurers. *Journal of Risk and Insurance, 72*(3), 441-478.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/ N Portfolio Strategy? *Rev. Financ. Stud., 22*(5), 1915-1953.

Diebold, F. X., and Mariano, R. S. (2002). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics, 20*(1), 134-144.

Easton, P. D. (2004). PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital. *The Accounting Review, 79*(1), 73-95.

Easton, P. D., and Monahan, S. J. (2005). An Evaluation of Accounting-Based Measures of Expected Returns. *The Accounting Review, 80*(2), 501-538.

Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, *6*(1), 122-157.

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

Fama, E. F., and French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance, 51*(1), 55-84.

Fama, E. F., and French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics, 43*(2), 153-193.

Fama, E. F., and MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*(3), 607-636.

Fan, K. (1949). On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations I. *Proceedings of the National Academy of Sciences, 35*(11), 652-655.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *Review of Economics and Statistics, 82*(4), 540-554.

Fuller, R. J., and Kerr, H. S. (1981). Estimating the Divisional Cost of Capital: An Analysis of the Pure-Play Technique. *The Journal of Finance, 36*(5), 997-1009.

Giacomini, R., and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica, 74*(6), 1545-1578.

Goldenberg, D. H., and Robin, A. J. (1991). The Arbitrage Pricing Theory and Cost-Of-Capital Estimation: The Case of Electric Utilities. *Journal of Financial Research, 14*(3), 181-196.

Goyal, A., Perignon, C., and Villa, C. (2008). How common are common return factors across the NYSE and Nasdaq?. *Journal of Financial Economics, 90*(3), 252-271.

Grinblatt, M., and Titman, S. (1987). The Relation Between Mean-Variance Efficiency and Arbitrage Pricing. *The Journal of Business, 60*(1), 97-112.

Harvey, C. R., Liu, Y., and Zhu, H. (2015). … and the Cross-Section of Expected Returns. *Rev. Financ. Stud., 29*(1), 5-68.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics, 119*(2), 249-283.

Hou, K., van Dijk, M. A., and Zhang, Y. (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics, 53*(3), 504-526.

Huberman, G., Kandel, S., and Stambaugh, R. F. (1987). Mimicking Portfolios and Exact Arbitrage Pricing. *The Journal of Finance, 42*(1), 1-9.

Jolliffe, I. (2002). *Principal component analysis*. John Wiley and Sons, Ltd.

Jong, J.-C., and Kotz, S. (1999). On a Relation between Principal Components and Regression Analysis. *The American Statistician, 53*(4), 349-351.

Kan, R., Robotti, C., and Shanken, J. (2013). Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *The Journal of Finance*, *68*(6), 2617-2649.

Kan, R., and Smith, D. R. (2008). The Distribution of the Sample Minimum-Variance Frontier. *Management Science, 54*(7), 1364-1380.

Kan, R., Wang, X., and Zhou, G. (2016) Optimal Portfolio Selection with and without Risk-Free Asset. *SSRN Working Paper*.

Kan, R., and Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis, 42*(3), 621-656.

Kavussanos, M., Marcoulis, S., and Arkoulis, A. (2002). Macroeconomic factors and international industry returns. *RAFE, 12*(12), 923-931.

Koedijk, K. G., and van Dijk, M. A. (2004). Global Risk Factors and the Cost of Capital. *Financial Analysts Journal, 60*(2), 32-38.

Kozak, S., Nagel, S., and Santosh, S. (2014). Interpreting factor models. Working Paper

Lee, C., Ng, D., and Swaminathan, B. (2009). Testing International Asset Pricing Models Using Implied Costs of Capital. *Journal of Financial and Quantitative Analysis, 44*(2), 307-335.

Lehmann, B. N., and Modest, D. M. (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics, 21*(2), 213-254.

Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review*, 81(3), 580-590.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics, 96*(2), 175-194.

Li, R.-C. (2015). Rayleigh Quotient Based Optimization Methods for Eigenvalue Problems. *Matrix Functions and Matrix Equations, 19*, 76-108.

Li, Y., Ng, D. T., and Swaminathan, B. (2013). Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics, 110*(2), 419-436.

Litzenberger, R., Ramaswamy, K., and Sosin, H. (1980). On the CAPM Approach to the Estimation of A Public Utility's Cost of Equity Capital. *The Journal of Finance, 35*(2), 369-383.

Ludvigson, S. C., and Ng, S. (2009). Macro Factors in Bond Risk Premia. *Rev. Financ. Stud., 22*(12), 5027-5067.

MacKinlay, C.A., and Pástor, Ľ. (2000). Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Rev. Financ. Stud., 13*(4), 883-916.

Mallows, C. L. (1973). Some Comments on C p. *Technometrics, 15*(4), 661-675.

Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal, 45*(1), 31-42.

Nardari, F., and Scruggs, J. T. (2007). Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions. *Journal of Financial and Quantitative Analysis, 42*(4), 857-891.

Patton, A. J., and Timmermann, A. (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, *140*(2), 884-918.

Pratt, S. P., and Grabowski, R. J. (2008). *Cost of capital*. John Wiley and Sons.

Pukthuanthong, K., and Roll, R. (2014) A Protocol for Factor Identification. *SSRN Working Paper*.

Pástor, Ľ., Sinha, M., and Swaminathan, B. (2008). Estimating the Intertemporal Risk-Return Tradeoff Using the Implied Cost of Capital. *The Journal of Finance, 63*(6), 2859-2897.

Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics, 4*(2), 129-176.

Roll, R., and Ross, S. A. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. *The Journal of Finance, 35*(5), 1073-1103.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory, 13*(3), 341-360.

Simin, T. (2008). The Poor Predictive Performance of Asset Pricing Models. *Journal of Financial and Quantitative Analysis, 43*(2), 355-380.

Stock, J. H., and Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association, 97*(460), 1167-1179.

Table 1: Bai and Ng (2002) results

Results are shown for the Bai and Ng (2002) criteria, using the 30 Fama-French industry portfolios with a sample of January 1927 - December 2015. The top panel shows the loss function results for each criterion and k factors from one to five. The bottom panel shows the optimum number of factors based on each criterion. The three different panel information criteria (IC) and panel C_p criteria (PC) are shown, and k* gives the optimal number of factors which gives the lowest loss function. See Bai and Ng (2002) for details.

k=	IC1	IC2	IC3	PC1	PC2	PC3
0	-5.2909	-5.2909	-5.2909	0.00504	0.00504	0.00504
1	-6.2904	-6.2894	-6.2926	0.00176	0.00176	0.00176
2	-6.3479	-6.346	-6.3524	0.00161	0.00161	0.00161
3	-6.3713	-6.3685	-6.378	0.00154	0.00155	0.00154
4	**-6.3733**	**-6.3695**	**-6.3823**	**0.00152**	**0.00152**	0.00151
5	-6.3664	-6.3616	-6.3775	0.00152	0.00153	**0.00151**

	k*
IC1	4
IC2	4
IC3	4
PC1	4
PC2	4
PC3	5

TABLE 2: In-Sample and Out-Of-Sample Performance

Panel A shows the full, in-sample results for the 30 Fama-French industry portfolios for the various models/approaches. The sample is January 1927 - December 2015. Panel B shows the one-month-ahead out-of-sample results with an anchored starting point and initial 50-year estimation window. The sample is January 1927 - December 2015, with out-of-sample hedging starting in January 1977 and ending in December 2015. Panel C shows the same OOS results, but with an initial 60-month window, with hedging starting in January 1932, ending again in December 2015. All panels show the mean squared error (MSE) and bias (absolute value of the average pricing error), averaged across all 30 portfolios. Bias is multiplied by 100 and reported in monthly percent form. MSE is multiplied by 10,000 to be consistent. The error term comes from directly applying weights and betas calculated from the particular approach to give a predicted return for each month (error is difference between realized return and predicted return). Note that MSE and Bias is not calculated via regression. The 2nd pass fit in Panel A is the adjusted R-squared value of a regression of mean returns regressed on 1st pass coefficients. The 2nd pass fit in Panels B and C is the adjusted R-squared value of a regression of mean returns regressed on mean predicted returns. Error ratio is the sum of squared errors (mean returns less mean predicted returns, squared) over the sum of squared mean returns, multiplied by 100. PCA is principal components analysis, C-PCA is the constrained PCA introduced in this paper, CAPM is capital asset pricing model, FF3 is the Fama-French three-factor model, and Carhart adds the momentum factor to the FF3 model. Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

Panel A: In-Sample Results

	PCA	C-PCA	CAPM	FF3	Carhart
MSE	10.75	11.09	18.03	16.52	16.40
Bias	0.11	0	0.14	0.18	0.15
2nd pass fit	-0.78	0.41	-1.51	-0.78	-0.42

Panel B: Out-Of-Sample Results Starting in 1977

	PCA	C-PCA	CAPM	FF3	Carhart
MSE	11.35	11.45	18.30	17.72	17.72
Bias	0.15	0.11	0.19	0.21	0.19
2nd pass fit	-0.09	0.44	-0.90	-1.14	-0.77
error ratio	8.15	4.20	14.24	16.06	13.20

Panel C: Out-Of-Sample Results Starting In 1932

	PCA	C-PCA	CAPM	FF3	Carhart
MSE	11.29	11.23	17.71	16.76	16.85
Bias	0.12	0.08	0.15	0.17	0.14

2nd pass fit	-0.82	0.00	-2.07	-2.93	-1.66
error ratio	3.27	1.79	5.74	7.07	4.87

Table 3: Varying Time Horizon Out of Sample

This table shows the mean squared error (MSE), root mean squared error (RMSE), bias/absolute value of the average error, bias squared (bias sq), variance of the error (var), and standard deviation of the error (std dev) for the constrained principal components analysis (C-PCA) and principal components analysis (PCA) models with the 30 Fama/French industry portfolios as the test assets. Also shown is the proportion of bias to RMSE (multiplied by 100) in order to give an economic interpretation of how much bias contributes to the loss. The sample is January 1927 - December 2015 (out-of-sample estimation ending at December 2015). Bias, standard deviation, and RMSE are multiplied by 100 and reported in percent form. MSE, bias squared, and variance are multiplied by 10,000 to be consistent. Bias is reported as monthly returns. The error term comes from directly applying weights and betas calculated from the particular approach to give a predicted return for each month (error is difference between realized return and predicted return). Note that MSE and Bias is not calculated via regression. Each panel shows the results when varying the forecast/hedging horizon from 1 month to 3, 12, and 48 months. The in-sample estimation window starts at January 1927, and out-of-sample estimation starts at January 1932 giving an initial estimation window of 60 months, which is expanding. Monthly returns are used to find the loadings in all cases.

1-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	11.23	3.35	0.08	0.01	11.22	3.35	2.39
PCA	11.29	3.36	0.12	0.01	11.28	3.36	3.57

3-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	37.59	6.13	0.25	0.06	37.53	6.13	4.08
PCA	38.02	6.17	0.36	0.13	37.89	6.16	5.84

12-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	192.18	13.86	1.11	1.23	190.95	13.82	8.01
PCA	197.97	14.07	1.53	2.34	195.63	13.99	10.87

48-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	1413.42	37.60	5.20	27.04	1386.38	37.23	13.83
PCA	1424.19	37.74	6.97	48.58	1375.61	37.09	18.47

TABLE 4: Out-Of-Sample Performance with a Rolling Window

This table shows the one-month-ahead out-of-sample results when using a rolling 60-month window for the 30 Fama-French industry portfolios for the various models/approaches. The sample is January 1927 - December 2015, with hedging starting in January 1932 and ending in December 2015. The above results show the mean squared error multiplied by 10,000 (MSE) and bias (absolute value of the average pricing error) multiplied by 100, averaged across all 30 portfolios. The error term comes from directly applying weights and betas calculated from the particular approach's in-sample estimation to give a one-month-ahead predicted return for each month (error is difference between realized return and predicted return). Note that MSE and Bias is not calculated via regression. 2nd pass fit is the adjusted R-squared value of a regression of mean returns regressed on mean predicted returns. Error ratio is the sum of squared errors (mean returns less mean predicted returns, squared) to the sum of squared mean returns (reported as multiplied by 100). PCA is principal components analysis, C-PCA is the constrained PCA introduced in this paper, CAPM is capital asset pricing model, FF3 is the Fama-French 3-factor model, and Carhart adds the momentum factor to the FF3 model. Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

	PCA	C-PCA	CAPM	FF3	Carhart
MSE	10.12	10.05	17.04	15.90	16.15
Bias	0.09	0.10	0.12	0.16	0.13
2nd pass fit	-0.01	-0.05	-0.89	-2.14	-1.30
error ratio	1.82	1.91	4.02	5.75	4.19

Table 5: Out-Of-Sample Factor Weights with In-Sample Betas

This table shows the results the regression results of returns on one-month-ahead factor returns. Reported are the mean for the 30 Fama-French industry portfolios. To form the one-month-ahead factor realizations, the factor weights are found in-sample and the one-month-ahead factor realization is stored based on those weights. The in-sample estimation period for the factor weights starts in January 1927, utilizing an expanding window with an anchored starting point. The regression sample is January 1977 to December 2015. F1 refers to the first factor, F2 the second, etc. Reported are coefficients and t-stats in parentheses. Also reported are the adjusted R-squared values and mean-squared error of the regression, multiplied by 10,000 (mean of the 30 portfolios). Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

	PCA				C-PCA			
constant	0.0018	(0.23)			0.0014	(0.11)		
F1	0.1810	(27.27)	0.1812	(27.58)	0.1817	(24.09)	0.1818	(25.05)
F2	0.0092	(0.91)	0.0095	(0.92)	-0.0081	(0.32)	-0.0070	(0.37)
F3	0.0052	(0.13)	0.0055	(0.14)	-0.0085	(-1.11)	-0.0087	(-1.12)
F4	0.0099	(0.36)	0.0103	(0.38)	-0.0015	(-0.50)	-0.0009	(-0.44)
adj. R-squared	0.63		0.63		0.68		0.68	
MSE	14.12		14.18		11.89		11.86	

Table 6: Out-Of-Sample Factor Weights with In-Sample Betas (Rolling Window)

This table shows the results the regression results of returns on one-month-ahead factor returns. Reported are the mean for the 30 Fama-French industry portfolios. To form the one-month-ahead factor realizations, the factor weights are found in-sample and the one-month-ahead factor realization is stored based on those weights. The in-sample estimation period for the factor weights starts in January 1927, utilizing a rolling 60-month window. The regression sample is January 1932 to December 2015. F1 refers to the first factor, F2 the second, etc. Reported are coefficients and t-stats in parentheses. Also reported are the adjusted R-squared values and mean-squared error of the regression, multiplied by 10,000 (mean of the 30 portfolios). Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

	PCA				C-PCA			
constant	0.0081	(4.01)			0.0071	(3.50)		
F1	-0.0536	(-8.03)	-0.0532	(-7.91)	0.0191	(1.81)	0.0227	(2.11)
F2	0.0109	(0.50)	0.0123	(0.56)	-0.0452	(-2.77)	-0.0422	(-2.55)
F3	-0.0286	(-0.86)	-0.0362	(-1.11)	-0.0699	(-2.54)	-0.0729	(-2.70)
F4	-0.0851	(-2.33)	-0.0923	(-2.53)	-0.0348	(-0.98)	-0.0376	(-1.06)
adj. R-squared	0.07		0.05		0.06		0.04	
MSE	43.64		44.30		44.25		44.76	

Appendix A

```
%===============================================================
MATLAB sample program designed to optimally choose k factors (portfolios
chosen from among n zero-investment assets) to minimize the unexplained
'first-pass' errors subject to a targeted fraction of 'second pass' pricing
errors (astar). Output provides Q as k vectors of portfolio weights
normalized to square to one. Further a and s provide the fractions of pricing
errors and unexplained variance, respectively, whereas a0 and s0 provide the
benchmark, unrestricted fractions of these.
===============================================================

clear all; clc;

%User Inputs k and astar:
k=2; %chosen number of factors; must be less than n
astar=0.00;  %chosen fraction of pricing errors (sum of squared alphas
relative to sum of squared mean returns).
%astar should logically be between 0 and 1, but also less than a0, the
implied a for lambda=0 for the optimization to make sense (otherwise the
constraint is not binding).

load FF25.txt;
FF=FF25(:,2:26);  %Fama-French excess returns for 25 size and value sorted
portfolios starting July 1963 ending June 2005
[t,n]=size(FF);

Sigma=cov(FF); %covariance matrix for the N asset returns
mu=mean(FF)'; %vector of mean returns over the T periods of the N assets
lambda=0; %lambda is the lagrangian multiplier constraint for the pricing
errors constraint.  Initial value is zero.
eps=0.0000000001; %tolerance level for deviations from the target fraction of
squared pricing errors.
diff=-5; %the experimental change in the lagrangian multiplier
a=0.10; %initial level for the fraction of pricing errors
its=0; %iterations counter
CHECK=[]; %storing results for all iterations

%Loop to converge to true lambda (and Q) contingent on choice of astar
while (a-astar)^2>eps && its<1000; %continue revisions until pricing errors
are within tolerance or the number of iterations becomes too large
V=Sigma+lambda*mu*mu'; %V is the matrix for which eigenvectors are found
[Q,Delta]=eigs(V,k); %'eigs' easier than 'eig' since it provides Qk
abar=a; %abar accounts for the lagged iteration of a
a=(mu'*mu - (mu'*Q*Q'*mu))/(mu'*mu); %defines fraction of pricing errors
s=1-trace(Q'*Sigma*Q)/trace(Sigma); %defines fraction of unexplained variance
if (a-astar)*(a-abar)>0   %if a worsens reverse the change in lambda
    diff=-diff/2; %reverse change in lambda but at half the pace
end
lambda=lambda+diff;  %update lambda in the direction that brings pricing
errors closer to target
its=its+1;
CHECK=[CHECK;[lambda a s]];
end
```

```
lambda=lambda-diff %corrected for the unnecessary change in lambda in the
final iteration.
%Note that lambda must be positive. If not, change the sign or lower the
value of the initial 'diff' choice.
lamcheck=trace((Delta-Q'*Sigma*Q)/((1-a)*mu'*mu))  %check that lambda is
indeed the equilibrium value

[Q0, Delta0]=eigs(Sigma,k); %find eigenvectors for unconstrained case
a0=(mu'*mu - (mu'*Q0*Q0'*mu))/(mu'*mu); %provides fraction pricing errors in
unconstrained case.
s0=1-trace(Q0'*Sigma*Q0)/trace(Sigma);
%Check to make sure that a0>astar

Q,[a,a0;s, s0] %factor weights (Q) plus fractions of pricing errors and
unexplained variance in optimized case (column 1) and benchmark case (column
2).
```

Appendix B

Consider the following optimization problem:

$$\text{(B1)} \qquad \underset{\mathbf{Q},\lambda}{Min} \quad \left[E(\mathbf{e}'\mathbf{e}) + \lambda(\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} - \boldsymbol{\gamma}\,\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})\right],$$

where $0 \le \gamma < 1$ represents the target level for the sum of squared alphas relative to the sum of squared mean returns. Using equations (1) and (2), the properties of traces, and taking expectations

$$\text{(B2)} \quad \underset{\mathbf{Q},\lambda}{Min} \quad Tr\left[(\mathbf{I} - \mathbf{BQ}')\,\Sigma\,(\mathbf{I} - \mathbf{QB}')\right]$$

$$+\lambda\left[\mu'(\mathbf{I} - \mathbf{BQ}')\,\Sigma^{-1}\,(\mathbf{I} - \mathbf{QB}')\mu - \gamma\,\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\right]$$

The $\mathbf{B}$ are set exogenously equal to the standard OLS betas of equation (4):

$$\text{(B3)} \qquad \mathbf{B} = \boldsymbol{\Sigma}\mathbf{Q}\,[\mathbf{Q}'\,\boldsymbol{\Sigma}\mathbf{Q}]^{-1},$$

Using the definition of $\mathbf{B}$ and the properties of traces in (4.B2), dropping constants and multiplying by -1 gives

$$\text{(B4)} \qquad \underset{\mathbf{Q}}{Max} \quad Tr\left([\mathbf{Q}'\boldsymbol{\Sigma}^2\mathbf{Q}][\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q}]^{-1} + \lambda\boldsymbol{\mu}\boldsymbol{\mu}'[\mathbf{Q}(\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})^{-1}\mathbf{Q}']\right),$$

and in addition λ is set to force $\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} = \boldsymbol{\gamma}\,\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. Using Trace properties:

$$\text{(B5)} \quad \underset{\mathbf{Q}}{Max} \quad Tr\left([\mathbf{Q}'(\boldsymbol{\Sigma}^2 + \lambda\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{Q}][\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q}]^{-1}\right).$$

This is equivalent to

(B6) $\underset{\mathbf{X}}{Max} \quad Tr\left([\mathbf{X}'(\mathbf{\Sigma}+\lambda\mathbf{mm}')\mathbf{X}]+\mathbf{\Delta}(\mathbf{I}-\mathbf{X}'\mathbf{X})\right),$

where $\mathbf{X}=\mathbf{\Sigma}^{1/2}\mathbf{Q}$, $\mathbf{m}=\mathbf{\Sigma}^{-1/2}\boldsymbol{\mu}$ and $\mathbf{X}'\mathbf{X}=\mathbf{Q}'\mathbf{\Sigma}\mathbf{Q}=\mathbf{I}$ and $\mathbf{\Delta}$ is the multiplier enforcing this constraint. Then the first-order conditions are

(B7) $(\mathbf{\Sigma}+\lambda\,\mathbf{mm}')\mathbf{X}=\mathbf{X}\mathbf{\Delta}$.

(B8) $\mathbf{V}(\lambda)\,\mathbf{X}=(\mathbf{\Sigma}+\lambda\,\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{\Sigma}^{-\frac{1}{2}})\mathbf{X}=\mathbf{X}\mathbf{\Delta}$

Pre-multiplying (4.B7) by $\mathbf{X}'$ gives

(B9) $\mathbf{\Delta}=\mathbf{X}'\mathbf{\Sigma}\mathbf{X}+\lambda\,\mathbf{X}'\mathbf{mm}'\mathbf{X}$

and subsequently pre- and post-multiplying both sides by $\mu_k=\boldsymbol{\mu}'\mathbf{Q}=\mathbf{m}'\mathbf{X}$ and its transpose, respectively, and using the normalization that $\mathbf{X}'\mathbf{X}=\mathbf{I}$ implies for the optimal choice of $\mathbf{X}$ that $\boldsymbol{\mu}_k'(\mathbf{\Delta}-\mathbf{X}'\mathbf{\Sigma}\mathbf{X})\boldsymbol{\mu}_k=\lambda(\mathbf{m}'\mathbf{XX}'\mathbf{m})^2$. To impose $\boldsymbol{\alpha}'\mathbf{\Sigma}^{-1}\boldsymbol{\alpha}=\boldsymbol{\gamma}\,\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}=\gamma\mathbf{m}'\mathbf{m}$ eliminate $\boldsymbol{\alpha}$ from equation (2). This produces $\mathbf{m}'\mathbf{XX}'\mathbf{m}=\boldsymbol{\mu}'\mathbf{QQ}'\boldsymbol{\mu}=(1-\gamma)\,\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}=(1-\gamma)\,\mathbf{m}'\mathbf{m}$. Then, $\mathbf{m}'\mathbf{XX}'\mathbf{m}=(1-\gamma)\,\mathbf{m}'\mathbf{m}$ and $\boldsymbol{\mu}_k'(\mathbf{\Delta}-\mathbf{X}'\mathbf{\Sigma}\mathbf{X})\boldsymbol{\mu}_k=\lambda(\mathbf{m}'\mathbf{XX}'\mathbf{m})^2$ jointly imply

(B10) $\lambda=\boldsymbol{\mu}_k'(\mathbf{\Delta}-\mathbf{X}'\mathbf{\Sigma}\mathbf{X})\boldsymbol{\mu}_k\,/(1-\gamma)^2(\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu})^2$.

Thus we use (4.B8) and (4.B10) together with $\mathbf{X}=\mathbf{\Sigma}^{1/2}\mathbf{Q}$, $\mathbf{X}'\mathbf{X}=\mathbf{Q}'\mathbf{\Sigma}\mathbf{Q}=\mathbf{I}$, and $\mu_k=\boldsymbol{\mu}'\mathbf{Q}$ to obtain the appropriate portfolios given OLS betas.