Statistical Risk Models

Zura Kakushadze[§†1] and Willie Yu[♯2]

[§] *Quantigic® Solutions LLC*
1127 High Ridge Road #135, Stamford, CT 06905 [3]
[†] *Free University of Tbilisi, Business School & School of Physics*
240, David Agmashenebeli Alley, Tbilisi, 0159, Georgia
[♯] *Centre for Computational Biology, Duke-NUS Medical School*
8 College Road, Singapore 169857

(February 14, 2016)

Abstract

We give complete algorithms and source code for constructing statistical risk models, including methods for fixing the number of risk factors. One such method is based on eRank (effective rank) and yields results similar to (and further validates) the method of (Kakushadze, 2015d). We also give a complete algorithm and source code for computing eigenvectors and eigenvalues of a sample covariance matrix which requires i) no costly iterations and ii) the number of operations linear in the number of returns. The presentation is intended to be pedagogical and oriented toward practical applications.

[1] Zura Kakushadze, Ph.D., is the President of Quantigic® Solutions LLC, and a Full Professor at Free University of Tbilisi. Email: zura@quantigic.com

[2] Willie Yu, Ph.D., is a Research Fellow at Duke-NUS Medical School. Email: willie.yu@duke-nus.edu.sg

[3] DISCLAIMER: This address is used by the corresponding author for no purpose other than to indicate his professional affiliation as is customary in publications. In particular, the contents of this paper are not intended as an investment, legal, tax or any other such advice, and in no way represent views of Quantigic® Solutions LLC, the website www.quantigic.com or any of their other affiliates.

1 Introduction

Multifactor risk models[4] are a popular risk management tool, e.g., in portfolio optimization. For stock portfolios, in their most popular incarnations, multifactor risk models are usually constructed based on industry and style risk factors.[5] However, in some cases such constructions are unavailable, e.g., because any industry classification (or similar) is lacking, any relevant style factors are impossible to define, etc. In fact, this is generally the case when the underlying returns are not for equities but some other "instruments", e.g., quantitative trading alphas (expected returns).

In such cases one usually resorts to statistical risk models. Often times these are thought of in the context of principal components of a sample covariance (or correlation) matrix of returns. More generally, one can think of statistical risk models as constructed solely based on the time series of the underlying returns and no additional information. The purpose of these notes is to provide a simple and pedagogical discussion of statistical risk models oriented toward practical applications.

In Section 2 we set up our discussion by discussing the sample covariance matrix, generalities of factor models, the requirement that factor models reproduce in-sample variances, and how a K-factor statistical risk model can be simply constructed by starting from the sample covariance (or correlation) matrix, writing down its spectral representation via principal components, truncating the sum by keeping only the first K principal components, and compensating for the deficit in the variances (i.e., on the diagonal of the resultant matrix) by adding specific (idiosyncratic) risk. This (generally)[6] results in a positive-definite (and thus invertible) risk model covariance

[4] For a partial list of related works, see (Acharya and Pedersen, 2005), (Ang *et al*, 2006), (Anson, 2013/14), (Asness, 1995), (Asness and Stevens, 1995), (Asness *et al*, 2001), (Bai, 2003), (Bai and Li, 2012), (Bai and Ng, 2002), (Bansal and Viswanathan, 1993), (Banz, 1981), (Basu, 1977), (Black, 1972), (Black *et al*, 1972), (Blume and Friend, 1973), (Brandt *et al*, 2010), (Briner and Connor, 2008), (Burmeister and Wall, 1986), (Campbell, 1987), (Campbell *et al*, 2001), (Campbell and Shiller, 1988), (Carhart, 1997), (Chamberlain and Rothschild, 1983), (Chan *et al*, 1985), (Chen *et al*, 1986, 1990), (Chicheportiche and Bouchaud, 2015), (Cochrane, 2001), (Connor, 1984, 1995), (Connor and Korajczyk, 1988, 1989, 2010), (Daniel and Titman, 1997), (DeBondt and Thaler, 1985), (Dhrymes *et al*, 1984), (Fama and French, 1992, 1993, 1996, 2015), (Fama and McBeth, 1973), (Ferson and Harvey, 1991, 1999), (Forni *et al*, 2000, 2005), (Forni and Lippi 2001), (Goyal *et al*, 2008), (Goyal and Santa-Clara, 2003), (Grinold and Kahn, 2000), (Hall *et al*, 2002), (Haugen, 1995), (Heaton and Lucas, 1999), (Heston and Rouwenhorst, 1994), (Jagannathan and Wang, 1996), (Jegadeesh and Titman, 1993, 2001), (Kakushadze 2015a,b,c,d, 2016), (Kakushadze and Liew, 2015), (Kakushadze and Yu, 2016), (King, 1966), (Korajczyk and Sadka, 2008), (Kothari and Shanken, 1997), (Lakonishok *et al*, 1994), (Lee and Stefek, 2008), (Lehmann and Modest, 1988), (Liew and Vassalou, 2000), (Lintner, 1965), (Lo, 2010), (Lo and MacKinlay, 1990), (MacKinlay, 1995), (MacQueen, 2003), (Markowitz, 1952, 1984), (Menchero and Mitra, 2008), (Merton, 1973), (Miller, 2006), (Motta *et al*, 2011), (Mukherjee and Mishra, 2005), (Ng *et al*, 1992), (Pastor and Stambaugh, 2003), (Roll and Ross, 1980), (Rosenberg, 1974), (Ross, 1976, 1978a, 1978b), (Scholes and Williams, 1977), (Schwert, 1990), (Shanken, 1987, 1990), (Shanken and Weinstein, 2006), (Sharpe, 1963, 1964), (Stock and Watson, 2002a, 2002b), (Stroyny, 2005), (Treynor, 1999), (Vassalou, 2003), (Whitelaw, 1997), (Zangari, 2003), (Zhang, 2010), and references therein.

[5] Shortcomings with traditional implementations are detailed in (Kakushadze and Yu, 2016).

[6] Assuming no two returns are 100% pair-wise (anti-)correlated.

matrix so long as $K < M$, where $M + 1$ is the number of observations in the time series. This holds even if $M < N$, in which case the sample covariance matrix is singular. In fact, one of the main motivations for considering factor models in the first instance is that in most practical applications $M < N$ (and often $M \ll N$), and even if $M \geq N$, in which case the sample covariance matrix is nonsingular, it is still out-of-sample unstable unless $M \gg N$, which is seldom (if ever) the case in practice. Factor models are intended to reduce this instability to a degree.[7]

The beauty of the statistical risk model construction is its simplicity. However, one must fix the number of risk factors K. We discuss two simple methods for fixing K in Section 3 (with variations). One is that of (Kakushadze, 2015d). Another, very different looking method, is based on our adaptation of eRank (effective rank) of (Roy and Vetterli, 2007) and yields results similar to (and further validates) that of (Kakushadze, 2015d). We use intraday alphas of (Kakushadze, 2015a) and backtest these methods out-of-sample. The method of (Kakushadze, 2015d) backtests better. We give R source code for computing a K-factor statistical risk model with K fixed via the aforementioned two methods (with variations) in Appendix A.[8]

In Section 4 we discuss how to compute principal components based on the returns. The "naïve" method is the power iterations method, which is applicable to more general matrices. However, it requires iterations and is computationally costly.[9] Because here we are dealing with sample covariance matrices, there is a simpler and faster way of computing principal components when $M \ll N$ that does not require any costly iterations and involves only $\mathcal{O}(M^2N)$ operations. We discuss this method in detail in Section 4 and give R source code for it in Appendix C. The main purpose of this exercise is to set up our further discussion in Section 4, where we explain that statistical risk models are simply certain deformations of the sample covariance matrix. We then also discuss "nontraditional" statistical risk models such as shrinkage (Ledoit and Wolf, 2004), which are also deformations of the sample covariance matrix, but involve M principal components as opposed to $K < M$ principal components. Generally, "nontraditional" models underperform.

We then take this a step further and explain that optimization using a statistical risk model is well-approximated by a weighted regression, where the regression is over the factor loadings matrix (i.e., the K principal components), and the weights are inverse specific variances. More precisely, this holds when the number of underlying returns $N \gg 1$, which is the case in most applications. In fact, optimization reduces to a weighted regression for $N \gg 1$ in a wider class of risk models that lack any "clustering" structure (we clarify the meaning of this statement in Section 4).

We briefly conclude in Section 5, where we discuss additional backtests, etc.

[7] Albeit in the case of statistical risk models, which are based on the very same returns used in computing the sample covariance matrix, this instability sizably seeps into the factor model.

[8] The source code given in Appendix A, Appendix B and Appendix C is not written to be "fancy" or optimized for speed or in any other way. Its sole purpose is to illustrate the algorithms described in the main text in a simple-to-understand fashion. See Appendix D for some legalese.

[9] We give R code for this method in Appendix B.

2 Statistical Risk Models

2.1 Sample Covariance Matrix

So, we have N instruments (e.g., stocks) with the time series of returns. Each time series contains $M+1$ observations corresponding to times t_s, and we will denote our returns as R_{is}, where $i = 1, \ldots, N$ and $s = 1, \ldots, M, M+1$ (t_1 is the most recent observation). The sample covariance matrix (SCM) is given by[10]

$$C_{ij} = \frac{1}{M} \sum_{s=1}^{M+1} X_{is} \; X_{js} \tag{1}$$

where $X_{is} = R_{is} - \overline{R}_i$ are serially demeaned returns; $\overline{R}_i = \frac{1}{M+1} \sum_{s=1}^{M+1} R_{is}$.

We are interested in cases where $M < N$, in fact, $M \ll N$. When $M < N$, C_{ij} is singular: we have $\sum_{s=1}^{M+1} X_{is} = 0$, so only M columns of the matrix X_{is} are linearly independent. Let us eliminate the last column: $X_{i,M+1} = -\sum_{s=1}^{M} X_{is}$. Then we can express C_{ij} via the first M columns:

$$C_{ij} = \sum_{s,s'=1}^{M} X_{is} \; \phi_{ss'} \; X_{js'} \tag{2}$$

Here $\phi_{ss'} = (\delta_{ss'} + u_s u_{s'})/M$ is a nonsingular $M \times M$ matrix ($s, s' = 1, \ldots, M$); $u_s \equiv 1$ is a unit M-vector. Note that $\phi_{ss'}$ is a 1-factor model (see below).

So, when $M < N$, the sample covariance matrix C_{ij} is singular with M nonzero eigenvalues. In this case we cannot invert C_{ij}, which is required in, e.g., optimization (mean-variance optimization (Markowitz, 1952), Sharpe ratio maximization (Sharpe, 1994), etc.). Furthermore, unless $M \gg N$, which is almost never (if ever) the case in practical applications, the off-diagonal elements of C_{ij} (covariances) generally are not expected to be stable out-of-sample. In contrast, the diagonal elements (variances) typically are much more stable out-of-sample and can be relatively reliably computed even for $M \ll N$ (which, in fact, is often the case in practical applications). So, we need to replace the sample covariance matrix C_{ij} by another *constructed* matrix – call it Γ_{ij} – that is much more stable out-of-sample and invertible (positive-definite). That is, we must build a risk model.

2.2 Factor Models

A popular method – at least in the case of equities – for constructing a nonsingular replacement Γ_{ij} for C_{ij} is via a factor model:

$$\Gamma_{ij} = \xi_i^2 \; \delta_{ij} + \sum_{A,B=1}^{K} \Omega_{iA} \; \Phi_{AB} \; \Omega_{jB} \tag{3}$$

[10] The difference between the unbiased estimate with M in the denominator vs. the maximum likelihood estimate with $M+1$ in the denominator is immaterial; in most applications $M \gg 1$.

Here: ξ_i is the specific (a.k.a. idiosyncratic) risk for each return; Ω_{iA} is an $N \times K$ factor loadings matrix; and Φ_{AB} is a $K \times K$ factor covariance matrix (FCM), $A, B = 1, \ldots, K$. The number of factors $K \ll N$ to have FCM more stable than SCM. And Γ_{ij} is positive-definite (and invertible) if FCM is positive-definite and all $\xi_i^2 > 0$.

2.3 Total Variances

The main objective of a risk model is to predict the covariance matrix out-of-sample as precisely as possible, including the out-of-sample variances. However, even though this requirement is often overlooked in practical applications, a well-built factor model had better reproduce the in-sample variances. That is, we require that the risk model variances Γ_{ii} be equal the in-sample variances C_{ii}:

$$\Gamma_{ii} = C_{ii} = \sigma_i^2 \tag{4}$$

Furthermore, as mentioned above, the N variances C_{ii} are relatively stable out-of-sample. It is therefore the $N(N-1)$ off-diagonal covariances, which are generally unstable out-of-sample, we must actually model. Put differently, we must model the *correlations* Ψ_{ij}, $i \neq j$, where $\Psi_{ij} = C_{ij}/\sigma_i\sigma_j$ is the sample correlation matrix, whose diagonal elements $\Psi_{ii} \equiv 1$. So, we need to replace the sample correlation matrix by another constructed matrix – let us call it $\widetilde{\Gamma}_{ij}$ – that is much more stable out-of-sample and invertible (positive-definite) subject to the conditions

$$\widetilde{\Gamma}_{ii} = \Psi_{ii} \equiv 1 \tag{5}$$

Once we build $\widetilde{\Gamma}_{ij}$, the risk model covariance matrix is given by $\Gamma_{ij} = \sigma_i\sigma_j\widetilde{\Gamma}_{ij}$. The advantage of modeling the correlation matrix Ψ_{ij} via $\widetilde{\Gamma}_{ij}$ as opposed to modeling the covariance matrix C_{ij} by Γ_{ij} is that the sample variances $C_{ii} = \sigma_i^2$ have a highly skewed (quasi log-normal) distribution, while Ψ_{ii} are uniform. In the following, in the main text[11] we will always focus on modeling Ψ_{ij} and for the sake of notational simplicity we will omit the twiddle on $\widetilde{\Gamma}_{ij}$, i.e., we will model

$$\Psi_{ij} = \sum_{s,s'=1}^{M} Y_{is}\ \phi_{ss'}\ Y_{js'} \tag{6}$$

via (3) subject to $\Gamma_{ii} \equiv 1$. Here $Y_{is} = X_{is}/\sigma_i$.

2.4 Principal Components

Looking at (6), it resembles a factor model, except that it has no specific risk (so it is singular). We cannot simply add some specific risk ad hoc to (6) as this would

[11] However, in Appendix A we give the source code which has an option to compute the factor model directly for C_{ij} as opposed to via Ψ_{ij}.

violate the requirement that $\Gamma_{ii} \equiv 1$ (as the resulting Γ_{ii} would be greater than 1). Therefore, to add some specific risk, we must simultaneously reduce the diagonal contribution from the factor risk. All values of $s, s' = 1, \ldots, M$ in the sum in (6) enter on the equal footing, in fact, we have a full $\mathbf{Z}_M$ permutational symmetry under which $s \to s+1$, $s' \to s'+1$, and $s\,(s') > M$ is identified with $s-M$ $(s'-M)$. When reducing factor risk we must either preserve this symmetry or somehow break it.

In fact, we can choose a different – but equivalent – basis, where this symmetry is not explicit and it is more evident how to "trim" the factor risk. Let $V_i^{(a)}$, $a = 1, \ldots, N$, be the principal components of Ψ_{ij} forming an orthonormal basis

$$\sum_{j=1}^{N} \Psi_{ij}\ V_j^{(a)} = \lambda^{(a)}\ V_i^{(a)} \tag{7}$$

$$\sum_{i=1}^{N} V_i^{(a)}\ V_i^{(b)} = \delta_{ab} \tag{8}$$

such that the eigenvalues $\lambda^{(a)}$ are ordered decreasingly: $\lambda^{(1)} > \lambda^{(2)} > \ldots$. More precisely, some eigenvalues may be degenerate. For simplicity – and this is not critical here – we will assume that all positive eigenvalues are non-degenerate. However, we can have multiple null eigenvalues. Typically, the number of nonvanishing eigenvalues[12] is M, where, as above, $M+1$ is the number of observations in the return time series. So, we have

$$\Psi_{ij} = \sum_{a=1}^{M} V_i^{(a)}\ \lambda^{(a)}\ V_j^{(a)} \tag{9}$$

This again resembles a factor model (with a diagonal factor covariance matrix). However, the $\mathbf{Z}_M$ symmetry is gone and we can readily "trim" the factor risk.

This is simply done by keeping only $K < M$ first principal components in the sum in (9) and replacing the diagonal contribution of the dropped $M-K$ principal components via the specific risk:

$$\Gamma_{ij} = \xi_i^2\ \delta_{ij} + \sum_{A=1}^{K} \lambda^{(A)}\ V_i^{(A)}\ V_j^{(A)} \tag{10}$$

$$\xi_i^2 = 1 - \sum_{A=1}^{K} \lambda^{(A)} \left(V_i^{(A)}\right)^2 \tag{11}$$

This corresponds to taking the factor loadings matrix and factor covariance matrix of the form

$$\Omega_{iA} = \sqrt{\lambda^{(A)}}\ V_i^{(A)}, \quad A = 1, \ldots, K \tag{12}$$

$$\Phi_{AB} = \delta_{AB} \tag{13}$$

This construction is nicely simple. However, what should K be?

[12] This number can be smaller if some returns are 100% correlated or anti-correlated. For the sake of simplicity – and this not critical here – we will assume that there are no such returns.

3 Fixing Factor Number

When $K = M$ we have $\Gamma_{ij} = \Psi_{ij}$, which is singular. Therefore, we must have $K \leq K_{max} < M$. So, what is K_{max}? And what is K_{min} (other than the evident $K_{min} = 1$)? It might be tempting to do complicated and convoluted things. We will not do this here. Instead, we will follow a pragmatic approach. One simple ("minimization" based) algorithm was set forth in (Kakushadze, 2015d). We review it below and then give yet another simple algorithm based on eRank (effective rank).

3.1 "Minimization" Algorithm

The idea is simple (Kakushadze, 2015d). It is based on the observation that, as K approaches M, $\min(\xi_i^2)$ goes to 0 (i.e., less and less of the total risk is attributed to the specific risk, and more and more of it is attributed to the factor risk), while as K approaches 0, $\max(\xi_i^2)$ goes to 1 (i.e., less and less of the total risk is attributed to the factor risk, and more and more of it is attributed to the specific risk). So, we can define K as follows:

$$|g(K) - 1| \to \min \tag{14}$$

$$g(K) = \sqrt{\min(\xi_i^2)} + \sqrt{\max(\xi_i^2)} \tag{15}$$

This simple algorithm works pretty well in practical applications.[13]

3.2 Effective Rank

Another simple method is to set (here $\text{Round}(\cdot)$ can be replaced by $\text{floor}(\cdot) = \lfloor \cdot \rfloor$)

$$K = \text{Round}(\text{eRank}(\Psi)) \tag{16}$$

Here $\text{eRank}(Z)$ is the effective rank (Roy and Vetterli, 2007) of a symmetric semi-positive-definite (which suffices for our purposes here) matrix Z. It is defined as

$$\text{eRank}(Z) = \exp(H) \tag{17}$$

$$H = -\sum_{a=1}^{L} p_a \ \ln(p_a) \tag{18}$$

$$p_a = \frac{\lambda^{(a)}}{\sum_{b=1}^{L} \lambda^{(b)}} \tag{19}$$

where $\lambda^{(a)}$ are the L *positive* eigenvalues of Z, and H has the meaning of the (Shannon a.k.a. spectral) entropy (Campbell, 1960), (Yang *et al*, 2005).

[13] The distribution of ξ_i^2 is skewed; typically, ξ_i^2 has a tail at higher values, while $\ln(\xi_i^2)$ has a tail at lower values, and the distribution is only roughly log-normal. So K is not (the floor/cap of) $M/2$, but somewhat higher, albeit close to it. See (Kakushadze, 2015d) for an illustrative example.

The meaning of eRank(Z) is that it is a measure of the effective dimensionality of the matrix Z, which is not necessarily the same as the number L of its positive eigenvalues, but often is lower. This is due to the fact that many returns can be highly correlated (which manifests itself by a large gap in the eigenvalues) thereby further reducing the effective dimensionality of the correlation matrix.

3.3 A Variation

When the average correlation[14] $\overline{\Psi} = \frac{1}{N^2}\sum_{i,j=1}^{N}\Psi_{ij}$ is high, then both the "minimization" and eRank based algorithms can produce low values of K (including 1). This is because in this case $\lambda^{(1)} \gg 1$ and there is a large gap in the eigenvalues. To circumvent this, we can define $K = K' + 1$, where K' is defined as above via the "minimization" or eRank based algorithms for the matrix

$$\Psi'_{ij} = \sum_{a=2}^{M} V_i^{(a)}\ \lambda^{(a)}\ V_j^{(a)} \tag{20}$$

I.e., we simply drop the first eigenvalue, determine the corresponding value of K', and add 1 to it. Appendix A provides R source code for both the "minimization" and eRank based algorithms with and without utilizing the K' based definition.

3.4 Some Backtests

Let us backtest the above algorithms for fixing K via utilizing the same backtesting procedure as in (Kakushadze, 2015d). The remainder of this subsection very closely follows most parts of Section 6 of (Kakushadze, 2015d).[15]

3.4.1 Notations

Let P_{is} be the time series of stock prices, where $i = 1, \ldots, N$ labels the stocks, and $s = 1, \ldots, M+1$ labels the trading dates, with $s = 1$ corresponding to the most recent date in the time series. The superscripts O and C (unadjusted open and close prices) and AO and AC (open and close prices fully adjusted for splits and dividends) will distinguish the corresponding prices, so, e.g., P_{is}^{C} is the unadjusted close price. V_{is} is the unadjusted daily volume (in shares). Also, for each date s we define the overnight return as the previous-close-to-open return:

$$E_{is} = \ln\left(P_{is}^{AO}/P_{i,s+1}^{AC}\right) \tag{21}$$

This return will be used in the definition of the expected return in our mean-reversion alpha. We will also need the close-to-close return

$$R_{is} = \ln\left(P_{is}^{AC}/P_{i,s+1}^{AC}\right) \tag{22}$$

[14] Instead we can define $\overline{\Psi} = \frac{1}{N(N-1)}\sum_{i,j=1;\ i\neq j}^{N}$. Since $N \gg 1$, the difference is immaterial.

[15] We "rehash" it here not to be repetitive but so that the presentation herein is self-contained.

An out-of-sample (see below) time series of these returns will be used in constructing the risk models. All prices in the definitions of E_{is} and R_{is} are fully adjusted.

We assume that: i) the portfolio is established at the open[16] with fills at the open prices P^O_{is}; ii) it is liquidated at the close on the same day – so this is a purely intraday alpha – with fills at the close prices P^C_{is}; and iii) there are no transaction costs or slippage – our aim here is not to build a realistic trading strategy, but to test relative performance of various risk models and see what adds value to the alpha and what does not. The P&L for each stock

$$\Pi_{is} = H_{is} \left[\frac{P^C_{is}}{P^O_{is}} - 1 \right] \tag{23}$$

where H_{is} are the *dollar* holdings. The shares bought plus sold (establishing plus liquidating trades) for each stock on each day are computed via $Q_{is} = 2|H_{is}|/P^O_{is}$.

3.4.2 Universe Selection

For the sake of simplicity,[17] we select our universe based on the average daily dollar volume (ADDV) defined via (note that A_{is} is out-of-sample for each date s):

$$A_{is} = \frac{1}{d} \sum_{r=1}^{d} V_{i,s+r} \; P^C_{i,s+r} \tag{24}$$

We take $d = 21$ (i.e., one month), and then take our universe to be the top 2000 tickers by ADDV. To ensure that we do not inadvertently introduce a universe selection bias, we rebalance monthly (every 21 trading days, to be precise). I.e., we break our 5-year backtest period (see below) into 21-day intervals, we compute the universe using ADDV (which, in turn, is computed based on the 21-day period immediately preceding such interval), and use this universe during the entire such interval. We do have the survivorship bias as we take the data for the universe of tickers as of 9/6/2014 that have historical pricing data on http://finance.yahoo.com (accessed on 9/6/2014) for the period 8/1/2008 through 9/5/2014. We restrict this universe to include only U.S. listed common stocks and class shares (no OTCs, preferred shares, etc.) with BICS (Bloomberg Industry Classification System) sector assignments as of 9/6/2014.[18] However, as discussed in detail in Section 7 of (Kakushadze, 2015a), the survivorship bias is not a leading effect in such backtests.[19]

[16] This is a so-called "delay-0" alpha: the same price, P^O_{is} (or adjusted P^{AO}_{is}), is used in computing the expected return (via E_{is}) and as the establishing fill price.

[17] In practical applications, the trading universe of liquid stocks typically is selected based on market cap, liquidity (ADDV), price and other (proprietary) criteria.

[18] The choice of the backtesting window is intentionally taken to be exactly the same as in (Kakushadze, 2015d) to simplify various comparisons, which include the results therefrom.

[19] Here we are after the *relative outperformance*, and it is reasonable to assume that, to the leading order, individual performances are affected by the survivorship bias approximately equally as the construction of all alphas and risk models is "statistical" and oblivious to the universe.

3.4.3 Backtesting

We run our simulations over a period of 5 years (more precisely, 1260 trading days going back from 9/5/2014, inclusive). The annualized return-on-capital (ROC) is computed as the average daily P&L divided by the intraday investment level I (with no leverage) and multiplied by 252. The annualized Sharpe Ratio (SR) is computed as the daily Sharpe ratio multiplied by $\sqrt{252}$. Cents-per-share (CPS) is computed as the total P&L divided by the total shares traded.[20]

3.4.4 Optimized Alphas

The optimized alphas are based on the expected returns E_{is} optimized via Sharpe ratio maximization using the risk models we are testing, i.e., the covariance matrix $\widehat{\Gamma}_{ij} = \sigma_i \sigma_j \Gamma_{ij}$ with Γ_{ij} given by (10), which we compute every 21 trading days (same as for the universe). For each date (we omit the index s) we maximize the Sharpe ratio subject to the dollar neutrality constraint:

$$\mathcal{S} = \frac{\sum_{i=1}^{N} H_i \; E_i}{\sqrt{\sum_{i,j=1}^{N} \widehat{\Gamma}_{ij} \; H_i \; H_j}} \to \max \tag{25}$$

$$\sum_{i=1}^{N} H_i = 0 \tag{26}$$

The solution is given by

$$H_i = -\eta \left[\sum_{j=1}^{N} \widehat{\Gamma}^{-1}_{ij} \; E_j - \sum_{j=1}^{N} \widehat{\Gamma}^{-1}_{ij} \; \frac{\sum_{k,l=1}^{N} \widehat{\Gamma}^{-1}_{kl} \; E_l}{\sum_{k,l=1}^{N} \widehat{\Gamma}^{-1}_{kl}} \right] \tag{27}$$

where $\widehat{\Gamma}^{-1}$ is the inverse of $\widehat{\Gamma}$, and $\eta > 0$ (mean-reversion alpha) is fixed via (we set the investment level I to \$20M in our backtests)

$$\sum_{i=1}^{N} |H_i| = I \tag{28}$$

Note that (27) satisfies the dollar neutrality constraint (26).

The simulation results are given in Table 1 for K obtained via the "minimization" and eRank based algorithms with and without utilizing the K' based definition (see Subsection 3.3) with Round$(\cdot)$ and floor$(\cdot)$ in (16). The "minimization" and eRank methods not based on K' produce similar results, which further validates the

[20] As mentioned above, we assume no transaction costs, which are expected to reduce the ROC of the optimized alphas by the same amount as all strategies trade the exact same amount by design. Therefore, including the transaction costs would have no effect on the actual *relative outperformance* in the horse race, which is what we are after here.

"minimization" method of (Kakushadze, 2015d). The slight improvement in CPS in the eRank method is immaterial and disappears when we impose position bounds (which in this case are the same as trading bounds as the strategy is purely intraday)

$$|H_{is}| \leq 0.01\ A_{is} \tag{29}$$

where A_{is} is ADDV defined in (24). These backtests use the R code in Appendix C of (Kakushadze, 2015d) for optimization with bounds. Table 2 gives the simulation results with these bounds for all of the above cases except for the eRank method with the K' based definition. In the latter case, because the typical value of K is close to or the same as the maximum allowed $K_{max} = M - 1 = 19$ (see rows 5 and 6 in Table 3), some of the desired holdings come out to be large compared with the (reasonable) bounds (29) (see row 7 in Table 3). The "average correlation" $\overline{\Psi}$ is not very high (nor is it very low – see row 8 in Table 3), so it is just as well that the value of K is not low and the K' based method is overkill, which is why it underperforms.

4 Principal Components, Deformations, etc.

4.1 Power Iterations

To construct the statistical risk models, we need to compute the first M principal components of Ψ_{ij}. One way is to successively use the power iterations method (Mises and Pollaczek-Geiringer, 1929), which usually costs more than $\mathcal{O}(M^2N)$ operations. To compute the first principal component costs $\mathcal{O}(n_{iter}MN)$, where n_{iter} is the number of iterations:

$$\left[V_i^{(1)}\right]_{r+1} = \frac{\widetilde{V}_i}{\sqrt{\sum_{j=1}^N \widetilde{V}_i^2}} \tag{30}$$

$$\widetilde{V}_i = \sum_{j=1}^N \Psi_{ij}\ \left[V_j^{(1)}\right]_r = \frac{1}{M}\sum_{j=1}^N \sum_{s=1}^{M+1} Y_{is}\ Y_{js}\ \left[V_j^{(1)}\right]_r \tag{31}$$

where r labels the iterations.[21] Each iteration costs $\mathcal{O}(MN)$ operations. However, we need to compute M principal components. This can be done as follows. Let

$$\Psi_{ij}^{(a)} = \frac{1}{M}\sum_{s=1}^{M+1} Y_{is}^{(a)}\ Y_{js}^{(a)} \tag{32}$$

$$Y_{is}^{(a+1)} = Y_{is}^{(a)} - V_i^{(a)} \sum_{j=1}^N V_j^{(a)}\ Y_{js}^{(a)} \tag{33}$$

$$Y_{is}^{(1)} = Y_{is} \tag{34}$$

[21] Note that the sum over s rums from 1 to $M + 1$ in (31).

Note that $\Psi_{ij}^{(1)} = \Psi_{ij}$. The first principal component of $\Psi_{ij}^{(a)}$ – which we can compute using the power iterations method – is the same as the a-th principal component $V_i^{(a)}$ of Ψ_{ij}. Each such computation costs $\mathcal{O}(n_{iter}^{(a)} MN)$ operations. So, computing the first M principal components costs $\mathcal{O}(n_{iter}^{tot} MN)$ operations, where $n_{iter}^{tot} = \sum_{a=1}^{M} n_{iter}^{(a)}$, and typically $n_{iter}^{tot} \gg M$. Source code for this procedure is given in Appendix B. Table 4 gives an analysis for a time series with $M = 19$ for $N = 2,339$ stock returns (close-to-close). The results show that $n_{iter}^{tot} \gg M$ for a reasonable computational precision, so the cost of computing the M principal components is substantially greater than $\mathcal{O}(M^2 N)$ operations. The power iterations method is not cheap...[22]

The above procedure simply amounts to successively removing the already-computed principal components from Ψ_{ij}. Indeed, $\Psi_{ij}^{(a+1)} = \Psi_{ij}^{(a)} - \lambda^{(a)} V_i^{(a)} V_j^{(a)}$. The reason to express $\Psi_{ij}^{(a)}$ via $Y_{is}^{(a)}$ in (32) is so we have a factorized form, which leads to a much smaller number of operations required to multiply this matrix by the iteration $\left[V_j^{(a)}\right]_r$. Note that $\Psi_{ij}^{(M)} = \lambda^{(M)} V_i^{(M)} V_j^{(M)}$, so irrespective of $\left[V_i^{(M)}\right]_{init}$, we get $V_i^{(M)}$ via (30) and (31) (with $V_i^{(1)}$, Ψ_{ij} and Y_{is} replaced by $V_i^{(M)}$, $\Psi_{ij}^{(M)}$ and $Y_{is}^{(M)}$) right at the first iteration (hence only 2 iterations in the last row in Table 4).

4.2 Computing Principal Components without Iterations

However, when $M \ll N$, we can compute the M principal components of the sample correlation matrix without any costly iterations (involving N-vectors) and the cost is $\mathcal{O}(M^2 N)$ operations. We start with (6). Let (in matrix notation) $\phi = \varphi\, \varphi^T$, where φ is the Cholesky decomposition of ϕ. ($\varphi_{ss} = \sqrt{(s+1)/sM}$; $\varphi_{ss'} = 1/\sqrt{s'(s'+1)M}$ for $s > s'$; $\varphi_{ss'} = 0$ for $s < s'$.) Let $\widetilde{Y} = Y\varphi$. Then we have

$$\Psi_{ij} = \sum_{s=1}^{M} \widetilde{Y}_{is}\ \widetilde{Y}_{js} \tag{35}$$

The columns of $\widetilde{Y}_{is}$ are not orthonormal. Let

$$G_{ss'} = \sum_{i=1}^{N} \widetilde{Y}_{is}\ \widetilde{Y}_{is'} \tag{36}$$

We can readily find its eigenpairs. (This costs only $\mathcal{O}(M^3)$ operations.[23]) Let the eigenvalues be $\rho^{(a)}$ and the principal components be $U_s^{(a)}$, $a = 1, \ldots, M$. Then the first M principal components of Ψ_{ij} are given by (this costs $\mathcal{O}(M^2 N)$ operations):

$$V_i^{(a)} = \frac{1}{\sqrt{\rho^{(a)}}} \sum_{s=1}^{M} \widetilde{Y}_{is}\ U_s^{(a)} \tag{37}$$

[22] Also see, e.g., (Gubernatis and Booth, 2008) and references therein.

[23] Technically, this involves iterations, but no costly iterations involving N-vectors ($M \ll N$).

Indeed, we have (in matrix notation) $\Psi\ V^{(a)} = \lambda^{(a)}\ V^{(a)}$, where $\lambda^{(a)} = \rho^{(a)}$, and $\sum_{i=1}^{N} V_i^{(a)}\ V_i^{(b)} = \delta_{ab}$. The R source code for this method is given in Appendix C.

4.3 Statistical Risk Model = Deformation

The purpose of the last subsection is not only to discuss an efficient method for computing eigenpairs, but also to rewrite the statistical risk model given by (10) and (11) directly in terms of the (normalized demeaned) returns Y_{is}. Thus, using (37) we have[24]

$$\Gamma_{ij} = \xi_i^2\ \delta_{ij} + \sum_{s,s'=1}^{M} Y_{is}\ \widetilde{\phi}_{ss'}\ Y_{js'} \tag{38}$$

$$\xi_i^2 = 1 - \sum_{s,s'=1}^{M} Y_{is}\ \widetilde{\phi}_{ss'}\ Y_{is'} \tag{39}$$

where

$$\widetilde{\phi}_{ss'} = \sum_{r,r'=1}^{M} \varphi_{sr}\ F_{rr'}\ \varphi_{s'r'} \tag{40}$$

$$F_{rr'} = \sum_{a=1}^{K} U_r^{(a)}\ U_{r'}^{(a)} \tag{41}$$

When $K = M$, we have $F_{rr'} = \delta_{rr'}$ and $\widetilde{\phi}_{ss'} = \phi_{ss'}$, so $\xi_i^2 \equiv 0$ and $\Gamma_{ij} = \Psi_{ij}$. So, the statistical risk model (10), as it can be rewritten via (38), is nothing but a *deformation* (or regularization) of the sample correlation matrix Ψ_{ij} given by (6): we deform $\phi_{ss'} \to \widetilde{\phi}_{ss'}$ thereby reducing the factor risk contribution into the total risk and replace the deficit by the specific risk. In this regard, note that we can consider more general deformations of the sample correlation matrix of the form

$$\widetilde{\Psi}_{ij} = \Delta_{ij} + \sum_{s,s'=1}^{M} Y_{is}\ \widetilde{\phi}_{ss'}\ Y_{js'} \tag{42}$$

subject to the requirements that $\widetilde{\Psi}_{ij}$ be positive-definite and $\widetilde{\Psi}_{ii} \equiv 1$. However, in practice, sticking to our basic premise that there is no information available beyond the returns (i.e., no style or industry factors can be constructed), there is no choice but to take diagonal Δ_{ij}, which is then completely fixed. So, we are left with the choice of deforming $\phi_{ss'} \to \widetilde{\phi}_{ss'}$. And (40) is just *one* of myriad such deformations.

In fact, (40) is not even the simplest such deformation. It is the one that arises in *traditional* statistical risk models based on principal components. However, a

[24] Note that this holds for any $M \leq N$, not just $M \ll N$.

choice to work with the principal components is by large simply a matter of taste (or even habit). It is just one basis, which a priori is not necessarily better or worse than any other basis. The *intuitive* justification behind a statistical risk model of the form (10) is clear: we keep the first K principal components with the largest contributions to the sample correlation matrix – based on the fact that the eigenvalues $\lambda^{(1)} > \lambda^{(2)} > \ldots$ – and replace the deficit (on the diagonal) by the specific risk. On the surface it all appears to make sense. However, the principal components beyond the first one are not stable out-of-sample, and this instability is inherited from that of the off-diagonal elements of the sample correlation matrix (i.e., pair-wise correlations). The first principal component also depends on the latter; however, for large N, in the leading approximation we have $V_i^{(1)} \approx 1/\sqrt{N}$ (the so-called "market mode" – see, e.g., (Bouchaud and Potters, 2011) and references therein), which is by definition stable, albeit subleading corrections are not. In any event, there is no reason to limit ourselves to the deformations of the form (40).

4.4 "Nontraditional" Statistical Risk Models

While a priori we can consider an arbitrary deformation $\phi_{ss} \to \widetilde{\phi}_{ss'}$ (subject to the requirement that $\xi_i^2 > 0$), in practice it ought to be reasonable in the sense that it has to work out-of-sample. In this regard, keeping the first K principal component of Ψ_{ij} can be argued to be reasonable in the sense that, while the principal components themselves are not stable out-of-sample (except for the quasi-stable first principal component – see above), tossing the higher principal components makes sense as their contributions are suppressed by the corresponding eigenvalues. This yields the deformation (40), which involves the matrix $F_{ss'}$ constructed from the returns Y_{is}.

So, can we deform $\phi_{ss'}$ directly, without any reference to the returns Y_{is}? The simplest such deformation is

$$\widetilde{\phi}_{ss'} = (1-q)\,\phi_{ss'} \tag{43}$$

$$\xi_i^2 \equiv q \tag{44}$$

This is nothing but shrinkage (Ledoit and Wolf, 2004). Here we are shrinking the sample correlation matrix (as opposed to the sample covariance matrix, which is what is usually done); $0 < q < 1$ is the "shrinkage constant"; the "shrinkage target" is the diagonal $N \times N$ unit matrix. So, in this case we have $K = M$, the factor loadings matrix is given by (12), but instead of (13) we have the factor covariance matrix $\Phi_{AB} = (1-q)\,\delta_{AB}$. So, shrinkage is a factor model (Kakushadze, 2016).

Unlike in the traditional statistical risk models, where we need to fix the number of factors K, in shrinkage the number of factors is fixed ($K = M$ for the diagonal shrinkage target; $\Delta_{ij} = \xi_i^2 \delta_{ij}$ – see below), but we must fix the shrinkage constant q instead.[25] However, contrary to an apparent common misconception, the value of q makes little difference if the number of returns $N \gg 1$. This is because in this case optimization using statistical models is well-approximated by a weighted regression.

[25] A way to fix q is discussed in (Ledoit and Wolf, 2004).

4.5 Optimization ≈ Regression

Optimization involves the inverse $\widehat{\Gamma}_{ij}^{-1}$ of the model covariance matrix $\widehat{\Gamma}_{ij} = \sigma_i \sigma_j \Gamma_{ij}$, where Γ_{ij} is given by (3). For our purposes here it is convenient to rewrite Γ_{ij} via $\Gamma_{ij} = \xi_i \xi_j \gamma_{ij}$, where

$$\gamma_{ij} = \delta_{ij} + \sum_{A=1}^{K} \beta_{iA}\ \beta_{jA} \tag{45}$$

and $\beta_{iA} = \widetilde{\beta}_{iA}/\xi_i$. Here (in matrix notation) $\widetilde{\beta} = \Omega\ \widetilde{\Phi}$, and $\widetilde{\Phi}$ is the Cholesky decomposition of Φ, so $\widetilde{\Phi}\ \widetilde{\Phi}^T = \Phi$. So, we have $\Gamma_{ij}^{-1} = \gamma_{ij}^{-1}/\xi_i \xi_j$, where

$$\gamma_{ij}^{-1} = \delta_{ij} - \sum_{A,B=1}^{K} \beta_{iA}\ Q_{AB}^{-1}\ \beta_{jB} \tag{46}$$

$$Q_{AB} = \delta_{AB} + q_{AB} \tag{47}$$

$$q_{AB} = \sum_{i=1}^{N} \beta_{iA}\ \beta_{iB} \tag{48}$$

It then follows that, if all $q_{AA} = \sum_{i=1}^{N} \beta_{iA}^2 \gg 1$, then $Q_{AB} \approx q_{AB}$ and

$$\gamma_{ij}^{-1} \approx \delta_{ij} - \sum_{A,B=1}^{K} \beta_{iA}\ q_{AB}^{-1}\ \beta_{jB} = \delta_{ij} - \frac{1}{\xi_i \xi_j} \sum_{A,B=1}^{K} \Omega_{iA}\ \widetilde{q}_{AB}^{-1}\ \Omega_{jB} \tag{49}$$

$$\widetilde{q}_{AB} = \sum_{i=1}^{N} \frac{1}{\xi_i^2}\ \Omega_{iA}\ \Omega_{iB} \tag{50}$$

I.e., in this case γ_{ij}^{-1} is (approximately) independent of the factor covariance matrix. Furthermore, for an arbitrary vector Z_i, we have

$$\sum_{i=1}^{N} \widehat{\Gamma}_{ij}^{-1}\ Z_j \approx \omega_i \left[Z_i - \sum_{j=1}^{N} \sum_{A,B=1}^{K} \widehat{\Omega}_{iA}\ \widetilde{q}_{AB}^{-1}\ \widehat{\Omega}_{jB}\ \omega_j\ Z_j \right] = \omega_i\ \varepsilon_i \tag{51}$$

Here: $\omega_i = 1/\sigma_i^2 \xi_i^2$; $\widehat{\Omega}_{iA} = \sigma_i \Omega_{iA}$; and ε_i are the residuals of the cross-sectional weighted regression (without the intercept) of Z_i over $\widehat{\Omega}_{iA}$ with the weights ω_i. So, in this regime, the optimization based on shrinkage reduces to a weighted regression.

The question is why – or, more precisely, when – all $q_{AA} \gg 1$. This is the case when: i) N is large, and ii) there is no "clustering" in the vectors β_{iA}. That is, we do not have vanishing or small values of β_{iA}^2 for most values of the index i with only a small subset thereof having $\beta_{iA}^2 \gtrsim 1$. Without "clustering", to have $q_{AA} \lesssim 1$, we would have to have $\beta_{iA}^2 \ll 1$, i.e., γ_{ij} and consequently Γ_{ij} would be almost diagonal. And such "clustering" is certainly absent if Ω_{iA} are the M principal components of the sample correlation matrix, so the above approximation holds in the case of shrinkage. The matrix $\widehat{\Omega}_{iA} = V_i^{(A)}$, $A = 1, \ldots, M$, and is independent of

the shrinkage constant q. The regression weights $\omega_i = 1/q\sigma_i^2$ do depend on q, but this dependence does not affect the desired holdings H_i in (27) as q simply rescales the overall normalization coefficient η in (27). That is, for large N, the desired holdings based on shrinkage are approximately independent of the shrinkage constant q. Table 5 gives the simulation results[26] for various values of q. For $q = 0$ we have $\Gamma_{ij} = \Psi_{ij}$, which is singular, so in this case the computation is done via the weighted regression – see (51).[27] Not only is the value of the shrinkage constant immaterial, but the shrinkage based models sizably underperform the traditional statistical risk model with K fixed via the "minimization" algorithm of (Kakushadze, 2015d).

To be clear, let us note that our observation that for large N the optimization reduces to the weighted regression also applies to the traditional statistical risk models. The difference between the latter and shrinkage is that i) fewer than M principal components are used as the columns of the $N \times K$ factor loadings matrix Ω_{iA} (i.e., $K < M$), and ii) the specific variances ξ_i^2 are no longer uniform (cf. (44)).

Let us mention that the shrinkage deformation (43) can be straightforwardly generalized via[28] $\widetilde{\phi}_{ss'} = \frac{1}{M}\left(\alpha\delta_{ij} + \beta u_s u_{s'}\right)$ $(u_s \equiv 1)$. In this case we still have the same M risk factors,[29] but we no longer have uniform $\xi_i^2 = 1 - \alpha - \frac{\beta-\alpha}{M}\left[\sum_{s=1}^{M} Y_{is}\right]^2$. For $\beta = \alpha = 1 - q$ this gives the standard shrinkage. Generally, we can have $\beta \neq \alpha$.

Finally, we can actually increase the number of risk factors beyond M. This can be done by considering Δ_{ij} in (42) that itself is a factor model. As before, let us continue assuming that we cannot construct any nontrivial style factors and there is no industry classification either. We can still construct a 1-factor model for Δ_{ij} with the intercept as the factor. If we choose the deformation $\widetilde{\phi}_{ss'} = (1-q)\phi_{ss'}$, then we can set $\Delta_{ij} = \xi_i^2\ \delta_{ij} + q\ \rho\ \nu_i\ \nu_j$ $(\nu_i \equiv 1,\ |\rho| < 1)$, so the factor model covariance matrix now reads[30]

$$\Gamma_{ij} = \xi_i^2\ \delta_{ij} + q\ \rho\ \nu_i\ \nu_j + (1-q)\sum_{s,s'=1}^{M} Y_{is}\ \phi_{ss'}\ Y_{js'} \tag{52}$$

[26] With the bounds (29) – not including the bounds does not change the qualitative picture.

[27] With no bounds the desired holdings (27) can be obtained in two steps: first we regress the expected returns E_i over $\widehat{\Omega}_{iA}$ (without the intercept) and weights $\omega_i = 1/\sigma_i^2$, and then we demean the residuals. With the bounds these two steps cannot be separated. So, we have two options. We can simply set q to a small number (e.g., $q = 10^{-6}$) and use the R code in Appendix C of (Kakushadze, 2015d). Or we can modify said code by (straightforwardly) replacing the optimization procedure therein via a weighted regression thereby arriving at "bounded regression with linear constraints". The result is slightly better if we simply regress E_i over $\widehat{\Omega}_{iA}$ *with* the intercept and weights $\omega_i = 1/\sigma_i^2$, in which case we have ROC 40.74%, SR 13.86, CPS 1.81.

[28] This choice uniquely preserves the $\mathbf{Z}_M$ permutational symmetry under which $s \to s+1$, $s' \to s'+1$, and $s\,(s') > M$ is identified with $s - M$ $(s' - M)$.

[29] I.e., the first M principal components of Ψ_{ij} – the deformation $\phi_{ss'} \to \widetilde{\phi}_{ss'}$ simply rotates the basis so long as $\widetilde{\phi}_{ss'}$ is nonsingular.

[30] This is the shrinkage model of (Ledoit and Wolf, 2004) where the shrinkage target corresponds to uniform correlations. If we take $\widetilde{\phi}_{ss'} = \frac{1}{M}\left(\alpha\delta_{ij} + \beta u_s u_{s'}\right)$, then ξ_i^2 are nonuniform for $\beta \neq \alpha$.

where $\xi_i^2 \equiv q(1-\rho)$. So, we have $M+1$ factors, the M principal components plus the intercept. For large N, as above, the optimization reduces to a weighted regression, except that now it is with the intercept. The result is essentially independent of the values of q and ρ (see Table 6) and expectedly somewhat better than in Table 5.

5 Concluding Remarks

To begin with, let us tie a loose end. We discussed the algorithms for fixing the number of statistical factors K (the "minimization" and eRank based algorithms and their variations). Here we can ask: how do we know that, say, the "minimization" based algorithm works better than picking some fixed value of K? This is tricky.

Indeed, how do we pick such "optimal" fixed K? We can do this by simply running $M-1$ backtests with fixed $K = 1, \ldots, M-1$ for a given alpha and picking the value of K that performs best. However, this "optimal" value would be in-sample. There is no guarantee that it will work out-of-sample. Furthermore, generally it will vary from alpha to alpha. The aforementioned "minimization" and eRank based algorithms by construction are oblivious to a choice of a sample, and even if they do not necessarily produce the "optimal" value of K for any given sample, they work for any sample. So, here we can ask whether they produce reasonable results for a given sample. Tables 7 and 8 give the simulation results for various fixed values of K without and with the bounds (29), respectively. Looking at these results (especially those with the bounds, as any outperformance without the bounds should be taken with a grain of salt) it is evident that the fixed K performance peaks around $K \approx 10$, while the "minimization" based algorithm compares closer to K between 12 and 13. This is consistent with the first row of Table 3. The important thing is that the "minimization" based algorithm produces results that are close to the in-sample "optimal" results for fixed $K = 10$.

In this regard it is instructive to run the following two series of backtests: i) taking the maximum $K_1 = M$ risk factors fixed but ad hoc basing the specific risks on the K-factor model with varying K (see Table 9 and Figure 1); and ii) varying the number K of the risk factors but ad hoc setting the specific risk equal the in-sample risk, i.e., $\xi_i^2 \equiv 1$ (see Table 10 and Figure 2). The simulation results indicate that both the specific risk and the number of risk factors make a difference. The performance in Table 9 peaks around $K = 13$ (also see Figure 1), which is the (approximate) number of risk factors fixed via the "minimization" based algorithm of (Kakushadze, 2015d). Not surprisingly, the Sharpe ratio in Table 10 (also see Figure 2) improves as K increases. However, the improvement rate slows down at higher K, so the specific risk effect is dominant. The performance in both Table 9 and Table 10 is worse than for the "minimization" based algorithm (see Table 1).

We already mentioned above (see footnote 13) that the "minimization" based algorithm produces K somewhat higher than $M/2$ due to a typically skewed ξ_i^2 distribution for a typical universe of stocks we dealt with in our backtests. Based

on the above discussion, it might be tempting to simply set K to the floor or cap of (or just rounded) $M/2$. Again, such a heuristic might work in-sample for some alphas, but not generally. Thus, if the underlying returns are highly correlated, a typical value of K produced by the "minimization" and eRank algorithms can be substantially lower than $M/2$ (including $K = 1$). In such cases we can use the variation described in Subsection 3.3, and then a priori there is no reason to expect $K \approx M/2$. A safer path would appear to be to use different methods, see if they produce consistent results out-of-sample, and pick one based thereon.

Let us also mention that in the R code in both functions in Appendix A we use the built-in R function `eigen()` to compute eigenpairs. For large N it is more efficient to replace it by the `qrm.calc.eignen.eff()` function given in Appendix C.

From our discussion above and backtests it is evident that "nontraditional" statistical risk models such as shrinkage underperform the traditional statistical risk models. The reason why is that in "nontraditional" models the rank of $\widetilde{\phi}_{ss'}$ equals M, while in traditional models it is reduced, which yields nonuniform specific risks.

Furthermore, in the case of equity portfolios for which well-built and granular enough industry classifications are available, statistical risk models simply have no chance against risk models utilizing an industry classification such as heterotic risk models (Kakushadze, 2015d) or heterotic CAPM (Kakushadze and Yu, 2016). The reason for this is twofold: i) industry factors are much more ubiquitous (thereby covering much more of the relevant risk space); and ii) principal components beyond the first one are unstable out-of-sample. In contrast, heterotic risk models use much more stable first principal components within each "cluster" (e.g., BICS subindustry), while heterotic CAPM uses a style factor.

A R Code for Statistical Risk Models

In this appendix we give the R (R Package for Statistical Computing, http://www.r-project.org) source code for building a purely statistical risk model (principal components) based on the algorithm we discuss in Sections 2 and 3, including the "minimization" and eRank based algorithms for fixing the number of factors K in Section 3. The two functions below are essentially self-explanatory and straightforward.

The function `qrm.cov.pc(ret, use.cor = T, excl.first = F)` corresponds to the "minimization" based method for fixing K. The input is: i) `ret`, an $N \times d$ matrix of returns (e.g., daily close-to-close returns), where N is the number of returns, $d = M + 1$ is the number of observations in the time series (e.g., the number of trading days), and the ordering of the dates is immaterial; ii) `use.cor`, where for `TRUE` (default) the risk factors are computed based on the principal components of the sample correlation matrix Ψ_{ij}, whereas for `FALSE` they are computed based on the sample covariance matrix C_{ij}; `excl.first`, where for `TRUE` the K' based method of Subsection 3.3 is used. The output is a list: `result$spec.risk` is the specific risk ξ_i (not the specific variance ξ_i^2) for `use.cor = F`, and $\sigma_i \xi_i$ (where $\sigma_i = \sqrt{C_{ii}}$) for `use.cor`

= T (recall that in this case ξ_i is the specific risk for the factor model for Ψ_{ij}, not C_{ij}); result$fac.load is the factor loadings matrix Ω_{iA} for use.cor = F, and $\sigma_i\Omega_{iA}$ for use.cor = T (recall that in this case Ω_{iA} is the factor loadings matrix for the factor model for Ψ_{ij}, not C_{ij}); result$fac.cov is the factor covariance matrix Φ_{AB} (with the normalization (12) for the factor loadings matrix, $\Phi_{AB} = \delta_{AB}$); result$cov.mat is the factor model covariance matrix Γ_{ij} for use.cor = F, and $\sigma_i\sigma_j\Gamma_{ij}$ for use.cor = T; result$inv.cov is the matrix inverse to result$cov.mat; result$pc are the first K principal components of a) C_{ij} for use.cor = F, and b) Ψ_{ij} for use.cor = T.

The second function is qrm.erank.pc(ret, use.cor = T, do.trunc = F, k = 0, excl.first = F) and corresponds to the eRank based method for fixing K for the default parameter k = 0. The input is the same as in the qrm.cov.pc() function except for the additional parameters do.trunc = F and k = 0. For a positive integer k the code simply takes its value as the number of factors K. For k = 0 (default) the code uses the eRank method: if do.trunc = F (default), then $K = \text{Round}(\text{eRank}(\cdot))$, while if do.trunc = T, then $K = \text{floor}(\text{eRank}(\cdot))$. (The argument of $\text{eRank}(\cdot)$ is the matrix C_{ij} if use.cor = F, and the matrix Ψ_{ij} if use.cor = T). The output is the same as in the qrm.cov.pc() function.

```
qrm.cov.pc <- function (ret, use.cor = T, excl.first = F)
{
   print("Running qrm.cov.pc()...")

   tr <- apply(ret, 1, sd)
   if(use.cor)
      ret <- ret / tr

   d <- ncol(ret)
   x <- t(ret)
   x <- var(x, x)
   tv <- diag(x)
   x <- eigen(x)
   if(excl.first)
   {
      k1 <- 2
      y1 <- sqrt(x$values[1]) * matrix(x$vectors[, 1], nrow(ret), 1)
      x1 <- y1 %*% t(y1)
      tv <- tv - diag(x1)
   }
   else
   {
      k1 <- 1
      x1 <- 0
   }
```

```
g.prev <- 999
for(k in k1:(d-1))
{
   u <- x$values[k1:k]
   v <- x$vectors[, k1:k]
   v <- t(sqrt(u) * t(v))
   x.f <- v %*% t(v)
   x.s <- tv - diag(x.f)
   z <- x.s / tv
   g <- abs(sqrt(min(z)) + sqrt(max(z)) - 1)

   if(is.na(g))
      break

   if(g > g.prev)
      break

   g.prev <- g

   spec.risk <- sqrt(x.s)
   if(excl.first)
      fac.load <- cbind(y1, v)
   else
      fac.load <- v
   fac.cov <- diag(1, k)
   cov.mat <- diag(x.s) + x.f + x1
}

y.s <- 1 / spec.risk^2
v <- fac.load
v1 <- y.s * v
inv.cov <- diag(y.s) - v1 %*%
   solve(diag(1, ncol(v)) + t(v) %*% v1) %*% t(v1)

if(use.cor)
{
   spec.risk <- tr * spec.risk
   fac.load <- tr * fac.load
   cov.mat <- tr * t(tr * cov.mat)
   inv.cov <- t(inv.cov / tr) / tr
}
```

```
   result <- new.env()
   result$spec.risk <- spec.risk
   result$fac.load <- fac.load
   result$fac.cov <- fac.cov
   result$cov.mat <- cov.mat
   result$inv.cov <- inv.cov
   result$pc <- x$vectors[, 1:ncol(fac.load)]
   result <- as.list(result)
   return(result)
}

qrm.erank.pc <- function (ret, use.cor = T, do.trunc = F,
      k = 0, excl.first = F)
{
   print("Running qrm.erank.pc()...")

   calc.erank <- function(x, excl.first)
   {
      take <- x > 0
      x <- x[take]
      if(excl.first)
         x <- x[-1]
      p <- x / sum(x)
      h <- - sum(p * log(p))
      er <- exp(h)
      if(excl.first)
         er <- er + 1
      return(er)
   }

   tr <- apply(ret, 1, sd)
   if(use.cor)
      ret <- ret / tr

   x <- t(ret)
   x <- var(x, x)
   tv <- diag(x)
   y <- eigen(x)
   if(k == 0)
   {
      er <- calc.erank(y$values, excl.first)

      if(do.trunc)
```

```
      k <- trunc(er)
   else
      k <- round(er)
}

k <- min(k, ncol(ret) - 2)
fac.load <- t(t(y$vectors[, 1:k]) * sqrt(y$values[1:k]))
fac.cov <- diag(1, k)
x.f <- fac.load %*% t(fac.load)
x.s <- tv - diag(x.f)
spec.risk <- sqrt(x.s)
cov.mat <- diag(x.s) + x.f

y.s <- 1 / spec.risk^2
v <- fac.load
v1 <- y.s * v
inv.cov <- diag(y.s) - v1 %*%
   solve(diag(1, k) + t(v) %*% v1) %*% t(v1)

if(use.cor)
{
   spec.risk <- tr * spec.risk
   fac.load <- tr * fac.load
   cov.mat <- tr * t(tr * cov.mat)
   inv.cov <- t(inv.cov / tr) / tr
}

result <- new.env()
result$spec.risk <- spec.risk
result$fac.load <- fac.load
result$fac.cov <- fac.cov
result$cov.mat <- cov.mat
result$inv.cov <- inv.cov
result$pc <- y$vectors[, 1:k]
result <- as.list(result)
return(result)
}
```

B R Code for Eigenpairs via Power Iterations

In this appendix we give the R source code for calculating the first M eigenpairs of the sample correlation matrix Ψ_{ij} based on the successive application of the power

iterations method as in Section 4. The code below is essentially self-explanatory and straightforward as it simply follows the formulas in Section 4. It consists of a single function `qrm.calc.eigen(ret, k, prec = 1e-3)`; `ret` is an $N \times (M+1)$ matrix of returns; N is the number of the underlying returns (e.g., alphas); $M+1$ is the number of data points in the time series (e.g., days); `k` is the number of the desired first k eigenpairs (in the decreasing order of the eigenvalues) to be computed (if $k > M$, only the first M eigenpairs are computed); `prec` is the convergence precision and is not the same as how close $\sum_{j=1}^{N} \Psi_{ij} V_j^{(a)} / \lambda^{(a)} V_i^{(a)}$ are to 1 or the approximate eigenvectors $V_i^{(a)}$ are to the true eigenvectors. That precision is lower owing to cumulative effects (due to sums, etc.). The output is a list: `result$count` is a k-vector whose elements are the numbers of iterations $n_{iter}^{(a)}$, $a = 1, \ldots, k$, for each eigenpair; `result$value` is a k-vector of the k eigenvalues; and `result$vector` is an $N \times k$ matrix whose columns are the k eigenvectors. The two lines `y[] <- rnorm(n, 0, 1)` and `y <- y / sqrt(sum(y^2))` are optional and used here for the purpose of generating randomness in the runs in Table 4; more economical ways of setting the initial iterations for the eigenpairs with $a > 1$ can be used.

```
qrm.calc.eigen <- function(ret, k, prec = 1e-3)
{
   pow.it.fac <- function(x, y, prec)
   {
      count <- 0
      repeat{
         count <- count + 1
         y.prev <- y
         y <- t(x) %*% y
         y <- x %*% y
         y <- y / sqrt(sum(y^2))
         if(max(abs(y/y.prev - 1)) < prec)
            break
      }
      result <- new.env()
      result$count <- count
      result$value <- sum((t(x) %*% y)^2)
      result$vector <- y
      return(result)
   }

   n <- nrow(ret)
   m <- ncol(ret)
   k <- min(k, m - 1)
   cor.mat <- cor(t(ret), t(ret))
```

```
   x <- ret - rowMeans(ret)
   s <- sqrt(rowSums(x^2))
   x <- x / s

   count <- value <- rep(NA, k)
   vector <- matrix(NA, n, k)
   y <- matrix(1/sqrt(n), n, 1)

   for(i in 1:k)
   {
      result <- pow.it.fac(x, y, prec)
      count[i] <- result$count
      value[i] <- result$value
      vector[, i] <- v <- result$vector

      for(j in 1:m)
         x[, j] <- x[, j] - v * sum(v * x[, j])

      y[] <- rnorm(n, 0, 1)
      y <- y / sqrt(sum(y^2))
   }

   result$count <- count
   result$value <- value
   result$vector <- vector
   return(result)
}
```

C R Code for Eigenpairs without Iterations

In this appendix we give the R source code for calculating the first M eigenpairs of the sample correlation matrix Ψ_{ij} based on the no-iterations method discussed in Subsection 4.2. The code below is essentially self-explanatory and straightforward as it simply follows the formulas in Subsection 4.2. It consists of a single function `qrm.calc.eigen.eff(ret, calc.cor = F)`; `ret` is an $N \times (M+1)$ matrix of returns; N is the number of the underlying returns; $M+1$ is the number of data points in the time series (e.g., days); for `calc.cor = F` (default), the code computes the eigenpairs for the covariance matrix; for `calc.cor = T` the code computes the eigenpairs for the correlation matrix. The method works only if $M \leq N$. The output is a list: `result$values` is an M-vector of the M eigenvalues; and `result$vectors` is an $N \times M$ matrix whose columns are the M eigenvectors.

```
qrm.calc.eigen.eff <- function (ret, calc.cor = F)
{
   calc.chol <- function(m)
   {
      x <- matrix(1, m, m)
      x <- upper.tri(x, T)
      z <- 1:m
      x <- x + diag(z)
      x <- x / sqrt(z * (z + 1))
      x <- t(x) / sqrt(m)
      return(x)
   }

   m <- ncol(ret) - 1
   n <- nrow(ret)

   if(m > n)
      stop("Too many observations...")

   if(calc.cor)
      ret <- ret / apply(ret, 1, sd)

   y <- ret - rowMeans(ret)
   y <- y[, -(m + 1)]
   p <- calc.chol(m)
   y <- y %*% p
   q <- t(y) %*% y
   q <- eigen(q)
   q.vec <- q$vectors
   q.val <- q$values
   q.vec <- t(t(q.vec) / sqrt(q.val))
   y <- y %*% q.vec

   result <- new.env()
   result$values <- q.val
   result$vectors <- y
   return(result)
}
```

D DISCLAIMERS

Wherever the context so requires, the masculine gender includes the feminine and/or neuter, and the singular form includes the plural and *vice versa*. The author of this paper ("Author") and his affiliates including without limitation Quantigic® Solutions LLC ("Author's Affiliates" or "his Affiliates") make no implied or express warranties or any other representations whatsoever, including without limitation implied warranties of merchantability and fitness for a particular purpose, in connection with or with regard to the content of this paper including without limitation any code or algorithms contained herein ("Content").

The reader may use the Content solely at his/her/its own risk and the reader shall have no claims whatsoever against the Author or his Affiliates and the Author and his Affiliates shall have no liability whatsoever to the reader or any third party whatsoever for any loss, expense, opportunity cost, damages or any other adverse effects whatsoever relating to or arising from the use of the Content by the reader including without any limitation whatsoever: any direct, indirect, incidental, special, consequential or any other damages incurred by the reader, however caused and under any theory of liability; any loss of profit (whether incurred directly or indirectly), any loss of goodwill or reputation, any loss of data suffered, cost of procurement of substitute goods or services, or any other tangible or intangible loss; any reliance placed by the reader on the completeness, accuracy or existence of the Content or any other effect of using the Content; and any and all other adversities or negative effects the reader might encounter in using the Content irrespective of whether the Author or his Affiliates is or are or should have been aware of such adversities or negative effects.

The R code included in Appendix A, Appendix B and Appendix C hereof is part of the copyrighted R code of Quantigic® Solutions LLC and is provided herein with the express permission of Quantigic® Solutions LLC. The copyright owner retains all rights, title and interest in and to its copyrighted source code included in Appendix A, Appendix B and Appendix C hereof and any and all copyrights therefor.

References

Acharya, V.V. and Pedersen, L.H. (2005) Asset pricing with liquidity risk. *Journal of Financial Economics* 77(2): 375-410.

Ang, A., Hodrick, R., Xing, Y. and Zhang, X. (2006) The Cross-Section of Volatility and Expected Returns. *Journal of Finance* 61(1): 259-299.

Anson, M. (2013/14) Performance Measurement in Private Equity: Another Look at the Lagged Beta Effect. *The Journal of Private Equity* 17(1): 29-44.

Asness, C.S. (1995) The Power of Past Stock Returns to Explain Future Stock Returns. Goldman Sachs Asset Management. Working paper.

Asness, C. and Stevens, R. (1995) Intra- and Inter-Industry Variation in the Cross-Section of Expected Stock Returns. Goldman Sachs Asset Management. Working paper.

Asness, C., Krail, R.J. and Liew, J.M. (2001) Do Hedge Funds Hedge? *The Journal of Portfolio Management* 28(1): 6-19.

Bai, J. (2003) Inferential theory for factor models of large dimension. *Econometrica* 71(11): 135-171.

Bai, J. and Li, K. (2012) Statistical Analysis of Factor Models of High Dimension. *The Annals of Statistics* 40(1): 436-465.

Bai, J. and Ng, S. (2002) Determining the number of factors in approximate factor models. *Econometrica* 70(1): 191-221.

Bansal, R. and Viswanathan, S. (1993) No Arbitrage and Arbitrage Pricing: A New Approach. *The Journal of Finance* 48(4): 1231-1262.

Banz, R. (1981) The relationship between return and market value of common stocks. *Journal of Financial Economics* 9(1): 3-18.

Basu, S. (1977) The investment performance of common stocks in relation to their price to earnings ratios: A test of the efficient market hypothesis. *Journal of Finance* 32(3): 663-682.

Black, F. (1972) Capital market equilibrium with restricted borrowing. *Journal of Business* 45(3): 444-455.

Black, F., Jensen, M. and Scholes, M. (1972) The capital asset pricing model: Some empirical tests. In: Jensen, M. (ed.) *Studies in the Theory of Capital Markets*. New York, NY: Praeger Publishers, pp. 79-121.

Blume, O. and Friend, L. (1973) A new look at the capital asset pricing model. *Journal of Finance* 28(1): 19-33.

Bouchaud, J.-P. and Potters, M. (2011) Financial applications of random matrix theory: a short review. In: Akemann, G., Baik, J. and Di Francesco, P. (eds.) *The Oxford Handbook of Random Matrix Theory.* Oxford, United Kingdom: Oxford University Press.

Brandt, M.W., Brav, A., Graham, J.R. and Kumar, A. (2010) The idiosyncratic volatility puzzle: Time trend or speculative episodes? *Review of Financial Studies* 23(2): 863-899.

Briner, B. and Connor, G. (2008) How much structure is best? A comparison of market model, factor model and unstructured equity covariance matrices. *Journal of Risk* 10(4), 3-30.

Burmeister, E. and Wall, K.D. (1986) The arbitrage pricing theory and macroeconomic factor measures. *Financial Review* 21(1): 1-20.

Campbell, J. (1987) Stock returns and the term structure. *Journal of Financial Economics* 18(2): 373-399.

Campbell, J.Y., Lettau, M., Malkiel, B.G. and Xu, Y. (2001) Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *Journal of Finance* 56(1): 1-43.

Campbell, J. and Shiller, R. (1988) The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1(3): 195-227.

Campbell, L.L. (1960) Minimum coefficient rate for stationary random processes. *Information and Control* 3(4): 360-371.

Carhart, M.M. (1997) Persistence in mutual fund performance. *Journal of Finance* 52(1): 57-82.

Chamberlain, G. and Rothschild, M. (1983) Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica* 51(5): 1281-1304.

Chan, K.C., Chen, N. and Hsieh, D. (1985) An Exploratory Investigation of the Firm Size Effect. *Journal of Financial Economics* 14(3): 451-471.

Chen, N., Grundy, B. and Stambaugh, R.F. (1990) Changing Risk, Changing Risk Premiums, and Dividend Yield Effects. *The Journal of Business* 63(1): 51-70.

Chen, N., Roll, R. and Ross, S. (1986) Economic forces and the stock market. *Journal of Business* 59(3): 383-403.

Chicheportiche, R. and Bouchaud, J.-P. (2015) A nested factor model for non-linear dependencies in stock returns. *Quantitative Finance* 15(11): 1789-1804.

Cochrane, J.H. (2001) *Asset Pricing.* Princeton, NJ: Princeton University Press.

Connor, G. (1984) A unified beta pricing theory. *Journal of Economic Theory* 34(1): 13-31.

Connor, G. (1995) The Three Types of Factor Models: A Comparison of Their Explanatory Power. *Financial Analysts Journal* 51(3): 42-46.

Connor, G. and Korajczyk, R. (1988) Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics* 21(2): 255-289.

Connor, G. and Korajczyk, R. (1989) An intertemporal beta pricing model. *Review of Financial Studies* 2(3): 373-392.

Connor, G. and Korajczyk, R. (2010) Factor Models in Portfolio and Asset Pricing Theory. In: Guerard Jr, J.B. (ed.) *Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques*. New York, NY: Springer, pp. 401-418.

Daniel, K. and Titman, S. (1997) Evidence on the Characteristics of Cross Sectional Variation in Stock Returns. *Journal of Finance* 52(1): 1-33.

DeBondt, W. and Thaler, R. (1985) Does the stock market overreact? *Journal of Finance* 40(3): 739-805.

Dhrymes, P.J., Friend, I. and Gultekin, N.B. (1984) A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory. *The Journal of Finance* 39(2): 323-346.

Fama, E. and French, K. (1992) The cross-section of expected stock returns. *Journal of Finance* 47(2): 427-465.

Fama, E.F. and French, K.R. (1993) Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* 33(1): 3-56.

Fama, E. and French, K. (1996) Multifactor explanations for asset pricing anomalies. *Journal of Finance* 51(1): 55-94.

Fama, E. and French, K. (2015) A Five-Factor Asset Pricing Model. *Journal of Financial Economics* (forthcoming), DOI: 10.1016/j.jfineco.2014.10.010.

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

Ferson, W. and Harvey, C. (1991) The variation in economic risk premiums. *Journal of Political Economy* 99(2): 385-415.

Ferson, W. and Harvey, C. (1999) Conditioning variables and the cross section of stock returns. *Journal of Finance* 54(4): 1325-1360.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000) The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics* 82(4): 540-554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005) The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* 100(471): 830-840.

Forni, M. and Lippi, M. (2001) The generalized dynamic factor model: Representation theory. *Econometric Theory* 17(6): 1113-1141.

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.

Goyal, A. and Santa-Clara, P. (2003) Idiosyncratic risk matters! *Journal of Finance* 58(3): 975-1007.

Grinold, R.C. and Kahn, R.N. (2000) *Active Portfolio Management.* New York, NY: McGraw-Hill.

Gubernatis, J.E. and Booth, T.E. (2008) Multiple Extremal Eigenpairs by the Power Method. *Journal of Computational Physics* 227(19): 8508-8522.

Hall, A.D., Hwang, S. and Satchell, S.E. (2002) Using bayesian variable selection methods to choose style factors in global stock return models. *Journal of Banking and Finance* 26(12): 2301-2325.

Haugen, R.A. (1995) *The New Finance: The Case Against Efficient Markets.* Upper Saddle River, NJ: Prentice Hall.

Heaton, J. and Lucas, D.J. (1999) Stock Prices and Fundamentals. *NBER Macroeconomics Annual* 14(1): 213-242.

Heston, S.L. and Rouwenhorst, K.G. (1994) Does Industrial Structure Explain the Benefits of International Diversification? *Journal of Financial Economics* 36(1): 3-27.

Hong, H., Torous, W. and Valkanov, R. (2007) Do industries lead stock markets? *Journal of Financial Economics* 83(2): 367-396.

Jagannathan, R. and Wang, Z. (1996) The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51(1): 3-53.

Jegadeesh, N. and Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48(1): 65-91.

Jegadeesh, N. and Titman, S. (2001) Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *Journal of Finance* 56(2): 699-720.

Kakushadze, Z. (2015a) Mean-Reversion and Optimization. *Journal of Asset Management* 16(1): 14-40.
Available online: http://ssrn.com/abstract=2478345.

Kakushadze, Z. (2015b) 4-Factor Model for Overnight Returns. *Wilmott Magazine* 2015(79): 56-62. Available online: http://ssrn.com/abstract=251187.

Kakushadze, Z. (2015c) Russian-Doll Risk Models. *Journal of Asset Management* 16(3): 170-185. Available online: http://ssrn.com/abstract=2538123.

Kakushadze, Z. (2015d) Heterotic Risk Models. *Wilmott Magazine* 2015(80): 40-55. Available online: http://ssrn.com/abstract=2600798.

Kakushadze, Z. (2016) Shrinkage = Factor Model. *Journal of Asset Management* 17(2): 69-72. Available online: http://ssrn.com/abstract=2685720.

Kakushadze, Z. and Liew, J.K.-S. (2015) Custom v. Standardized Risk Models. *Risks* 3(2): 112-138. Available online: http://ssrn.com/abstract=2493379.

Kakushadze, Z. and Yu, W. (2016) Multifactor Risk Models and Heterotic CAPM. *The Journal of Investment Strategies* 5(4): 1-49. Available online: http://ssrn.com/abstract=2722093.

King, B.F. (1966) Market and Industry Factors in Stock Price Behavior. *Journal of Business* 39(1): 139-190.

Korajczyk, R.A. and Sadka, R. (2008) Pricing the Commonality Across Alternative Measures of Liquidity. *Journal of Financial Economics* 87(1): 45-72.

Kothari, S. and Shanken, J. (1997) Book-to-market, dividend yield and expected market returns: A time series analysis. *Journal of Financial Economics* 44(2): 169-203.

Lakonishok, J., Shleifer, A. and Vishny, R.W. (1994) Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance* 49(5): 1541-1578.

Lee, J.-H. and Stefek, D. (2008) Do Risk Factors Eat Alphas? *The Journal of Portfolio Management* 34(4): 12-24.

Ledoit, O. and Wolf, M. (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10(5): 603-621.

Ledoit, O. and Wolf, M. (2004) Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management* 30(4): 110-119.

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

Liew, J. and Vassalou, M. (2000) Can Book-to-Market, Size and Momentum be Risk Factors that Predict Economic Growth? *Journal of Financial Economics* 57(2): 221-245.

Lintner, J. (1965) The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47(1): 13-37.

Lo, A.W. (2010) *Hedge Funds: An Analytic Perspective.* Princeton, NJ: Princeton University Press.

Lo, A.W. and MacKinlay, A.C. (1990) Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies* 3(3): 431-468.

MacKinlay, A.C. (1995) Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics* 38(1): 3-28.

MacQueen, J. (2003) The structure of multifactor equity risk models. *Journal of Asset Management* 3(4) 313-322.

Markowitz, H.M. (1952) Portfolio Selection. *Journal of Finance* 7(1): 77-91.

Markowitz, H.M. (1984) The Two-Beta Trap. *Journal of Portfolio Management* 11(1): 12-19.

Menchero, J. and Mitra, I. (2008) The Structure of Hybrid Factor Models. *Journal of Investment Management* 6(3): 35-47.

Merton, R. (1973) An intertemporal capital asset pricing model. *Econometrica* 41(5): 867-887.

Miller, G. (2006) Needles, Haystacks, and Hidden Factors. *The Journal of Portfolio Management* 32(2): 25-32.

Mises, R.V. and Pollaczek-Geiringer. H. "Praktische Verfahren der Gleichungsauflösung." ZAMM – Zeitschrift für Angewandte Mathematik und Mechanik, 9(2) (1929), pp. 152-164.

Motta, G., Hafner, C. and von Sachs, R. (2011) Locally stationary factor models: identification and nonparametric estimation. *Econometric Theory* 27(6): 1279-1319.

Mukherjee, D. and Mishra, A.K. (2005) Multifactor Capital Asset Pricing Model Under Alternative Distributional Specification. SSRN Working Papers Series, http://ssrn.com/abstract=871398 (December 29, 2005).

Ng, V., Engle, R.F. and Rothschild, M. (1992) A multi-dynamic-factor model for stock returns. *Journal of Econometrics* 52(1-2): 245-266.

Pastor, L. and Stambaugh, R.F. (2003) Liquidity Risk and Expected Stock Returns. *The Journal of Political Economy* 111(3): 642-685.

Rebonato, R. and Jäckel, P. (1999) The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *The Journal of Risk* 2(2): 17-27.

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

Rosenberg, B. (1974) Extra-Market Components of Covariance in Security Returns. *Journal of Financial and Quantitative Analysis* 9(2): 263-274.

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

Ross, S.A. (1978a) A Simple Approach to the Valuation of Risky Streams. *Journal of Business* 51(3): 453-475.

Ross, S.A. (1978b) Mutual Fund Separation in Financial Theory – The Separating Distributions. *Journal of Economic Theory* 17(2): 254-28

Roy, O. and Vetterli, M. (2007) The effective rank: A measure of effective dimensionality. In: *European Signal Processing Conference (EUSIPCO)*. Poznań, Poland (September 3-7, 2007), pp. 606-610.

Scholes, M. and Williams, J. (1977) Estimating Betas from Nonsynchronous Data. *Journal of Financial Economics* 5(3): 309-327.

Schwert, G. (1990) Stock returns and real activity: A century of evidence. *Journal of Finance* 45(4): 1237-1257.

Shanken, J. (1987) Nonsynchronous data and the covariance-factor structure of returns. *Journal of Finance* 42(2): 221-231.

Shanken, J. (1990) Intertemporal Asset Pricing: An Empirical Investigation. *Journal of Econometrics* 45(1-2): 99-120.

Shanken, J. and Weinstein, M.I. (2006) Economic Forces and the Stock Market Revisited. *Journal of Empirical Finance* 13(2): 129-144.

Sharpe, W.F. (1963) A simplified model for portfolio analysis. *Management Science* 9(2): 277-293.

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

Sharpe, W.F. (1994) The Sharpe Ratio. *The Journal of Portfolio Management* 21(1): 49-58.

Stock, J.H. and Watson, M.W. (2002a) Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20(2): 147-162.

Stock, J.H. and Watson, M.W. (2002b) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460): 1167-1179.

Stroyny, A.L. (2005) Estimating a Combined Linear Factor Model. In: Knight, J. and Satchell, S.E. (eds.) *Linear Factor Models in Finance.* Oxford: Elevier Butterworth-Heinemann.

Treynor, J.L. (1999) Towards a Theory of Market Value of Risky Assets. In: Korajczyk, R. (ed.) *Asset Pricing and Portfolio Performance: Models, Strategy, and Performance Metrics.* London: Risk Publications.

Vassalou, M. (2003) News Related to Future GDP Growth as a Risk Factor in Equity Returns. *Journal of Financial Economics.* 68(1): 47-73.

Whitelaw, R. (1997) Time variations and covariations in the expectation and volatility of stock market returns. *Journal of Finance* 49(2): 515-541.

Yang, W., Gibson, J.D. and He, T. (2005) Coefficient rate and lossy source coding. *IEEE Transactions on Information Theory* 51(1): 381-386.

Zangari, P. (2003) Equity factor risk models. In: Litterman, B. (ed.) *Modern Investment Management: An Equilibrium Approach.* New York, NY: John Wiley & Sons, Inc., pp. 334-395.

Zhang, C. (2010) A Re-examination of the Causes of Time-varying Stock Return Volatilities. *Journal of Financial and Quantitative Analysis* 45(3): 663-684.

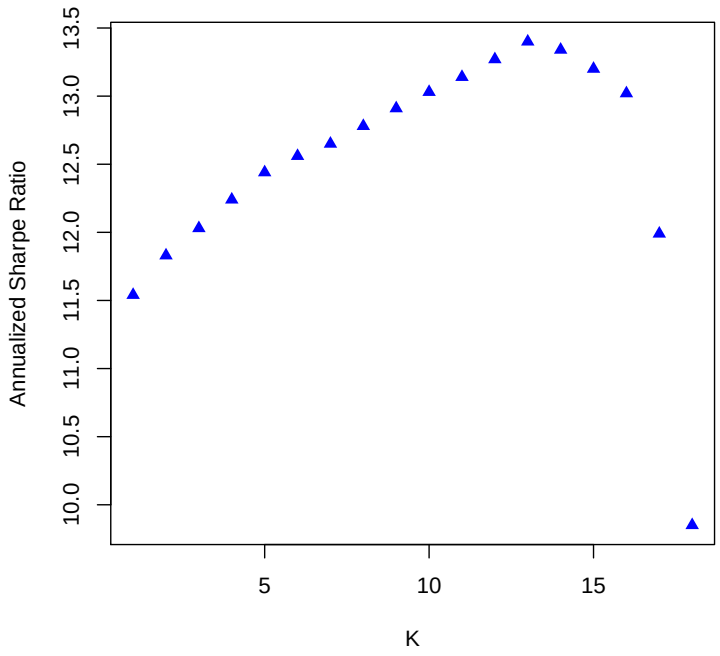

Figure 1. Graph of the values of the Sharpe ratio (SR) from Table 9 vs. the number of risk factors K (as defined in said table).

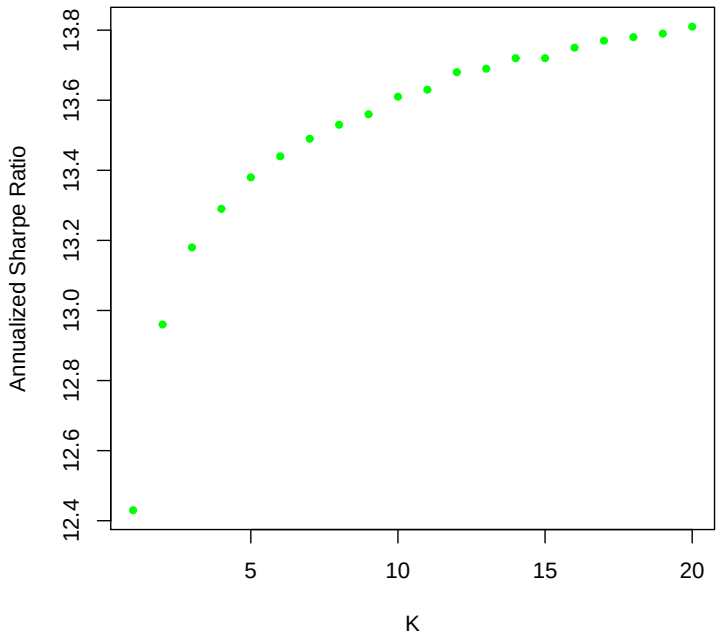

Figure 2. Graph of the values of the Sharpe ratio (SR) from Table 10 vs. the number of risk factors K (as defined in said table).

Table 1: Simulation results for the optimized alphas without bounds using the statistical risk models. See Subsection 3.4 for details. The result in the first row is the same as in (Kakushadze, 2015d). We label the risk models M1-M6 for notational convenience in Table 3.

Risk Model	ROC	SR	CPS
M1: K, "minimization" algorithm	47.74%	11.88	2.26
M2: $K = K' + 1$, "minimization" algorithm	46.87%	11.59	2.23
M3: $K = \text{Round}(\text{eRank}(\Psi))$	47.89%	11.20	2.28
M4: $K = \text{floor}(\text{eRank}(\Psi))$	47.13%	11.10	2.24
M5: $K = \text{Round}(\text{eRank}(\Psi')) + 1$	28.51%	4.67	1.32
M6: $K = \text{floor}(\text{eRank}(\Psi')) + 1$	32.46%	5.53	1.49

Table 2: Simulation results for the optimized alphas with bounds. See Subsection 3.4 for details. The result in the first row is the same as in (Kakushadze, 2015d).

Risk Model	ROC	SR	CPS
K, "minimization" algorithm	40.92%	14.33	1.96
$K = K' + 1$, "minimization" algorithm	40.36%	14.04	1.94
$K = \text{Round}(\text{eRank}(\Psi))$	40.78%	14.01	1.96
$K = \text{floor}(\text{eRank}(\Psi))$	40.84%	14.06	1.96

Table 3: Summaries of various quantities from out backtests. The M1-M6 models are defined in Table 1. 1st Qu. = 1st Quartile, 3rd Qu. = 3rd Quartile, StDev = standard deviation, MAD = mean absolute deviation. Summaries in the first 7 rows are computed based on the 60 data points corresponding to 60 21-trading-day intervals in our 1,260 trading-day backtesting interval (see Subsection 3.4.3). The "average correlation" $\overline{\Psi} = \frac{1}{N^2}\sum_{i,j=1}^{N}\Psi_{ij}$. The parameter $\zeta = \max(|H_{is}|/0.01\ A_{is})$, where $\max(\cdot)$ is cross-sectional for each date s (s takes 1,260 values) and the summary in the last row is taken over the 1,260 trading days.

Quantity	Min	1st Qu.	Median	Mean	3rd Qu.	Max	StDev	MAD
M1, K	6	11.75	12	12.17	13	14	1.57	1.48
M2, K	11	13	14	13.77	14	16	0.91	1.48
M3, K	3	10	12	11.93	14	17	3.31	2.97
M4, K	3	10	12	11.5	14	17	3.18	2.97
M5, K	17	19	19	18.92	19	19	0.33	0
M6, K	17	18	19	18.57	19	19	0.53	0
$\overline{\Psi}$ (%)	11.68	22.86	31.14	32.93	38.43	74.78	13.44	11.46
M6, ζ	3.26	16.66	25.22	29.66	37.39	189.5	19.19	14.35

Table 4: The number of iterations $n_{iter}^{(a)}$, $a = 1, \ldots, M$, required to compute the first $M = 19$ principal components of the sample correlation matrix based on a time series of $N = 2,339$ stock returns with $M + 1 = 20$ data points (trading days). There are 10 runs. In each run the initial iteration $[V_i^{(1)}]_{init}$ for the first principal component is taken as $[V_i^{(1)}]_{init} = u_i/\sqrt{N}$, where $u_i \equiv 1$ is the unit N-vector. For the other principal components the initial iterations are taken as random N-vectors with 0 mean and unit variance further normalized to have the quadratic norm 1, i.e., $\sum_{i=1}^{N} \left([V_i^{(a)}]_{init}\right)^2 = 1$, $a = 2, \ldots, M$. The mean (median) of the total iteration n_{iter}^{tot} (i.e., of the last row) is 5,037 (5,038). See Subsection 4 for details. We use the code in Appendix B to generate this table. The convergence precision is set to the default, `prec = 1e-3`. See Appendix B.

Run #:	1	2	3	4	5	6	7	8	9	10
$n_{iter}^{(1)}$	13	13	13	13	13	13	13	13	13	13
$n_{iter}^{(2)}$	24	24	25	24	23	24	25	23	23	22
$n_{iter}^{(3)}$	162	168	199	150	158	209	211	193	221	148
$n_{iter}^{(4)}$	181	165	167	211	219	158	160	207	185	166
$n_{iter}^{(5)}$	144	107	117	97	119	100	101	97	101	118
$n_{iter}^{(6)}$	457	298	371	287	387	387	392	359	426	448
$n_{iter}^{(7)}$	147	175	151	162	179	137	120	185	172	174
$n_{iter}^{(8)}$	430	309	372	443	361	440	437	366	395	412
$n_{iter}^{(9)}$	387	265	326	423	327	361	343	323	321	351
$n_{iter}^{(10)}$	222	209	217	209	273	229	316	208	210	234
$n_{iter}^{(11)}$	203	182	243	225	284	205	141	235	307	226
$n_{iter}^{(12)}$	261	242	235	288	365	296	286	284	259	252
$n_{iter}^{(13)}$	456	607	489	711	598	726	520	620	552	645
$n_{iter}^{(14)}$	1059	1029	1122	1089	1092	1296	1190	1169	1032	988
$n_{iter}^{(15)}$	307	403	392	371	441	438	364	437	392	409
$n_{iter}^{(16)}$	126	144	153	127	161	147	155	132	197	150
$n_{iter}^{(17)}$	156	192	203	157	206	177	167	171	197	177
$n_{iter}^{(18)}$	81	80	87	67	82	99	77	80	114	83
$n_{iter}^{(19)}$	2	2	2	2	2	2	2	2	2	2
n_{iter}^{tot}	4818	4614	4884	5056	5290	5444	5020	5104	5119	5018

Table 5: Simulation results for the optimized alphas (with bounds) based on shrinkage with a diagonal shrinkage target for various values of the shrinkage constant q. See Subsection 4.5 for details.

q	ROC	SR	CPS
10^{-6}	40.67%	13.84	1.81
0.3	40.66%	13.83	1.81
0.6	40.66%	13.82	1.81
0.9	40.65%	13.81	1.81

Table 6: Simulation results for the optimized alphas (with bounds) based on shrinkage with a uniform correlation shrinkage target for various values of the shrinkage constant q and the correlation ρ. See Subsection 4.5 for details.

q	ρ	ROC	SR	CPS
10^{-6}	0.1	41.03%	14.10	1.83
0.3	0.1	41.02%	14.09	1.83
0.6	0.1	41.01%	14.07	1.82
0.9	0.1	40.92%	13.92	1.82
10^{-6}	0.5	41.03%	14.11	1.83
0.3	0.5	41.03%	14.10	1.82
0.6	0.5	41.02%	14.09	1.82
0.9	0.5	40.98%	14.01	1.82
10^{-6}	0.9	41.03%	14.11	1.82
0.3	0.9	41.03%	14.11	1.82
0.6	0.9	41.03%	14.11	1.82
0.9	0.9	41.02%	14.09	1.82

Table 7: Simulation results for the optimized alphas without bounds for fixed values of the number of risk factors K. See Section 5 for details.

K	ROC	SR	CPS
1	46.69%	11.17	2.18
2	47.77%	11.69	2.24
3	47.83%	11.90	2.25
4	47.83%	11.90	2.26
5	48.26%	12.15	2.28
6	48.48%	11.89	2.29
7	48.73%	11.93	2.31
8	48.71%	12.02	2.31
9	48.68%	12.01	2.30
10	48.66%	12.16	2.30
11	48.21%	11.88	2.28
12	48.22%	11.79	2.29
13	47.54%	11.66	2.26
14	46.47%	10.98	2.21
15	45.90%	10.51	2.18
16	45.19%	10.56	2.15
17	45.96%	10.04	2.19
18	40.57%	6.98	1.89
19	27.78%	4.48	1.29

Table 8: Simulation results for the optimized alphas with bounds for fixed values of the number of risk factors K. See Section 5 for details.

K	ROC	SR	CPS
1	39.14%	12.97	1.85
2	40.20%	13.64	1.90
3	40.64%	13.97	1.93
4	40.78%	14.15	1.94
5	41.05%	14.30	1.96
6	41.07%	14.36	1.96
7	41.13%	14.36	1.96
8	41.19%	14.37	1.97
9	41.26%	14.40	1.97
10	41.37%	14.37	1.98
11	41.14%	14.32	1.97
12	41.19%	14.35	1.97
13	40.90%	14.31	1.96
14	40.29%	14.05	1.93
15	39.81%	13.71	1.91
16	38.96%	13.35	1.88
17	37.14%	12.11	1.81
18	30.96%	9.84	1.52

Table 9: Simulation results for the optimized alphas with bounds for $K_1 = M$ risk factors (K_1 is fixed) but the specific risks ad hoc based on the K-factor model (K varies). See Section 5 for details.

K	ROC	SR	CPS
1	37.53%	11.54	1.77
2	37.93%	11.83	1.80
3	38.20%	12.03	1.82
4	38.37%	12.24	1.83
5	38.70%	12.44	1.85
6	38.81%	12.56	1.86
7	38.97%	12.65	1.86
8	39.14%	12.78	1.87
9	39.35%	12.91	1.88
10	39.60%	13.03	1.89
11	39.58%	13.14	1.89
12	39.78%	13.27	1.91
13	39.79%	13.40	1.91
14	39.39%	13.34	1.89
15	39.18%	13.20	1.88
16	38.52%	13.02	1.86
17	36.99%	11.99	1.81
18	30.98%	9.85	1.52

Table 10: Simulation results for the optimized alphas with bounds for varying number K of risk factors and the specific risk ad hoc set to the in-sample risk. See Section 5 for details. We expect the $K = 20$ case to be the same as shrinkage with the shrinkage parameter $q \to 1$, which is why line 20 in this table is the same (within rounding, which in all tables herein for all non-integer quantities such as ROC, SR, CPS, etc., is to 2 decimal points) as the last line in Table 5.

K	ROC	SR	CPS
1	38.80%	12.43	1.73
2	39.73%	12.96	1.77
3	40.04%	13.18	1.79
4	40.25%	13.29	1.79
5	40.32%	13.38	1.80
6	40.39%	13.44	1.80
7	40.43%	13.49	1.80
8	40.48%	13.53	1.80
9	40.54%	13.56	1.81
10	40.58%	13.61	1.81
11	40.64%	13.63	1.81
12	40.67%	13.68	1.81
13	40.63%	13.69	1.81
14	40.65%	13.72	1.81
15	40.63%	13.72	1.81
16	40.65%	13.75	1.81
17	40.66%	13.77	1.81
18	40.64%	13.78	1.81
19	40.65%	13.79	1.81
20	40.65%	13.81	1.81