Multifactor Risk Models and Heterotic CAPM

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

(January 24, 2016)

ZK: To my children Mirabelle and Maximilien Kakushadze

WY: To my parents Albert and Ribena Yu

Abstract

We give a complete algorithm and source code for constructing general multifactor risk models (for equities) via any combination of style factors, principal components (betas) and/or industry factors. For short horizons we employ the Russian-doll risk model construction to obtain a nonsingular factor covariance matrix. This generalizes the heterotic risk model construction to include arbitrary non-industry risk factors as well as industry risk factors with generic "weights". The aim of sharing our proprietary know-how with the investment community is to encourage organic risk model building. The presentation is intended to be essentially self-contained and pedagogical. So, stop wasting money and complaining, start building risk models and enjoy![4]

---

[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.

[4] This is the last paper in the trilogy, which contains "Russian-Doll Risk Models" (Kakushadze, 2015c) and "Heterotic Risk Models" (Kakushadze, 2015d).

1 Introduction

Quantitative traders build their own alphas and trading strategies. However, only a small fraction of quantitative traders build their own risk models. How come? Especially considering that it is precisely the risk models – one way or another – that turn alphas (or expected returns, trading signals, etc.) into portfolio holdings.

A typical answer is "we focus on alphas",[5] "we do not have the required data", "we do not have time/human resources", etc. So, most quantitative traders end up buying off-the-shelf standardized commercial risk models with at least some understanding that what they are buying is not built for their kind of trading or tested on real-life alphas.[6] In fact, many quant traders even complain about this.

The prosaic answer to the above question is actually much simpler. Most quant traders do not know how to build high quality risk models from scratch. It takes time and experience to develop expertise for such model building and most quant traders indeed spend much of their time developing alphas. There are some quant shops that develop their own risk models, but they appear to be in the minority.

The purpose of this paper is to change the aforesaid landscape by giving a complete algorithm and source code for constructing general multifactor risk models[7] for equities[8] via any combination of style factors, principal components and/or industry factors. For short horizons we employ the Russian-doll risk model construc-

---

[5] To traders, alphas make money. Risk models just help make money. Hence the priorities.

[6] The majority of clients of the standardized commercial risk model providers are large institutionals such as mutual and pension funds that have no need for customization but require standardization for inter- and intra-institutional risk reporting, etc. (Kakushadze and Liew, 2015).

[7] For a partial list of works related to factor risk models, 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), (Kakushadze and Liew, 2015), (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.

[8] Albeit the general methods we discuss here can be applied to other instruments as well.

---

tion (Kakushadze, 2015c) to obtain a nonsingular factor covariance matrix. This generalizes the heterotic risk model construction (Kakushadze, 2015d) to include arbitrary non-industry as well as industry risk factors with generic "weights". By sharing our proprietary know-how we hope to encourage organic risk model building.

The main difficulty in dealing with generic risk factors is that using the standard "lore" for computing the factor covariance matrix and specific (idiosyncratic) risk, which is done via a (weighted) regression, yields total variances for stocks which do not agree with sample variances. The exception is when the risk factors are based on principal components of the sample covariance (or correlation) matrix or its sub-matrices corresponding to "clusters" such as sub-industries as in the heterotic risk model construction. In this case the total risks automatically work out. However, if we include, e.g., style factors, this method produces incorrect total variances, which is often overlooked in practice. Our algorithm rectifies this shortcoming by design.

In Section 2 we review multifactor models, the requirement that a factor model reproduce sample variances, and why it is (not) satisfied for principal component (general) risk factors. In Section 3 we discuss industry risk factors and the heterotic risk model construction. In Section 4 we give our algorithm for dealing with general risk factors. However, the paper does not end there. This is because for short historical lookbacks, which is the case in short-horizon models, the sample factor covariance matrix is singular and itself needs to be modeled. This, in turn, is due to the fact that the number of industry factors is too large (in three digits). In Section 5 we apply the Russian-doll construction to general factor models and give a complete embedding resulting in a nonsingular factor covariance matrix and consequently, an invertible multifactor model covariance matrix. We also explain why adding a few style or other non-industry risk factors (e.g., principal components) to ubiquitous industry risk factors adds little to no value.[9] We further illustrate this via out-of-sample backtests using intraday mean-reversion alphas. In Section 6 we introduce risk models we refer to as "heterotic CAPM", which are constructed similarly to the heterotic risk models, but the "weights" for industry risk factors are based on style factors instead of principal components. In this novel, nontraditional implementation some style factors do add value. We briefly conclude in Section 7. Appendix A contains R source code,[10] which implements our algorithm for general risk models, the complete Russian-doll embedding, style factor computation, etc.

---

[9] And this is how style factors are traditionally treated including in standardized commercial risk models. These offerings aver that style factors do add value. However, there are a few caveats there. First, in such models the industry factors are not taken at the most granular level and consequently are not as numerous, typically, around 60 or fewer, so augmenting them with 10 or so style factors a priori may add some value. Second, these models are not necessarily tested on real-life trading alphas. Third, they are geared toward longer horizons, i.e., lower turnover strategies with lower Sharpe ratios. As we discuss in Section 5, adding industry classification granularity increases Sharpe ratios. (In fact, for short-horizon models there are fewer (4 or so) style factors.)

[10] The source code given in Appendix A 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. Important legalese relating to this code is given in Appendix B.

2

2 Multifactor Risk Models

2.1 Sample Covariance Matrix

So, we have N stocks and the associated time series of N returns $R_i(t_s)$, $i = 1, \ldots, N$, where t_s, $s = 0, 1, 2, \ldots, M$ ($s = 0$ corresponds to the most recent time), are the $M + 1$ points in time in our time series (i.e., we have $M + 1$ observations for each return). E.g., we could consider daily close-to-close returns, albeit the frequency is not critical here. Let C_{ij} be the $N \times N$ sample covariance matrix computed based on these time series of returns. When $M < N$, the sample covariance matrix C_{ij} is singular with M positive eigenvalues.[11] In this case we cannot invert C_{ij}, which is required in 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 Total Variances

In this regard, a lot of headache (see below) can be avoided by making the following simple observation. The main goal 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 risk model had better reproduce the in-sample variances. That is, we require that the risk model variances Γ_{ii} coincide with the in-sample variances C_{ii}:

$$\Gamma_{ii} = C_{ii} \tag{1}$$

Furthermore, as mentioned above, the N variances C_{ii} are relatively stable out-of-sample. It is therefore (a half of) 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 $\Psi_{ij} = C_{ij}/\sqrt{C_{ii}}\sqrt{C_{jj}}$ 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 – call it $\widetilde{\Gamma}_{ij}$ – that is much more stable out-of-sample and invertible (positive-definite) subject to the conditions

$$\widetilde{\Gamma}_{ii} = \Psi_{ii} \equiv 1 \tag{2}$$

Once we build $\widetilde{\Gamma}_{ij}$, the risk model covariance matrix is given by $\Gamma_{ij} = \sqrt{C_{ii}}\sqrt{C_{jj}}\,\widetilde{\Gamma}_{ij}$.

---

[11] The positive eigenvalue count can be lower if some returns are 100% (anti-)correlated.

2.3 Factor Models

A popular way of constructing Γ_{ij} is via a factor model. However, here we wish to construct $\widetilde{\Gamma}_{ij}$, not Γ_{ij}. Nonetheless, the factor model construction applies just as well. Instead of the returns $R_i(t_s)$ we simply use the "normalized" returns $\widetilde{R}_i(t_s) = R_i(t_s)/\sqrt{C_{ii}}$. Then Ψ_{ij} is the sample covariance matrix of these normalized returns: $\Psi_{ij} = \text{Cov}(\widetilde{R}_i(t_s), \widetilde{R}_j(t_s)) = \text{Cor}(R_i(t_s), R_j(t_s))$, where $\text{Cov}(\cdot, \cdot)$ and $\text{Cor}(\cdot, \cdot)$ are serial. I.e., we build a factor model for Ψ_{ij} by treating $\widetilde{R}_i(t_s)$ as returns.

Then, we have N random processes Υ_i, which model the returns $\widetilde{R}_i$ (we omit t_s dependence for notational convenience):

$$\Upsilon_i = \chi_i + \sum_{A=1}^{K} \Omega_{iA} \; f_A \tag{3}$$

$$\text{Cov}(\chi_i, \chi_j) = \Xi_{ij} = \xi_i^2 \; \delta_{ij} \tag{4}$$

$$\text{Cov}(\chi_i, f_A) = 0 \tag{5}$$

$$\text{Cov}(f_A, f_B) = \Phi_{AB} \tag{6}$$

$$\text{Cov}(\Upsilon_i, \Upsilon_j) = \widetilde{\Gamma}_{ij} \tag{7}$$

where (in matrix notation)

$$\widetilde{\Gamma} = \Xi + \Omega \; \Phi \; \Omega^T \tag{8}$$

and δ_{ij} is the Kronecker delta; ξ_i is the specific risk (a.k.a. idiosyncratic risk) for each stock; Ω_{iA} is an $N \times K$ factor loadings matrix; and Φ_{AB} is a $K \times K$ factor covariance matrix, $A, B = 1, \dots, K$, where $K \ll N$. I.e., the random processes Υ_i are modeled via N random processes χ_i (specific risk) together with K random processes f_A (factor risk). Assuming all $\xi_i > 0$ and Φ_{AB} is positive-definite, then $\widetilde{\Gamma}_{ij}$ is automatically positive-definite (and invertible). Furthermore, $\widetilde{\Gamma}_{ij}$ is expected to be much more stable out-of-sample than Ψ_{ij} as the number of risk factors, for which the factor covariance matrix Φ_{AB} needs to be computed, is $K \ll N$.

Since $\widetilde{R}_i$ are normalized, so are the specific risk and factor risk; indeed, (2) gives the following N conditions[12] for $N + K(K+1)/2$ unknowns:

$$\xi_i^2 + \sum_{A,B=1}^{K} \Omega_{iA} \; \Phi_{AB} \; \Omega_{iB} \equiv 1 \tag{9}$$

There are no "natural" $K(K+1)/2$ conditions we can impose in terms of out-of-sample unstable off-diagonal correlations Ψ_{ij}, $i \neq j$. So, we need additional assumptions to compute ξ_i and Φ_{AB}. *Intuitively,* one such assumption is that the total risk should be attributed to the factor risk to the greatest extent possible, i.e., the part of the total risk attributed to the specific risk should be minimized.

---

[12] With additional assumptions not all of these conditions are independent (see below).

4

2.4 Linear Regression

One way to formulate this requirement mathematically is via least squares. Thus, mimicking (3), we decompose the returns $\widetilde{R}_i$ via a linear model

$$\widetilde{R}_i = \epsilon_i + \sum_{A=1}^{K} \Omega_{iA} \, f_A \tag{10}$$

Here the residuals ϵ_i are *not* the same as χ_i in (3); in particular, generally the covariance matrix $\text{Cov}(\epsilon_i, \epsilon_j)$ is not diagonal (see below). We can require that

$$\sum_{i=1}^{N} \epsilon_i^2 \to \min \tag{11}$$

where the minimization is w.r.t. f_A. This produces a linear regression.[13] The solution to (11) is given by (in matrix notation)

$$f = \left(\Omega^T \, \Omega \right)^{-1} \Omega^T \, \widetilde{R} \tag{12}$$

$$\epsilon = [1 - Q] \, \widetilde{R} \tag{13}$$

$$Q = \Omega \left(\Omega^T \, \Omega \right)^{-1} \Omega^T \tag{14}$$

where $Q = Q^T$ is a projection operator: $Q^2 = Q$. Consequently, we have:

$$\widehat{\Xi} = \text{Cov} \left(\epsilon, \epsilon^T \right) = [1 - Q] \, \Psi \, [1 - Q] \tag{15}$$

$$\Omega \, \Phi \, \Omega^T = \Omega \, \text{Cov} \left(f, f^T \right) \Omega^T = Q \, \Psi \, Q \tag{16}$$

Note that the matrix $\widehat{\Xi}$ is not diagonal. However, the idea is to identify ξ_i^2 with the diagonal part of $\widehat{\Xi}$:

$$\xi_i^2 = \widehat{\Xi}_{ii} = ([1 - Q] \, \Psi \, [1 - Q])_{ii} \tag{17}$$

and we have

$$\widetilde{\Gamma}_{ij} = \xi_i^2 \, \delta_{ij} + (Q \, \Psi \, Q)_{ij} \tag{18}$$

Note that ξ_i^2 defined via (17) are automatically positive (nonnegative, to be precise – see below). However, we must satisfy the conditions (2), which reduce to

$$X_{ii} = 0 \tag{19}$$

$$X = 2 \, Q \, \Psi \, Q - Q \, \Psi - \Psi \, Q \tag{20}$$

The N conditions (19) are not all independent. Thus, we have $\text{Tr}(X) = 0$.

---

[13] Without the intercept, that is, unless the intercept is already subsumed in Ω_{iA}. Also, this is a regression with unit weights. Since the returns $\widetilde{R}_i$ are normalized, there are no "natural" out-of-sample stable candidates for the regression weights excepting the unit weights. This is the beauty of factoring out the sample volatilities $\sqrt{C_{ii}}$: the sample correlation matrix Ψ_{ij} knows nothing about volatility, hence – naturally – unit weights. Also, note that for a large enough universe the sample volatilities have a skewed – roughly log-normal – distribution cross-sectionally. In contrast, Ψ_{ii} are nicely uniform. This makes (9) easier to satisfy (see below).

2.5 Principal Components

The conditions (19) are nontrivial. They are not satisfied for an arbitrary factor loadings matrix Ω_{iA}. Thus, consider the simplest case of a single risk factor ($K = 1$), so that the loadings matrix Ω_{iA} consists of a single column, call it ω_i. Without loss of generality we can normalize it such that $\sum_{i=1}^N \omega_i^2 = 1$. Further, without loss of generality we can assume that all $\omega_i \neq 0$ – indeed, for any vanishing ω_i factor risk vanishes and we only have specific risk, the matrix $\widetilde{\Gamma}_{ij}$ is block-diagonal, and we can ignore the block corresponding to the vanishing ω_i. Then (19) reduces to

$$\sum_{j=1}^N \Psi_{ij}\, \omega_j = \omega_i \sum_{j,k=1}^N \Psi_{jk}\, \omega_j\, \omega_k \tag{21}$$

This invariably implies that ω_i is an eigenvector (principal component) of Ψ_{ij}.

2.5.1 Multiple Principal Components

More generally, (19) is automatically satisfied if we take the columns of the factor loadings matrix to be, say, the first K of the N principal components $V_i^{(a)}$, $a = 1, \ldots, N$, of Ψ_{ij} forming an orthonormal basis

$$\sum_{j=1}^N \Psi_{ij}\, V_j^{(a)} = \lambda^{(a)}\, V_i^{(a)} \tag{22}$$

$$\sum_{i=1}^N V_i^{(a)}\, V_i^{(b)} = \delta_{ab} \tag{23}$$

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[14] is M, where, as above, $M+1$ is the number of observations in the stock return time series. We can readily construct a factor model with $K < M$:[15]

$$\Omega_{iA} = \sqrt{\lambda^{(A)}}\, V_i^{(A)}, \quad A = 1, \ldots, K \tag{24}$$

Then the factor covariance matrix $\Phi_{AB} = \delta_{AB}$ and we have

$$\widetilde{\Gamma}_{ij} = \xi_i^2\, \delta_{ij} + \sum_{A=1}^K \lambda^{(A)}\, V_i^{(A)}\, V_j^{(A)} \tag{25}$$

$$\xi_i^2 = 1 - \sum_{A=1}^K \lambda^{(A)} \left(V_i^{(A)}\right)^2 \tag{26}$$

---

[14] This number can be smaller if some stock 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.

[15] For $K = M$ we have $\widetilde{\Gamma} = \Psi$, which is singular. Null eigenvalues do not contribute into (25).

6

so $\widetilde{\Gamma}_{ii} = \Psi_{ii} \equiv 1$. See (Kakushadze, 2015d) for source code for building a risk model based on principal components (including for fixing the value of K).[16]

An evident limitation of the principal component approach is that the number of risk factors is limited by M. If long lookbacks are unavailabe/undesirable, as, e.g., in short-holding quantitative trading strategies, then typically $M \ll N$. Yet, the number of the actually relevant underlying risk factors can be substantially greater than M, and most of these risk factors are missed by this approach.[17] Thus, it is unclear what to do with the principal components with null eigenvalues. They do not contribute to any sample factor covariance matrix. We need additional input.

3 Industry Risk Factors

While the number of risk factors based on principal components is limited, especially for shorter lookbacks,[18] risk factors based on a granular enough industry classification can be ubiquitous. Furthermore, they are independent of the pricing data – industry classification is based on fundamental/economic data (such as companies' products and services and more generally their revenue sources, suppliers, competitors, partners, etc.) – and, in this regard, are essentially insensitive to the lookback. In fact, typically industry classification based risk factors tend to be rather stable out-of-sample as companies seldom jump industries (let alone sectors).

For terminological definiteness, here we will use the BICS[19] nomenclature for the levels in the industry classification, albeit this is not critical here. Also, BICS has three levels "sector $\rightarrow$ industry $\rightarrow$ sub-industry" (where "sub-industry" is the most granular level). The number of levels in the industry hierarchy is not critical here either. So, we have: N stocks labeled by $i = 1, \ldots, N$; K sub-industries labeled by $A = 1, \ldots, K$; F industries labeled by $a = 1, \ldots, F$;[20] and L sectors labeled by $\alpha = 1, \ldots, L$. More generally, we can think of such groupings as "clusters". Sometimes, loosely, we will refer to such "cluster" based factors as "industry" factors.[21]

3.1 "Binary" Property

The binary property implies that each stock belongs to one and only one sub-industry, industry and sector (or, more generally, "cluster"). Let G be the map

---

[16] Note that $\xi_i^2 = \sum_{a=K+1}^{M} \lambda^{(a)} \left(V_i^{(a)} \right)^2 \geq 0$. We are assuming $\lambda^{(a)} \geq 0$, which (up to computational precision) is the case if there are no N/As in the stock return times series.

[17] Another limitation is that the principal components are based on off-diagonal elements of Ψ_{ij} and tend to be unstable out-of-sample, the first principal component typically being the most stable. In many applications one uses only the first principal component (see below).

[18] The number of style factors is also limited, of order 10 or fewer (see below).

[19] Bloomberg Industry Classification System.

[20] The subscript index a should not be confused with the superscript (a) in (22).

[21] Albeit in the BICS context we may be referring to, e.g., sub-industries, while in other classification schemes the actual naming may be altogether different.

between stocks and sub-industries, S be the map between sub-industries and industries, and W be the map between industries and sectors:

$$G : \{1, \ldots, N\} \mapsto \{1, \ldots, K\} \tag{27}$$

$$S : \{1, \ldots, K\} \mapsto \{1, \ldots, F\} \tag{28}$$

$$W : \{1, \ldots, F\} \mapsto \{1, \ldots, L\} \tag{29}$$

The beauty of the binary property is that the "clusters" (sub-industries, industries and sectors) can be used to identify blocks (sub-matrices) in the correlation matrix Ψ_{ij}. E.g., for sub-industries the binary matrix $\delta_{G(i),A}$ defines such blocks.

3.2 Binary Models

Consider the following factor loadings matrix:

$$\Omega_{iA} = \frac{1}{\sqrt{N_A}} \, \delta_{G(i),A} \tag{30}$$

where $J(A) = \{i | G(i) = A\}$ is the set of tickers (whose number $N(A) = |J(A)|$) in the sub-industry labeled by A. However, the factor loadings (30) do not satisfy the conditions (19). I.e., we cannot identify the factor covariance matrix with the sample covariance matrix of the regression coefficients and the specific risks with the variances of the regression residuals. There is a way around this difficulty. What if we continue to identify the factor covariance matrix with the sample covariance matrix of the regression coefficients and *define* the specific risks via (as opposed to (17))

$$\xi_i^2 = 1 - (Q \ \Psi \ Q)_{ii} \tag{31}$$

so we still have (18) and (2)? The issue with the definition (31) is that for a general factor loadings matrix Ω_{iA} some of the resulting ξ_i^2 unacceptably are *negative*. However, for the binary factor loadings (30) ξ_i^2 are nonnegative. Indeed, we have

$$\xi_i^2 = 1 - \frac{1}{N^2(G(i))} \sum_{k,l \in J(G(i))} \Psi_{kl} \geq 1 - \frac{1}{N^2(G(i))} \sum_{k,l \in J(G(i))} |\Psi_{kl}| \geq 0 \tag{32}$$

In fact, since we are assuming that we have no 100% (anti-)correlated pairs of stocks, the only way we can have $\xi_i^2 = 0$ is for single-ticker sub-industries ($N(A) = 1$). These, in fact, are not problematic (see below). So, for binary factor loadings we can construct a factor model this way; however, with a caveat: if $M < K$, which is often the case, the factor covariance matrix

$$\Phi_{AB} = \frac{1}{\sqrt{N_A N_B}} \sum_{i \in J(A)} \sum_{j \in J(B)} \Psi_{ij} \tag{33}$$

is singular. We will address this issue below via the Russian-doll risk model construction. But first let us discuss heterotic risk models (Kakushadze, 2015d).

8

3.3 Heterotic Models

Consider the following factor loadings matrix:

$$\Omega_{iA} = \delta_{G(i),A} \, U_i \tag{34}$$
$$U_i = [U(A)]_i, \quad i \in J(A), \quad A = 1, \ldots, K \tag{35}$$

where the $N(A)$-vector $U(A)$ is the *first* principal component of the $N(A) \times N(A)$ matrix $\Psi(A)$ defined via $[\Psi(A)]_{ij} = \Psi_{ij}, i, j \in J(A)$. (Note that $\sum_{i \in J(A)}[U(A)]_i^2 = 1$; also, let the corresponding (largest) eigenvalue of $\Psi(A)$ be $\lambda(A)$.)[22] With this factor loadings matrix we can compute the factor covariance matrix and specific risk via the above linear regression:[23]

$$\xi_i^2 = 1 - \lambda(G(i)) \, U_i^2 \tag{36}$$
$$\widetilde{\Gamma}_{ij} = \left[1 - \lambda(G(i)) \, U_i^2\right] \delta_{ij} + U_i \, U_j \sum_{k \in J(G(i))} \sum_{l \in J(G(j))} U_k \, \Psi_{kl} \, U_l \tag{37}$$

and (19) is automatically satisfied. This simplicity is due to the use of the (first) principal components corresponding to the blocks $\Psi(A)$ of the sample correlation matrix.[24] However, here too, as in the binary models, for short lookbacks we have $M < K$, so (note that $\Phi_{AA} = \lambda(A)$) the factor covariance matrix

$$\Phi_{AB} = \sum_{i \in J(A)} \sum_{j \in J(B)} U_i \, \Psi_{ij} \, U_j \tag{38}$$

is singular. This again brings us to the nested Russian-doll risk model construction we discuss below. However, first we need to introduce one more ingredient.

4 How to Deal with General Risk Factors?

In heterotic risk models, and also when the factor loadings matrix is just a collection of principal components, the conditions (19) are automatically satisfied. In binary risk models (19) is not satisfied, but we found a workaround by simply defining the specific risk via (31). However, in more general cases this trick does not do. E.g., if we combine heterotic or binary risk factors with some style factors (see below) and/or principal components, (19) is not satisfied, and (31) generally will produce ξ_i^2 at least some of which are negative. However, not all is lost. Thanks to modeling

---

[22] If $N(A) = 1$, i.e., we have only one ticker in the sub-industry labeled by A, then $[U(A)]_i = 1$ and $\lambda(A) = \Psi_{ii} = 1, i \in J(A)$.

[23] For single-ticker sub-industries ($N(A) = 1$) the specific risk vanishes: $\xi_i^2 = 0$; however, this does not pose a problem as this does not cause the matrix $\widetilde{\Gamma}_{ij}$ to be singular (see below).

[24] In fact, we can generalize this construction by using multiple principle components per "cluster"; however, out-of-sample instability of the higher principal components mentioned above poses an issue. See (Kakushadze, 2015d) for details.

the *correlation* matrix Ψ_{ij} via $\widetilde{\Gamma}_{ij}$ (as opposed to the covariance matrix C_{ij} via Γ_{ij}), if we stick to the definition (17) via the linear regression, then (19) is not satisfied for general Ω_{iA}, but the violations of the condition (2) are of order 1 as opposed to some large numbers.[25] Then we can simply rescale the matrix $\widetilde{\Gamma}_{ij}$ such that (2) is satisfied, i.e., we can define $\widetilde{\Gamma}_{ij}$ via (as opposed to (18))

$$\widetilde{\Gamma}_{ij} = \frac{1}{\gamma_i \gamma_j} \left[\xi_i^2 \ \delta_{ij} + \sum_{A,B=1}^{K} \Omega_{iA} \ \Phi_{AB} \ \Omega_{jB} \right] = \frac{1}{\gamma_i \gamma_j} \left[\xi_i^2 \ \delta_{ij} + (Q \ \Psi \ Q)_{ij} \right] \quad (39)$$

$$\gamma_i^2 = \xi_i^2 + (Q \ \Psi \ Q)_{ii} \quad (40)$$

where, as before, ξ_i^2 is defined via (17) and is expressly nonnegative. This is a simple but powerful prescription that does the job: now we have (2). (The *actual* factor loadings and specific risks are $\widehat{\Omega}_{iA} = \Omega_{iA}/\gamma_i$ and $\widehat{\xi}_i = \xi_i/\gamma_i$.) However, we still need to deal with the factor covariance matrix Φ_{AB} being singular for short lookbacks.

5 Russian-Doll Construction

5.1 General Idea

The simple idea behind the Russian-doll construction is to model such Φ_{AB} itself via yet another factor model matrix Γ'_{AB} (as opposed to computing it as a sample covariance matrix of the risk factor returns f_A):[26]

$$\Gamma'_{AB} = \sqrt{\Phi_{AA}} \sqrt{\Phi_{BB}} \ \widetilde{\Gamma}'_{AB} \quad (41)$$

$$\widetilde{\Gamma}'_{AB} = \frac{1}{\gamma'_A \gamma'_B} \left[(\xi'_A)^2 \ \delta_{AB} + \sum_{a,b=1}^{F} \Omega'_{Aa} \ \Phi'_{ab} \ \Omega'_{Bb} \right] \quad (42)$$

$$(\gamma'_A)^2 = (\xi'_A)^2 + \sum_{a,b=1}^{F} \Omega'_{Aa} \ \Phi'_{ab} \ \Omega'_{Ab} \quad (43)$$

where ξ'_A is the specific risk for the "normalized" factor return $\widetilde{f}_A = f_A/\sqrt{\Phi_{AA}}$; Ω'_{Aa}, $A = 1, \ldots, K$, $a = 1, \ldots, F$ is the corresponding factor loadings matrix; and Φ'_{ab} is the factor covariance matrix for the underlying risk factors f'_a, $a = 1, \ldots, F$, where we assume that $F \ll K$. If the smaller factor covariance matrix Φ'_{ab} is still singular, we model it via yet another factor model with fewer risk factors, and so on – until the resulting factor covariance matrix is nonsingular. If, at the final stage, we are left with a single factor, then the resulting 1×1 factor covariance matrix is automatically nonsingular – it is simply the sample variance of the remaining factor.

---

[25] Had we tried to do the same in the context of Γ_{ij}, the violations of the condition (1) would be much more dramatic and nonuniform owing to the fact that the variances C_{ii} have a highly skewed (quasi-log-normal) distribution. It is therefore more convenient to model Ψ_{ij} via $\widetilde{\Gamma}_{ij}$.

[26] We use a prime on $\widetilde{\Gamma}'_{AB}$, γ'_A, ξ'_A, Φ'_{ab}, etc., to avoid confusion with $\widetilde{\Gamma}_{ij}$, γ_i, ξ_i, Φ_{AB}, etc.

5.1.1 When We Have Industry Factors Only

Moving forward we will need to distinguish between industry and non-industry[27] factors. Examples of the latter are style factors and principal components. When we have only industry factors, we can use the natural hierarchy in the industry classification to implement the Russian-doll construction. For concreteness we will use the BICS terminology for the levels in the industry classification, albeit this is not critical here. Also, BICS has three levels "sector $\rightarrow$ industry $\rightarrow$ sub-industry" (where "sub-industry" is the most granular level). For definiteness, we will assume three levels here, albeit the generalization to more levels is straightforward. So, in the BICS terminology, we can use sub-industries as the risk factors for stocks, industries as the risk factors for sub-industries, sectors as the risk factors for industries, and (if need be) the "market" as the single risk factor for sectors (Kakushadze, 2015c,d). That is, a binary industry classification provides a natural nested hierarchy of risk factors. The binary models (30) and the heterotic models (34) discussed above both contain only industry factors.[28]

5.1.2 Including Non-Industry Factors

When we have both industry and non-industry factors, things are trickier. Thus, for non-industry risk factors it is not always as straightforward to identify a nested hierarchy of risk factors. E.g., for the principal component based risk factors there is no evident guiding principle to do so. However, not all is lost. In practice, the number of relevant style factors typically is substantially smaller than the number of industry factors (especially for (ultra-)short horizon models (Kakushadze, 2015b)), of order 10 or fewer. The number of principal components one may wish to include is also limited, either because of a short lookback, or out-of-sample instability of the higher principal components, or both. Thanks to this, we can start with a few non-industry (style and/or principal component) factors plus ubiquitous industry factors (typically, ~ 100 or more), and build a Russian-doll risk model.

The idea here is simple: there is no need to reduce the number of already-few non-industry factors, only that of the ubiquitous industry factors. That is, we apply the sequential nested Russian-doll embedding to the industry factors only, leaving the non-industry factors intact. At the final stage, we will have to compute the factor covariance matrix for the remaining factors, which will include all non-industry factors (along with, e.g., the sectors or the "market").

We will use the mid-Greek symbols $\mu, \nu, \ldots$ to label the non-industry risk factors, and we use the i, A, a, α labels as above to label stocks, sub-industries, industries and sectors. Let $\widetilde{A} = (A, \mu)$, $\widetilde{a} = (a, \mu)$ and $\widetilde{\alpha} = (\alpha, \mu)$. Let Y be the number of non-binary risk factors. Let $\widetilde{K} = K + Y$, $\widetilde{F} = F + Y$ and $\widetilde{L} = L + Y$. We will also use the index $\widetilde{\mu} = (0, \mu)$ with $\widetilde{Y} = Y + 1$ values, where "0" labels the "market".

---

[27] Here "non-industry" means "not based on the industry classification".

[28] In the heterotic risk models the columns of the factor loadings matrix are "weighted" by the first principal components in the sub-industries, industries, sectors and (if need be) "market".

We then have:

$$\Omega_{iA} = \omega_i \; \delta_{G(i),A} \tag{44}$$

$$\Omega_{i\widetilde{A}} = (\Omega_{iA}, \Omega_{i\mu}) \tag{45}$$

$$\Omega'_{Aa} = \omega'_A \; \delta_{S(A),a} \tag{46}$$

$$\Omega'_{\widetilde{A}\widetilde{a}} = \mathrm{diag}\,(\Omega'_{Aa}, \delta_{\mu\nu}) \tag{47}$$

$$\Omega''_{a\alpha} = \omega''_a \; \delta_{W(a),\alpha} \tag{48}$$

$$\Omega''_{\widetilde{a}\widetilde{\alpha}} = \mathrm{diag}\,(\Omega''_{a\alpha}, \delta_{\mu\nu}) \tag{49}$$

$$\Omega'''_{\alpha 0} = \omega'''_\alpha \tag{50}$$

$$\Omega'''_{\widetilde{\alpha}\widetilde{\mu}} = \mathrm{diag}\,(\Omega'''_{\alpha 0}, \delta_{\mu\nu}) \tag{51}$$

where: the maps G (tickers to sub-industries), S (sub-industries to industries) and W (industries to sectors) are defined in (27), (28) and (29); the columns $\Omega_{i\mu}$ of the factor loadings matrix $\Omega_{i\widetilde{A}}$ correspond to the non-industry factors; and we will specify the "weights" ω_i, ω'_A, ω''_a and ω'''_α for the industry factors below.

5.2   Complete Russian-Doll Embedding

We will continue to use the BICS terminology for the levels in the industry classification, albeit this is not critical here. Also, BICS has three levels "sector $\rightarrow$ industry $\rightarrow$ sub-industry" (where "sub-industry" is the most granular level). For definiteness, we will assume three levels here, and the generalization to more levels is straightforward. A nested Russian-doll risk model then is constructed as follows:

$$\Gamma_{ij} = \sqrt{C_{ii}}\sqrt{C_{jj}}\; \widetilde{\Gamma}_{ij} \tag{52}$$

$$\widetilde{\Gamma}_{ij} = \widehat{\xi}_i^2 \; \delta_{ij} + \sum_{\widetilde{A},\widetilde{B}=1}^{\widetilde{K}} \widehat{\Omega}_{i\widetilde{A}} \; \Gamma'_{\widetilde{A}\widetilde{B}} \; \widehat{\Omega}_{j\widetilde{B}} \tag{53}$$

$$\Gamma'_{\widetilde{A}\widetilde{B}} = \sqrt{\Phi_{\widetilde{A}\widetilde{A}}}\sqrt{\Phi_{\widetilde{B}\widetilde{B}}}\; \widetilde{\Gamma}'_{\widetilde{A}\widetilde{B}} \tag{54}$$

$$\widetilde{\Gamma}'_{\widetilde{A}\widetilde{B}} = (\widehat{\xi}'_{\widetilde{A}})^2 \; \delta_{\widetilde{A}\widetilde{B}} + \sum_{\widetilde{a},\widetilde{b}=1}^{\widetilde{F}} \widehat{\Omega}'_{\widetilde{A}\widetilde{a}} \; \Gamma''_{\widetilde{a}\widetilde{b}} \; \widehat{\Omega}'_{\widetilde{B}\widetilde{b}} \tag{55}$$

$$\Gamma''_{\widetilde{a}\widetilde{b}} = \sqrt{\Phi'_{\widetilde{a}\widetilde{a}}}\sqrt{\Phi'_{\widetilde{b}\widetilde{b}}}\; \widetilde{\Gamma}''_{\widetilde{a}\widetilde{b}} \tag{56}$$

$$\widetilde{\Gamma}''_{\widetilde{a}\widetilde{b}} = (\widehat{\xi}''_{\widetilde{a}})^2 \; \delta_{\widetilde{a}\widetilde{b}} + \sum_{\widetilde{\alpha},\widetilde{\beta}=1}^{\widetilde{L}} \widehat{\Omega}''_{\widetilde{a}\widetilde{\alpha}} \; \Gamma'''_{\widetilde{\alpha}\widetilde{\beta}} \; \widehat{\Omega}''_{\widetilde{b}\widetilde{\beta}} \tag{57}$$

$$\Gamma'''_{\widetilde{\alpha}\widetilde{\beta}} = \sqrt{\Phi''_{\widetilde{\alpha}\widetilde{\alpha}}}\sqrt{\Phi''_{\widetilde{\beta}\widetilde{\beta}}}\; \widetilde{\Gamma}'''_{\widetilde{\alpha}\widetilde{\beta}} \tag{58}$$

$$\widetilde{\Gamma}'''_{\widetilde{\alpha}\widetilde{\beta}} = (\widehat{\xi}'''_{\widetilde{\alpha}})^2 \; \delta_{\widetilde{\alpha}\widetilde{\beta}} + \sum_{\widetilde{\mu},\widetilde{\nu}=1}^{\widetilde{Y}} \widehat{\Omega}'''_{\widetilde{\alpha}\widetilde{\mu}} \; \Phi'''_{\widetilde{\mu}\widetilde{\nu}} \; \widehat{\Omega}'''_{\widetilde{\mu}\widetilde{\nu}} \tag{59}$$

12

where

$$\widehat{\xi}_i = \xi_i/\gamma_i \tag{60}$$

$$\widehat{\Omega}_{i\widetilde{A}} = \Omega_{i\widetilde{A}}/\gamma_i \tag{61}$$

$$\gamma_i^2 = \xi_i^2 + (Q\ \Psi\ Q)_{ii} \tag{62}$$

$$\xi_i^2 = ([1-Q]\ \Psi\ [1-Q])_{ii} \tag{63}$$

$$Q = \Omega\left(\Omega^T\Omega\right)^{-1}\Omega^T \tag{64}$$

$$\widehat{\xi}'_{\widetilde{A}} = \xi'_{\widetilde{A}}/\gamma'_{\widetilde{A}} \tag{65}$$

$$\widehat{\Omega}'_{\widetilde{A}\widetilde{a}} = \Omega'_{\widetilde{A}\widetilde{a}}/\gamma'_{\widetilde{A}} \tag{66}$$

$$(\gamma'_{\widetilde{A}})^2 = (\xi'_{\widetilde{A}})^2 + (Q'\ \Psi'\ Q')_{\widetilde{A}\widetilde{A}} \tag{67}$$

$$(\xi'_{\widetilde{A}})^2 = ([1-Q']\ \Psi'\ [1-Q'])_{\widetilde{A}\widetilde{A}} \tag{68}$$

$$Q' = \Omega'\left(\Omega'^T\Omega'\right)^{-1}\Omega'^T \tag{69}$$

$$\widehat{\xi}''_{\widetilde{a}} = \xi''_{\widetilde{a}}/\gamma''_{\widetilde{a}} \tag{70}$$

$$\widehat{\Omega}''_{\widetilde{a}\widetilde{\alpha}} = \Omega''_{\widetilde{a}\widetilde{\alpha}}/\gamma''_{\widetilde{a}} \tag{71}$$

$$(\gamma''_{\widetilde{a}})^2 = (\xi''_{\widetilde{a}})^2 + (Q''\ \Psi''\ Q'')_{\widetilde{a}\widetilde{a}} \tag{72}$$

$$(\xi''_{\widetilde{a}})^2 = ([1-Q'']\ \Psi''\ [1-Q''])_{\widetilde{a}\widetilde{a}} \tag{73}$$

$$Q'' = \Omega''\left(\Omega''^T\Omega''\right)^{-1}\Omega''^T \tag{74}$$

$$\widehat{\xi}'''_{\widetilde{\alpha}} = \xi'''_{\widetilde{\alpha}}/\gamma'''_{\widetilde{\alpha}} \tag{75}$$

$$\widehat{\Omega}'''_{\widetilde{\alpha}\widetilde{\mu}} = \Omega'''_{\widetilde{\alpha}\widetilde{\mu}}/\gamma'''_{\widetilde{\alpha}} \tag{76}$$

$$(\gamma'''_{\widetilde{\alpha}})^2 = (\xi'''_{\widetilde{\alpha}})^2 + (Q'''\ \Psi'''\ Q''')_{\widetilde{\alpha}\widetilde{\alpha}} \tag{77}$$

$$(\xi'''_{\widetilde{\alpha}})^2 = ([1-Q''']\ \Psi'''\ [1-Q'''])_{\widetilde{\alpha}\widetilde{\alpha}} \tag{78}$$

$$Q''' = \Omega'''\left(\Omega'''^T\Omega'''\right)^{-1}\Omega'''^T \tag{79}$$

In (64) Ω is the $N \times \widetilde{K}$ matrix $\Omega_{i\widetilde{A}}$; in (69) Ω' is the $\widetilde{K} \times \widetilde{F}$ matrix $\Omega'_{\widetilde{A}\widetilde{a}}$; in (74) Ω'' is the $\widetilde{F} \times \widetilde{L}$ matrix $\Omega''_{\widetilde{a}\widetilde{\alpha}}$; and in (79) Ω''' is the $\widetilde{L} \times \widetilde{Y}$ matrix $\Omega'''_{\widetilde{\alpha}\widetilde{\mu}}$. Also,

$$\Phi = \left(\Omega^T\ \Omega\right)^{-1}\Omega^T\Psi\ \Omega\left(\Omega^T\ \Omega\right)^{-1} \tag{80}$$

$$\Phi' = \left(\Omega'^T\Omega'\right)^{-1}\Omega'^T\Psi'\ \Omega'\left(\Omega'^T\Omega'\right)^{-1} \tag{81}$$

$$\Phi'' = \left(\Omega''^T\Omega''\right)^{-1}\Omega''^T\Psi''\ \Omega''\left(\Omega''^T\Omega''\right)^{-1} \tag{82}$$

$$\Phi''' = \left(\Omega'''^T\Omega'''\right)^{-1}\Omega'''^T\Psi'''\ \Omega'''\left(\Omega'''^T\Omega'''\right)^{-1} \tag{83}$$

$$\Psi_{ij} = C_{ij}/\sqrt{C_{ii}}\sqrt{C_{jj}} \tag{84}$$

$$\Psi'_{\widetilde{A}\widetilde{B}} = \Phi_{\widetilde{A}\widetilde{B}}/\sqrt{\Phi_{\widetilde{A}\widetilde{A}}}\sqrt{\Phi_{\widetilde{B}\widetilde{B}}} \tag{85}$$

$$\Psi''_{\widetilde{a}\widetilde{b}} = \Phi'_{\widetilde{a}\widetilde{b}}/\sqrt{\Phi'_{\widetilde{a}\widetilde{a}}}\sqrt{\Phi'_{\widetilde{b}\widetilde{b}}} \tag{86}$$

$$\Psi'''_{\widetilde{\alpha}\widetilde{\beta}} = \Phi''_{\widetilde{\alpha}\widetilde{\beta}}/\sqrt{\Phi''_{\widetilde{\alpha}\widetilde{\alpha}}}\sqrt{\Phi''_{\widetilde{\beta}\widetilde{\beta}}} \tag{87}$$

13

Finally, the "weights" ω_i, ω'_A, ω''_a and ω'''_α in (44), (46), (48) and (50) are give by

$$\omega_i = [U(A)]_i, \quad i \in J(A) \tag{88}$$

$$\omega'_A = [U'(a)]_A, \quad A \in J'(a) \tag{89}$$

$$\omega''_a = [U''(\alpha)]_a, \quad a \in J''(\alpha) \tag{90}$$

Here $J(A) = \{i|G(i) = A\}$ ($N_A = |J(A)|$ tickers in sub-industry A), $J'(a) = \{A|S(A) = a\}$ ($N'(a) = |J'(a)|$ sub-industries in industry a), $J''(\alpha) = \{a|W(a) = \alpha\}$ ($N''(\alpha) = |J''(\alpha)|$ industries in sector α). The "market" contains all N tickers. The $N(A)$-vector $U(A)$ is the first principal component of $\Psi(A)$ with the eigenvalue $\lambda(A)$ ($[\Psi(A)]_{ij} = \Psi_{ij}$, $i, j \in J(A)$); the $N'(a)$-vector $U'(a)$ is the first principal component of $\Psi'(a)$ with the eigenvalue $\lambda'(a)$ ($[\Psi'(a)]_{AB} = \Psi'_{AB}$, $A, B \in J'(a)$); the $N''(\alpha)$-vector $U''(\alpha)$ is the first principal component of $\Psi''(\alpha)$ with the eigenvalue $\lambda''(\alpha)$ ($[\Psi''(\alpha)]_{ab} = \Psi''_{ab}$, $a, b \in J''(\alpha)$); finally, ω'''_α is the first principal component of $\Psi'''_{\alpha\beta}$ with the eigenvalue λ'''. The vectors $U(A)$, $U'(a)$ and $U''(\alpha)$ are normalized, so $\sum_{i \in J(A)} U_i^2 = 1$, $\sum_{A \in J'(a)} (U'_A)^2 = 1$, $\sum_{a \in J''(\alpha)} (U''_a)^2 = 1$, and also $\sum_{\alpha=1}^L (\omega'''_\alpha)^2 = 1$.

For the sake of completeness, above we included the step where the sample factor covariance matrix $\Phi''_{\widetilde{\alpha}\widetilde{\beta}}$ for the sectors plus non-industry factors is further approximated via a $\widetilde{Y}$-factor model $\widetilde{\Gamma}'''_{\widetilde{\alpha}\widetilde{\beta}}$, where $\widetilde{Y} = Y + 1$, i.e., the L risk factors corresponding to the sectors are modeled via a 1-factor model (the "market"), while the Y non-industry risk factors are untouched (see above). If $\Phi''_{\widetilde{\alpha}\widetilde{\beta}}$ computed via (82) is nonsingular, then this last step can be omitted,[29] so at the last stage we have $\widetilde{L} = L + Y$ factors (as opposed to $\widetilde{Y}$ factors).[30] Similarly, if we have enough observations to compute the sample covariance matrix $\Phi'_{\widetilde{a}\widetilde{b}}$ for the industries plus non-industry factors, we can stop at that stage. Finally, note that in the above construction we are guaranteed to have $(\widetilde{\xi}'''_\alpha)^2 > 0$, $(\widetilde{\xi}''_a)^2 > 0$, $(\widetilde{\xi}'_A)^2 > 0$ and $\widetilde{\xi}_i^2 \geq 0$. The last equality occurs only for single-ticker sub-industries (see below).

5.3 Model Covariance Matrix and Its Inverse

The model covariance matrix is given by Γ_{ij} defined in (52). For completeness, let us present it in the "canonical" form:

$$\Gamma_{ij} = \widetilde{\xi}_i^2 \, \delta_{ij} + \sum_{\widetilde{A}, \widetilde{B}=1}^{\widetilde{K}} \widetilde{\Omega}_{i\widetilde{A}} \, \Phi^*_{\widetilde{A}\widetilde{B}} \, \widetilde{\Omega}_{j\widetilde{B}} \tag{91}$$

---

[29] That is, assuming there are enough observations in the time series for out-of-sample stability.

[30] For the sake of completeness, the definitions of the factor returns at each stage are as follows (in matrix notation): (i) for the sub-industry plus non-industry factor returns $f_{\widetilde{A}}$: $f = (\Omega^T \Omega)^{-1} \Omega^T \widetilde{R}$, where $\widetilde{R}_i = R_i/\sqrt{C_{ii}}$; (ii) for the industry plus non-industry factor returns $f'_{\widetilde{a}}$: $f' = (\Omega'^T \Omega')^{-1} \Omega'^T \widetilde{f}$, where $\widetilde{f}_{\widetilde{A}} = f_{\widetilde{A}}/\sqrt{\Phi_{\widetilde{A}\widetilde{A}}}$; (iii) for the sector plus non-industry factor returns $f''_{\widetilde{\alpha}}$: $f'' = (\Omega''^T \Omega'')^{-1} \widetilde{f}'$, where $\widetilde{f}'_{\widetilde{a}} = f'_{\widetilde{a}}/\sqrt{\Phi'_{\widetilde{a}\widetilde{a}}}$; and (iv) for the "market" plus non-industry factor returns $f'''_{\widetilde{\mu}}$: $f''' = (\Omega'''^T \Omega''')^{-1} \Omega'''^T \widetilde{f}''$, where $\widetilde{f}''_{\widetilde{\alpha}} = f''_{\widetilde{\alpha}}/\sqrt{\Phi''_{\widetilde{\alpha}\widetilde{\alpha}}}$.

14

where

$$\widetilde{\xi}_i^2 = C_{ii} \, \widehat{\xi}_i^2 \tag{92}$$

$$\widetilde{\Omega}_{i\widetilde{A}} = \sqrt{C_{ii}} \, \widehat{\Omega}_{i\widetilde{A}} \tag{93}$$

$$\Phi^*_{\widetilde{A}\widetilde{B}} = \Gamma'_{\widetilde{A}\widetilde{B}} \tag{94}$$

where $\widehat{\xi}_i^2$ is defined in (60), U_i is defined in (61), $\Gamma'_{\widetilde{A}\widetilde{B}}$ is defined in (54), and we use the star superscript in the factor covariance matrix $\Phi^*_{\widetilde{A}\widetilde{B}}$ (which is nonsingular) to distinguish it from the sample factor covariance matrix $\Phi_{\widetilde{A}\widetilde{B}}$ (which is singular).

In many applications, such as portfolio optimization, one needs the inverse of the matrix Γ. When we have no single-ticker sub-industries, the inverse is given by (in matrix notation)

$$\Gamma^{-1} = \widetilde{\Xi}^{-1} - \widetilde{\Xi}^{-1} \, \widetilde{\Omega} \, \Delta^{-1} \, \widetilde{\Omega}^T \, \widetilde{\Xi}^{-1} \tag{95}$$

$$\Delta = (\Phi^*)^{-1} + \widetilde{\Omega}^T \, \widetilde{\Xi}^{-1} \, \widetilde{\Omega} \tag{96}$$

$$\widetilde{\Xi} = \mathrm{diag}(\widetilde{\xi}_i^2) \tag{97}$$

However, when there are some single-ticker sub-industries, the corresponding $\widetilde{\xi}_i^2 = 0$, $i \in H$, where $H = \{i | N(G(i)) = 1\}$, so (95) "breaks". Here is a simple "fix". We can rewrite Γ_{ij} via

$$\Gamma_{ij} = \overline{\xi}_i^2 \, \delta_{ij} + \sum_{\widetilde{A},\widetilde{B}=1}^{\widetilde{K}} \widetilde{\Omega}_{i\widetilde{A}} \, \overline{\Phi}^*_{\widetilde{A}\widetilde{B}} \, \widetilde{\Omega}_{j\widetilde{B}} \tag{98}$$

where: $\overline{\xi}_i^2 = \widetilde{\xi}_i^2$ for $i \notin H$; $\overline{\xi}_i^2 = \sum_{A=1}^{K} \widetilde{\Omega}_{iA}^2 \, \Phi^*_{AA}$ for $i \in H$; and $\overline{\Phi}^*_{AA} = 0$ for $A \in E$, where $E = \{A | N(A) = 1\}$, whereas for all other values of $\widetilde{A}$ and $\widetilde{B}$ we have $\overline{\Phi}^*_{\widetilde{A}\widetilde{B}} = \Phi^*_{\widetilde{A}\widetilde{B}}$. Now we can invert Γ via

$$\Gamma^{-1} = \overline{\Xi}^{-1} - \overline{\Xi}^{-1} \, \widetilde{\Omega} \, \overline{\Delta}^{-1} \, \widetilde{\Omega}^T \, \overline{\Xi}^{-1} \tag{99}$$

$$\overline{\Delta} = (\overline{\Phi}^*)^{-1} + \widetilde{\Omega}^T \, \overline{\Xi}^{-1} \, \widetilde{\Omega} \tag{100}$$

$$\overline{\Xi} = \mathrm{diag}(\overline{\xi}_i^2) \tag{101}$$

Note that, due to the factor model structure, to invert the $N \times N$ matrix Γ, we only need to invert two $\widetilde{K} \times \widetilde{K}$ matrices $\overline{\Phi}^*$ and $\overline{\Delta}$. If there are no single-ticker sub-industries, then Φ^* itself has a factor model structure and involves inverting two $\widetilde{F} \times \widetilde{F}$ matrices, one of which has a factor model structure, and so on.

5.4 Do Non-industry Factors Add Value?

Above we described an algorithm for constructing general multifactor risk models containing both industry and non-industry based risk factors. More precisely, for the industry based risk factors we chose the "weights" ω_i, ω'_A, ω''_a and ω'''_α in (44), (46),

15

(48) and (50) as the first principal components of the correlation (sub-)matrices for the corresponding "clusters" (sub-industries, industries, sectors, "market"). However, these "weights" can be chosen at will, e.g., we could have chosen some or all of them to be binary (i.e., as in (30)). In fact, these choices need not even be uniform across different "clusters", e.g., we could choose binary "weights" for some sub-industries and first (or higher) principal components for other sub-industries. It is a separate issue whether the resulting model would perform better.

The question we address in this subsection is whether non-industry factors actually add value. An off-the-cuff answer might appear to be that they do, at least for style factors.[31] However, this "intuition" based on longer-horizon models does us no good for short-lookback models we are after here. And there is a simple way to test this, to wit, by utilizing the same intraday mean-reversion alphas as in (Kakushadze, 2015d). In the remainder of this subsection, which very closely follows most parts of Section 6 of (Kakushadze, 2015d),[32] we describe the backtesting procedure.

5.4.1 Notations

Let P_{is} be the time series of stock prices, where $i = 1, \ldots, N$ labels the stocks, and $s = 0, 1, \ldots, M$ labels the trading dates, with $s = 0$ 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:

$$E_{is} = \ln \left(P_{is}^{AO} / P_{i,s+1}^{AC} \right) \tag{102}$$

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

$$R_{is} = \ln \left(P_{is}^{AC} / P_{i,s+1}^{AC} \right) \tag{103}$$

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[33] with fills at the open prices P_{is}^O; 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_{is}^C; 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

---

[31] If the non-industry factors are based on principal components, one may have a priori doubts whether they add value considering out-of-sample stability issues (see above).

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

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

Electronic copy available at: https://ssrn.com/abstract=2722093

and what does not. The P&L for each stock

$$\Pi_{is} = H_{is} \left[\frac{P_{is}^C}{P_{is}^O} - 1 \right] \tag{104}$$

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 $Q_{is} = 2|H_{is}|/P_{is}^O$.

5.4.2 Universe Selection

For the sake of simplicity,[34] 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):

$$A_{is} = \frac{1}{d} \sum_{r=1}^{d} V_{i,s+r} \ P_{i,s+r}^C \tag{105}$$

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 sector, industry and sub-industry assignments as of 9/6/2014.[35] However, as discussed in detail in Section 7 of (Kakushadze, 2015a), the survivorship bias is not a leading effect in such backtests.[36]

5.4.3 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 (in cents, not dollars) divided by the total shares traded.[37]

---

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

[35] 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.

[36] 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.

[37] 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

5.4.4 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 Γ_{ij} given by (52), 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:

$$\mathcal{S} = \frac{\sum_{i=1}^{N} H_i \ E_i}{\sqrt{\sum_{i,j=1}^{N} \Gamma_{ij} \ H_i \ H_j}} \to \max \tag{106}$$

$$\sum_{i=1}^{N} H_i = 0 \tag{107}$$

The solution is given by

$$H_i = -\gamma \left[\sum_{j=1}^{N} \Gamma_{ij}^{-1} \ E_j - \sum_{j=1}^{N} \Gamma_{ij}^{-1} \ \frac{\sum_{k,l=1}^{N} \Gamma_{kl}^{-1} \ E_l}{\sum_{k,l=1}^{N} \Gamma_{kl}^{-1}} \right] \tag{108}$$

where Γ^{-1} is the inverse of Γ (see Subsection 5.3), and $\gamma > 0$ (mean-reversion alpha) is fixed via (we set the investment level I to \$20M in our backtests)

$$\sum_{i=1}^{N} |H_i| = I \tag{109}$$

Note that (108) satisfies the dollar neutrality constraint (107).

The simulation results are given in Table 1 for (i) the heterotic risk model based on BICS sectors only (i.e., we have only one level), (ii) the heterotic risk model based on BICS industries (i.e., we have two levels, sectors and industries), (iii) the heterotic risk model based on BICS sub-industries (i.e., we have three levels, sectors, industries and sub-industries), (iv) the heterotic risk model based on BICS sub-industries plus the "prc" style risk factor of (Kakushadze, 2015b),[38] (v) the heterotic risk model based on BICS sub-industries plus the "hlv" style risk factor of (Kakushadze, 2015b), (vi) the heterotic risk model based on BICS sub-industries plus the "vol" style risk factor of (Kakushadze, 2015b),[39] (vii) the heterotic risk model based on BICS sub-

---

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.

[38] The corresponding column in the factor loadings matrix $\Omega_{i\widetilde{A}}$ is defined as $\ln(P_{i,s+1}^{AC})$ for the first (i.e., earliest) date labeled by s in the aforesaid 21-day backtesting period and kept the same for the entire such period. Other similar definitions of this factor (such as averaging over, say, the 21-day period immediately preceding the backtesting period) give very similar results.

[39] The intraday volatility based "hlv" and average volume based "vol" factors are recomputed based on 21-day periods, same as the price based "prc" factor. For the formulaic definitions of the "prc", "mom", "hlv" and "vol" factors see (Kakushadze, 2015b) and the `qrm.style()` function in Appendix A hereof, where the R code is essentially formulaic (i.e., reads like formulas).

industries plus the "mom" style risk factor of (Kakushadze, 2015b),[40] (viii) the heterotic risk model based on BICS sub-industries plus the "prc", "hlv" and "vol" style risk factors,[41] (ix) the heterotic risk model based on BICS sub-industries plus the first principal component of the sample correlation matrix Ψ_{ij}, (x) the heterotic risk model based on BICS sub-industries plus the first 10 principal components of the sample correlation matrix,[42] and (xi) the heterotic risk model based on BICS sub-industries plus the first principal component of the sample correlation matrix and the "prc" style risk factor. The results in Table 1 unequivocally suggest that adding granularity to the industry classification adds sizable value. However, any improvement from style factors and/or principal components is at best marginal.

How come? The reason is simple. By construction, when we add a new column to the factor loadings matrix, only its part orthogonal to the already present columns contributes to the risk model – indeed, the risk model is invariant under the linear transformations of the form $\Omega_{i\widetilde{A}} \to \sum_{\widetilde{B}=1}^{\widetilde{K}} \Omega_{i\widetilde{B}} \ D_{\widetilde{B}\widetilde{A}}$, where $D_{\widetilde{B}\widetilde{A}}$ is a nonsingular matrix. BICS sub-industries cover a large subspace of the (relevant)[43] risk space. It is difficult for a few style risk factors and/or principal components to provide a new relevant direction not covered by a few hundred sub-industries. And the number of relevant style risk factors is at most of order 10 for longer-horizon models and even fewer (around 4) for shorter-lookback risk models. Prosaically, it boils down to a numbers game, which the ubiquitous industry factors win hands down...[44]

6 Heterotic CAPM

Does this mean that the style risk factors are useless? Not quite. It is just that the way to use them is not to trivially append them to the much more numerous industry factors, which is how it is done traditionally. In this section we discuss a different way of utilizing the style factors, which does appear to add sizable value.

The idea – or the thought process – is quite simple. We know the industry factors add value. In fact, we know that adding granularity (i.e., going from sectors to industries to sub-industries) adds value. I.e., breaking up the universe of tickers

---

[40] For using the prior day's open-to-close momentum based risk factor (see below) to make sense, it is necessary to recompute the risk model every day, which is what – unlike all other rows based on 21-day recalculations – the "mom" row in Table 1 shows. The drop in SR and CPS is largely due to the noise from daily recalculations. Thus, the heterotic risk model based on BICS sub-industries (case (iii) above) recomputed daily has ROC 55.94%, SR 14.80 and CPS 2.66. The "mom" factor for the date labeled by s is defined as $\ln(P^C_{i,s+1}/P^O_{i,s+1})$ (Kakushadze, 2015b).

[41] We deliberately omit the "mom" factor – see footnote 40.

[42] The principal components are computed out-of-sample based on the 21-day windows, same as for the heterotic (industry classification based) risk factors.

[43] While there is no such thing as a "perfect" industry classification, well-constructed industry classifications do provide value in covering the risk space as they are based on "economic" considerations, to wit, the products, services, revenue sources, competitors, partners, suppliers, clients, etc., of the companies they classify.

[44] For long horizons, cf. (Hong, Torous and Valkanov, 2007).

19

into "clusters" (sectors, industries, sub-industries) makes a lot of sense. So, why try to apply the style risk factors to the entire universe when we can apply them at the level of each "cluster" (e.g., sub-industry)? This is analogous to what we did in the heterotic risk models. Instead of using principal components (of the sample correlation matrix) for the entire universe, we used them for each "cluster". So, the idea is this. What if we use the style factors to define the "weights" ω_i in (44) at the (most granular) level of sub-industries?

6.1   One Factor per Sub-industry

In this subsection we discuss a construction, which we term "Heterotic CAPM" for the reasons we explain below, which utilizes a single style factor, e.g., "prc", to define the "weights" ω_i in (44). We will generalize this construction to include multiple style factors below. Also, we will keep the "weights" ω'_A, ω''_a and ω'''_α in (46), (48) and (50) based on principal components. We could also utilize style factors at the levels of industries, sectors and the "market", but we will not do so here.

Table 2 gives the backtest results for ω_i defined as (i) the intercept,[45] (ii) "prc", (iii) "hlv", (iv) "vol", and (v) "mom".[46] These results appear to be all over the place. And they are for a good reason. Let us focus on the "prc" based model. Suppressing the time series index, we have $\omega_i = \ln(P_i^{AC})$. However, there is absolutely no reason why we could not define $\omega_i = \ln(P_i^{AC}/\mu)$, where μ is some constant. I.e., ω_i is only defined up to a shift proportional to the intercept $\nu_i \equiv 1$: $\omega_i = \ln(P_i^{AC}) - \ln(\mu)\,\nu_i$, and a priori there is no reason to set $\mu = 1$. The backtest results depend on this choice.[47] So, what should this μ be? There is no magic bullet here. We simply have a 1-parameter family, and this parameter can be optimized based on backtests. However, neither is it guaranteed to be out-of-sample stable (i.e., it will have to be recompute frequently), nor is it guaranteed to be independent of the alpha model which is used in the backtesting. *Así es la vida.* In fact, it is even worse: we could define $\omega_i = \ln(P_i^{AC}) - \ln(\mu_A)\,\nu_i$, $i \in J(A)$, i.e., we can have a different normalization factor μ_A for each sub-industry labeled by A. However, that would be too many parameters to deal with, so let us stick to μ uniform over all sub-industries.

In this regard, it is convenient to parameterize the "weights" ω_i as follows:

$$\omega_i = q\,\omega_i^* + (1 - 2q)\,\overline{\omega}^*\,\nu_i \tag{110}$$

where $\omega_i^* = \ln(P_i^{AC})$ and $\overline{\omega}^* = \frac{1}{N}\sum_{j=1}^{N}\omega_j^*$. So, for $q = 0$ we have the intercept (the overall normalization factor is immaterial), for $q = 1$ we have demeaned $\omega_i = \omega_i^* - \overline{\omega}^*$, and for $q = 1/2$ we have $\omega_i = \omega_i^*$. The simulation results for the values of q ranging

---

[45]   I.e., we take $\omega_i = \nu_i \equiv 1$. The intercept ν_i is just the simplest style factor.

[46]   In cases (i)-(iv) we recalculate the risk model every 21 days, and daily in case (v) (see above).

[47]   Thus, when $\mu = 1$, we get the second row in Table 2. However, when μ is large (or close to zero), we get the first row in Table 2. The normalizations (i.e., their means) of "prc", "hlv" and "vol" are all very different, hence the disparate results in Table 2. For "mom" we have an additional effect from the extra noise due to daily recomputations.

from 0 (which is equivalent to the intercept) to 0.5 (the style factor "as-is") to 1.0 (the demeaned style factor) are given for "prc" in Table 3 and "hlv" in Table 4. For "vol" the best simulated performance is for $|q| \gg 1$ (ROC 56.72%, SR 15.23, CPS 2.67) with inferior performance in between. It is evident that the demeaned style factor has inferior performance,[48] so does the intercept, and the optimal value of q corresponds to a combination of the two, except that this combination is not necessarily the same as ω_i^*, i.e., it does not necessarily correspond to $q = 0.5$.

The theoretical reason for this is clear. The factor risk model construction is not invariant under the shifts $\Omega_{iA} \to \Omega_{iA} + \chi_A$, which implies that the "centering" of the style factors matters – the overall multiplicative normalization invariance due to the invariance under the homogeneous linear transformations $\Omega_{iA} \to \sum_{A=1}^{K} \Omega_{iB} \ D_{BA}$, where D_{BA} is a nonsingular matrix, does not extend inhomogeneously.

In any event, it is evident that the "prc" style factor potentially adds most value by sizably improving CPS. However, to make sure we are not dealing with a spurious improvement, we must check whether it persists once we add position/liquidity bounds. Therefore, in Table 5 we give the simulation results with bounds for the vanilla heterotic and the "prc" based heterotic CAPM risk models. The strict bounds are

$$|H_{is}| \leq 0.01 \ A_{is} \tag{111}$$

where A_{is} is ADDV defined in (105). For the sake of simplicity we take $q = 0.5$ (i.e., the "prc" factor "as-is", $\omega_i = \omega_i^*$) for the heterotic CAPM model as this choice, albeit not optimal, does not raise out-of-sample stability questions. We use the R code in Appendix C of (Kakushadze, 2015d) for optimization with bounds. Table 5 shows that, happily, the improvement in CPS persists even with the bounds (111).

6.2 Multiple Factors per Sub-industry

In the above construction we used only one style factor per sub-industry. We can readily generalize this construction to utilize multiple style factors. Suppose we have Y style factors labeled by $\mu = 1, \ldots, Y$ and K sub-industries. Naïvely, we can take an $N \times (KY)$ factor loadings matrix of the form $\Omega_{i\widehat{A}} = \omega_{i\mu} \delta_{G(i),A}$, where $\widehat{A} = (A, \mu)$ takes KY values, and $\omega_{i\mu}$ are the factor loadings for the style factors. However, this choice will not work for small sub-industries. We can try to do overly complicated things, but in practice it does not pay off. So, let us do something simple. First, let Y_A be the number of style factors for the sub-industry A. So, $Y_A \leq Y$. Thus,

---

[48] This is because roughly half of the values of a demeaned style factor are positive, and the rest are negative, so roughly half of the correlations within each sub-industry come out to be negative, which makes little sense financially considering stocks in the same sub-industry on average should be positively correlated. The same applies to the "mom" factor, which (without demeaning) has roughly equal numbers of positive and negative values. More precisely, this is the case on days when the market is not running up or sliding down. In the heterotic CAPM models we have $\widetilde{\Gamma}_{ij} = \widehat{\xi}_i^2 \ \delta_{ij} + \omega_i \ \omega_j \ \Gamma'_{G(i),G(j)}$, so for the pair-wise ($i \neq j$) correlations within the same sub-industry ($G(i) = G(j)$) we have $\text{sign}(\widetilde{\Gamma}_{ij}) = \text{sign}(\omega_i \ \omega_j)$, hence the aforesaid roughly 50-50 split.

if $N(A) < Y$, then we can simply have a single style factor ($Y_A = 1$) for such a sub-industry and identify it with the first style factor in $\omega_{i\mu}$, i.e., we can order the columns in $\omega_{i\mu}$ such that the first column is the default column (e.g., "prc"). If $N(A) \geq Y$, we can set $Y_A = \min(N(A) - 1, Y)$ and take the first Y_A of the Y style factors. We must also make sure that the resulting style factors $\omega_{i\mu}$, $\mu = 1, \ldots, Y_A$, for each sub-industry are linearly independent (see the source code in Appendix A). The simulated performance (without bounds) for, e.g., "prc + hlv" ($Y = 2$, "prc" is the default) has the following characteristics: ROC 53.15%, SR 16.39%, CPS 2.54. The increase in the Sharpe ratio is not surprising – there are substantially more risk factors. However, it comes at a price, to wit, the reduction in ROC and CPS.

6.3 A Simple Test for Adding New Factors

As we discussed above, it is understandable that appending a few style factors to the much more numerous sub-industries makes a little difference. However, using multiple style factors per each sub-industry as in the heterotic CAPM models does not appear to add much value either. How come? There are two parts to this story.

First, the entire idea – which is essentially how most (if not all) "fundamental" commercial multifactor risk models are built – of appending style factors to industries is not all that justified. Here is why. Consider a 1-factor model for the correlation matrix: $\widetilde{\Gamma}_{ij} = \xi_i^2 \delta_{ij} + \beta_i \beta_j$, where β_i is the single column of our factor loadings matrix (and we have absorbed the 1×1 factor covariance matrix into the definition of β_i). It makes a lot of sense to take β_i to be proportional to the first principal component $V_i^{(1)}$ of the sample correlation matrix Ψ_{ij}. This is because $\Psi_{ij} = \lambda^{(1)} V_i^{(1)} V_j^{(1)} + \ldots$, where the ellipses stand for the higher principal component terms, which are subleading when N is large.[49] It further makes a lot of sense to do this for each sub-industry, i.e., model the correlation matrix $[\Psi(A)]_{ij}$ for each sub-industry labeled by A via its first principal component – and this is nothing but the heterotic risk model of (Kakushadze, 2015d) (see Subsection 3.3). However, a priori there is no reason why any given style factor β_i should be a good candidate for why the off-diagonal elements of the correlation matrix should be well-approximated by a bilinear tensor $\beta_i \beta_j$. The "lore" for justifying using style factors as β_i goes as follows. Suppose we take historical returns and regressed them over some style factors. If the correlations are sufficiently high (e.g., if we have high Fama-MacBeth (1973) t-statistic), then using such style factors in a multifactor model is justified. However, this argument has an evident caveat. A factor model – by construction – assumes that the residuals of that regression should have low correlations with the factor returns and the pair-wise correlations between different residuals should also be low. Furthermore, this should persist out-of-sample. None of this is guaranteed by simply having sufficiently high correlations between returns and style factors.

---

[49] The out-of-sample stability of $V_i^{(1)}$ is a separate issue. For large N, in the leading approximation, we have $V_i^{(1)} \approx 1/\sqrt{N}$ (the so-called "market mode"), which is evidently stable.

Second, a simple test for checking whether a style (or any other) factor may add value is to compute the following coefficient:

$$\kappa = \frac{\sum_{i,j=1}^{N} \beta_i \ \Psi_{ij} \ \beta_j}{\lambda^{(1)} \sum_{i=1}^{N} \beta_i^2} \tag{112}$$

As above, $\lambda^{(1)}$ is the largest eigenvalue of the sample correlation matrix Ψ_{ij}. By definition, $\kappa \leq 1$ (and $\kappa = 1$ only when $\beta_i = V_i^{(1)}$, up to an overall normalization factor). If κ is small, or is not small in-sample, but is unstable out-of-sample (i.e., takes small values from sample to sample), then such β_i is not going to add much value. In (112), whenever possible, β_i should be computed based on a time period prior to the time period based on which Ψ_{ij} is computed. For short lookbacks, which we are primarily interested in here, usually this is possible.

We can rewrite (112) as follows:

$$\kappa = \sum_{a=1}^{M} \frac{\lambda^{(a)}}{\lambda^{(1)}} \left[\sum_{i=1}^{N} \beta_i' \ V_i^{(a)} \right]^2 \tag{113}$$

where $\beta_i' = \beta_i / \sqrt{\sum_{j=1}^{N} \beta_j^2}$. If $N \gg 1$, we usually have $\lambda^{(1)} \gg \lambda^{(2)}$, so assuming $\left| \beta_i' \ V_i^{(1)} \right|$ is not small, we have

$$\kappa \approx \left[\sum_{i=1}^{N} \beta_i' \ V_i^{(1)} \right]^2 \tag{114}$$

and κ approximately measures the inner product (i.e., the cross-sectional "correlation") of β_i' with the first principal component.

In fact, we can compute κ for each sub-industry:

$$\kappa(A) = \frac{\sum_{i,j \in J(A)} \beta_i \ [\Psi(A)]_{ij} \ \beta_j}{\lambda(A) \sum_{i \in J(A)} \beta_i^2} \tag{115}$$

where $\lambda(A)$ is the largest eigenvalue of $[\Psi(A)]_{ij}$ and, as above, $J(A)$ is the set of the values of the index i in the sub-industry labeled by A. For sub-industries with small enough numbers of tickers $N(A)$ (compared with M), the sample correlation matrix $[\Psi(A)]_{ij}$ is reliable enough, and therefore so is $\kappa(A)$. For large sub-industries $[\Psi(A)]_{ij}$ is not as reliable and, in fact, is singular (when $M < N(A)$). However, the first principal component $[V(A)]_i$ of $[\Psi(A)]_{ij}$ is relatively reliable, and if $N(A)$ is large enough, we can apply the approximation (114) to such sub-industries

$$\kappa(A) \approx \left[\sum_{i \in J(A)} \beta_i' \ [V(A)]_i \right]^2 \tag{116}$$

23

assuming that $\left| \sum_{i \in J(A)} \beta_i' \, [V(A)]_i \right|$ is not small (see below). The bottom line is that κ and $\kappa(A)$ measure how much value β_i is expected to add in a factor model.[50]

Now, if $\left| \sum_{i \in J(A)} \beta_i' \, [V(A)]_i \right| \ll 1$, then this factor cannot possibly give a sizable improvement. We can always express β_i', $i \in J(A)$, as a linear combination of the eigenvectors of $[\Psi(A)]_{ij}$. We have no control over the eigenvectors with null eigenvalues, which are present for $M < N(A)$. We simply do not have enough data (owing to the short lookback) to deal with those directions in the risk space. As to the higher principal components with positive eigenvalues, their contribution to the performance is at best subleading and unstable out-of-sample. This is aptly illustrated by taking the vanilla heterotic model construction (Subsection 3.3) and replacing the first principal component for each sub-industry by the k_{pc}-th principal component, where $k_{pc} > 1$. More precisely, we need to deal with small sub-industries, so for a fixed k_{pc} we replace the first principal component for the sub-industry labeled by A by the k_{pc}'-th principal component, where $k_{pc}' = \min(k_{pc}, N(A))$. The result of the simulation for the values of k_{pc} from 1 to 5 is given in Table 6. These results are unequivocal: higher principal components by themselves are suboptimal. Why? For the same reason as why demeaned style factors are suboptimal: for large N we know that $V_i^{(1)} \approx 1/\sqrt{N}$, so $V_i^{(a)}$, $a > 1$, have approximately vanishing sums. The same holds for large sub-industries, and also roughly for the rest of the sub-industries (because the tickers within each sub-industry should be relatively highly correlated).[51] Consequently, as we discussed above, roughly half of the pair-wise correlations within each sub-industry are negative, which makes little sense.[52]

So, in plain English, the game of picking a style (or whatever other) factor is in trying to find a linear combination of the first principal component (with a sizable weight) and some higher principal components in such a way that it is as stable out-of-sample as possible. This is a tough game to play, especially when the number of observations (and, therefore, of the available principal components) is limited. So, style factors are simply shortcuts or rough *intelligent guesses* based on some financial/economic considerations. However, they do not necessarily work well in

---

[50] Another way of approaching this issue is, as in (Kakushadze, 2015e), to pull the off-diagonal elements of Ψ_{ij} into a vector and regress it over the similarly-pulled three tensors $\nu_i \nu_j$, $\nu_i \beta_j + \beta_i \nu_j$ and $\beta_i \beta_j$, where, as above, $\nu_i \equiv 1$ is the intercept. However, this regression makes sense only if Ψ_{ij} itself is reliably computed, i.e., if $M \gg N$, which is not the case here. We could attempt to do this for small sub-industries, but this would only give a partial picture, so here we opt to work with the coefficient κ instead. (In (Kakushadze, 2015e) this regression was applied to alphas and M was large enough compared with N.)

[51] This statement should be understood statistically. Even for a well-built industry classification, sometimes tickers are "misclassified", and even if they are properly classified, the correlations do not always aline with sub-industries, but *on average* they do, or else the vanilla heterotic risk model would not work as well as it does.

[52] On the flip side, if we replace the first principal component by the intercept, the performance also worsens (see above). This is because in this case we have uniform correlations within each sub-industry. The first principal component is close to the rescaled intercept but still captures the variability in the pair-wise correlations, hence better performance.

24

the factor model context. E.g., using long-horizon style factors such as value or growth for short-horizon risk models (for use with trading strategies with holding periods of order of a few days or intraday) makes little sense (Kakushadze and Liew, 2015), and even factors such as size (log of the market cap) must be stripped off of long-horizon elements (for size one strips off the shares outstanding and arrives at log of the price) (Kakushadze, 2015b). However, it goes beyond that. The results in Table 6 indicate that it is just a single linear combination that is expected to add value, not multiple style factors. This is consistent with our results above, that "prc" adds most value, albeit this does not imply that "prc" is the most optimal combination, in fact, above we saw that it is not. However, there is no simple or "natural" way to determine an out-of-sample stable "optimal" combination.

As they say, a picture is worth a thousand words. In Figure 1 we plot the value of the coefficient $\kappa(A)$ vs. $N(A)$, where $\kappa(A)$ is defined in (115). We emphasize that the computation of $\kappa(A)$ in Figure 1 is: (i) 100% out-of-sample (see below), and (ii) wholly independent of any trading strategy, alpha, etc. To compute it, we do the following. We take the same backtesting period as in Subsections 5.4.2 and 5.4.3, break it up into 21 day periods as above, take 2,000 top tickers by ADDV for each period, calculate the correlation matrix $[\Psi(A)]_{ij}$ for each sub-industry[53] for each such period, and calculate the "prc", "mom", "hlv" and "vol" style factors based on a 21-day period[54] immediately preceding such period (so everything is out-of-sample). Figure 1 combines data for all such 21-day periods. As expected, "mom" is most unstable, while "prc", "hlv" and "vol" have similar behavior, $\kappa(A)$ is close to 1 for large sub-industries (as these style factors have a large intercept component), and for smaller industries it is less stable. Again, the bottom line is that the intercept (or the first principal component) is the *Konzertmeister* here.[55]

7   Concluding Remarks

So – at long last – in this paper we give a complete algorithm and source code for constructing general multifactor risk models (for equities) via any combination of style factors, principal components and/or industry factors. For short lookbacks, so

---

[53] More precisely, we drop small sub-industries with fewer than 4 tickers.

[54] The "prc" factor is computed based on the day immediately preceding said 21-day period (averaging over the entire preceding 21-day period makes little difference). The "mom" factor normally would be computed as an intraday open-to-close return based on the day immediately preceding said 21-day period; however, for this test we define "mom" as a 21-day moving average of such intraday open-to-close returns to smooth out the noise (albeit this turns out to make a little difference). The "hlv" and "vol" factors are defined as above, via 21-day moving averages.

[55] For the sake of completeness, let us note that if β_i corresponds to a sub-industry factor of the form (30), assuming $V_i^{(1)} \approx 1/\sqrt{N}$, we have $\kappa \approx N(A)/N \ll 1$, where $N(A)$ is the number of tickers in the sub-industry and we are assuming that the number of sub-industries is large and $N(A) \ll N$. I.e., in this case it takes a combination of essentially all or most sub-industries to get a sizable contribution into the factor model – individual sub-industries have subleading contributions. As they say, there is strength in numbers...

that the factor covariance matrix is nonsingular, we employ the Russian-doll risk model construction (Kakushadze, 2015c). This generalizes the heterotic risk model construction (Kakushadze, 2015d) to include arbitrary non-industry risk factors as well as industry risk factors with generic "weights". When these "weights" for the sub-industry risk factors are based on a style factor (as opposed to principal components), we refer to such a construction as heterotic CAPM. The name is due to the fact that for each sub-industry we have a 1-factor model analogous to CAPM as the factor loadings are based on a style factor. In the case of the original CAPM it is (log of) the market cap; in the case of short lookbacks we have log of the price instead (the "prc" style factor (Kakushadze, 2015b)), albeit in principle it can be any other style factor such as "hlv" or "vol" or even log of the market cap.[56]

As we discussed above, generally appending a few style (or principal component) factors to the much more ubiquitous industry factors (numbering in a few hundred at the most granular level) adds little value. This is because well-constructed industry classifications such as BICS or GICS[57] already capture a lot of the relevant risk space at their most granular levels. So, a way to squeeze value from the style factors is to use them as the "weights" in the industry factors as in our heterotic CAPM models. However, even there the improvement is not earthshaking: thus, the best performing "prc" factor based heterotic CAPM does sizably improve CPS, but at the expense of lowering ROC and SR a bit. Overall, for short-horizon models the heterotic risk model construction appears to be a safe and stable choice.[58]

Another remark concerns the factor models generally. When the specific risk is computed via (17),[59] there is some degree of the inherent out-of-sample instability in the sample correlation matrix Ψ_{ij} that seeps into $\xi_i^2 = ([1-Q]\Psi[1-Q])_{ii}$ via the $Q\Psi$ and ΨQ terms. In this regard, note that the bilinear in Q term $Q\Psi Q = \Omega\Phi\Omega^T$ is different, because it is expressed via the (much more stable) factor covariance matrix Φ_{AB}. The linear in Q terms $Q\Psi$ and ΨQ cannot be rewritten via Φ_{AB}, they are related to the correlations between stock returns and factor returns, which are not as stable as factor correlations. One way to mitigate this instability is by using the definition $\xi_i^2 = 1 - (Q\Psi Q)_{ii}$ instead. However, as we discussed above, this definition is not guaranteed to yield positive ξ_i^2. One way to deal with this is to "squash" the factor risk as in the `qrm.fac()` function in Appendix A. However, this "squashing" itself introduces its own instabilities with a worse net result (see Appendix A).

So, now that we have provided a complete algorithm and source code for con-

---

[56] Log of the market cap would make sense for long horizons. For short horizons "prc" appears to add most value (compared with the "hlv", "vol" and "mom" factors).

[57] Global Industry Classification Standard.

[58] As we discussed above, there is no simple, "natural" way to pick the "centering" of a style factor such as "prc" or "hlv", i.e., where its mean should be. Furthermore, any given choice is not guaranteed to be out-of-sample stable. In this regard, using the first principal components as the "weights" for the industry risk factors as in the heterotic construction is "natural", albeit even the first principal components are not 100% out-of-sample stable. However, unlike higher principal components, they are sufficiently stable, hence the stability of the heterotic risk models.

[59] That in general we must rescale it as in Section 4 does not alter the point we make here.

structing general multifactor risk models, we hope quantitative traders will build their own risk models as opposed to wasting money on off-the-shelf alternatives, which are not optimized for (short-horizon) quantitative trading but are mostly geared toward standardized risk management in the context of the longer-horizon investments (mutual/pension funds, etc.) (Kakushadze and Liew, 2015). The usual excuse that some hard-to-get data is required to build such models is just that, an excuse. Most serious quant traders already have all the data required to build a stable short-horizon risk model, such as price-volume and industry classification data. And complicated "fundamental" data going back many years is not required as it is not relevant at short-horizons (Kakushadze and Liew, 2015).[60] So, stop wasting money and complaining, start building risk models and enjoy!

A   R Code: General Russian-Doll Risk Model

In this appendix we give the R (R Package for Statistical Computing, http://www.r-project.org) source code for building general risk models, which has options for building the heterotic risk models, heterotic CAPM models (based on a single or multiple style factors), and models with the traditional treatment of non-industry (style and/or principal component) risk factors (whereby they are appended to the industry based risk factors). See Sections 2-6 for details. The code below is essentially self-explanatory and straightforward as it simply follows the formulas therein.

The function `qrm.cov.gen(ret, load, calc.inv = T)` calculates the covariance matrix $\Gamma_{ij} = \sigma_i \sigma_j \widetilde{\Gamma}_{ij}$ ($\sigma_i = \sqrt{C_{ii}}$) for a general factor loadings matrix Ω_{iA} using the algorithm of Section 4. The input is as follows: `ret` is an $N \times d$ matrix of returns R_{is} (e.g., daily close-to-close returns), where N is the number of tickers, $d = M + 1$ 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) `load` is the $N \times K$ factor loadings matrix Ω_{iA}; if `calc.inv = T` the code also computes and returns the inverse of Γ_{ij}; otherwise it does not. The output is a list: `result$spec.risk` is the actual specific risk $\widetilde{\xi}_i = \sigma_i \xi_i / \gamma_i$, `result$fac.load` is the actual factor loadings matrix $\widetilde{\Omega}_{iA} = \sigma_i \Omega_{iA} / \gamma_i$ (see Section 4), `result$fac.cov` is the factor covariance matrix Φ_{AB}, `result$cov.mat` is the factor model covariance matrix Γ_{ij}, `result$inv.cov` is the matrix Γ_{ij}^{-1} inverse to Γ_{ij} (populated only if `calc.inv = T`), and `result$fac.ret` is the $K \times d$ matrix of factor returns f_{As}. If $M < N$, the factor loadings matrix is singular (so the inverse matrix Γ_{ij}^{-1} cannot be computed). This is dealt with by employing the Russian-doll embedding, which is implemented via the next function.

The function `qrm.gen.het(ret, ind, mkt.fac = F, rm.sing.tkr = F, p = 0, append.style = F, k.style = 0, style = 0)` reduces to the function `qrm.het(ret,`

---

[60]   Nor is the implied volatility data, as its relevance at short horizons is expected to be at best marginal, and there is no empirical evidence that it adds value for short-horizon strategies (Kakushadze, 2015d). (Cf. an older paper (Ederington and Guan, 2002), which sometimes is referred to in the context of using implied volatility for "fundamental" (long-horizon) risk models.)

`ind, mkt.fac = F, rm.sing.tkr = F)` in Appendix B of (Kakushadze, 2015d) for the default values `p = 0, append.style = F, k.style = 0, style = 0`. The input is: i) `ret`, which is the same as above; ii) `ind`, a list whose length a priori is arbitrary, and its elements are populated by the binary matrices (with rows corresponding to tickers, so `dim(ind[[·]])[1]` is N) corresponding to the levels in the input binary industry classification hierarchy in the order of decreasing granularity (for BICS `ind[[1]]` is the $N \times K$ matrix $\delta_{G(i),A}$ (sub-industries), `ind[[2]]` is the $N \times F$ matrix $\delta_{G'(i),a}$ (industries), and `ind[[3]]` is the $N \times L$ matrix $\delta_{G''(i),\alpha}$ (sectors), where the map G is defined in (27) (tickers to sub-industries), $G' = GS$ (tickers to industries), and $G'' = GSW$ (tickers to sectors), with the map S (sub-industries to industries) defined in (28), and the map W (industries to sectors) defined in (29)); iii) `mkt.fac`, where for `TRUE` at the final step we have a single industry factor ("market"), while for `FALSE` (default) the industry factors correspond to the least granular level in the industry classification (sectors for BICS); and iv) `rm.sing.tkr`, where for `TRUE` the tickers corresponding to the single-ticker "clusters" at the most granular level in the industry classification (in the BICS case this would be the sub-industry level) are dropped altogether, while for `FALSE` (default) the output universe is the same as the input universe. When the parameter `p` is greater than 0, then the first p principal components of the sample correlation matrix Ψ_{ij} are appended (via `cbind()`) to the $N \times K$ sub-industry factor covariance matrix Ω_{iA}. Similarly, when the parameter `append.style = T`, then the $N \times Y$ style factor matrix `style` is appended to Ω_{iA}. When both `p` is greater than 0 and `append.style = T`, then both the first p principal components and the style factors are appended. When `p = 0` and `append.style = F`, there is another option, to wit, to have positive integer `k.style`, in which case the code computes a heterotic CAPM model based on the first k_{style} columns in `style`. More precisely, for small sub-industries the code adjusts the number of style factors such that the resulting factor model for a given sub-industry is nonsingular (see Section 5). The output is a list: `result$spec.risk` is the actual specific risk $\overline{\xi}_i$, `result$fac.load` is the actual factor loadings matrix $\widetilde{\Omega}_{i\widetilde{A}}$, `result$fac.cov` is the factor covariance matrix $\overline{\Phi}^*_{\widetilde{A}\widetilde{B}}$ (see Subsection 5.3), `result$cov.mat` is the factor model covariance matrix Γ_{ij}, and `result$inv.cov` is the matrix Γ_{ij}^{-1} inverse to Γ_{ij}.

The function `qrm.style(prc, do.norm = T)` computes the style factors "prc", "mom", "hlv" and "vol" of (Kakushadze, 2015b). The input is: i) `prc`, a list with daily price-volume data (`prc$close` is an $N \times d$ matrix with closing prices for the d days in the time series (first column = most recent date), and similarly for open, high, low and volume); ii) if `do.norm = T` the "hlv" and "vol" factors are normalized using the auxiliary function `qrm.normalize(x, center = mean(x), sdev = sd(x))`, which conforms a vector `x` to a normal distribution with the mean and standard deviation equal `center` and `sdev`, respectively; otherwise, said factors are unnormalized. In all backtests (see the main text), all tables and Figure 1 we have set `do.norm = F`.

Finally, we give an alternative version of the function `qrm.cov.gen(ret, load, calc.inv = T)` named `qrm.cov.gen.alt(ret, load, calc.inv = T)`. The input and output are unchanged. The difference is in how the latter handles the $\widetilde{\Gamma}_{ii} \neq 1$ cases.

<div align="center">28</div>

The former uses the scale factors γ_i as in Section 4 to simultaneously rescale the specific risk and the factor risk. The latter rescales only the factor risk in such a way that the specific variances defined as sample variances less diagonal elements of the factor risk are positive. The function `qrm.cov.gen.alt()` calls another (self-explanatory) function `qrm.fr(fr, tv, low = (.1)^2, high = (.9)^2)`, which rescales the factor risk. Generally, `qrm.cov.gen.alt()` appears to underperform `qrm.cov.gen()`. E.g., the "prc" factor based heterotic CAPM model using `qrm.cov.gen.alt()` backtests as follows (without any bounds): ROC 52.23%, SR 13.85, CPS 2.84 (cf. Table 2, row 3).

```
qrm.cov.gen <- function (ret, load, calc.inv = T)
{
   print("Running qrm.cov.gen()...")

   tr <- sqrt(apply(ret, 1, var))
   r1 <- ret / tr

   reg <- lm(r1 ~ -1 + load)
   z <- t(coef(reg))
   x <- residuals(reg)
   g <- var(z, z)
   x.f <- load %*% g %*% t(load)
   x.s <- apply(x, 1, var)
   tr1 <- sqrt(x.s + diag(x.f))
   tr <- tr / tr1
   sv <- x.s * tr^2
   load <- load * tr
   cov.mat <- diag(sv) + t(x.f * tr) * tr

   if(calc.inv)
   {
      v <- load / sv
      d <- solve(g) + t(load) %*% v
      inv <- diag(1 / sv) - v %*% solve(d) %*% t(v)
   }

   result <- new.env()
   result$spec.risk <- sqrt(sv)
   result$fac.load <- load
   result$fac.cov <- g
   result$cov.mat <- cov.mat
   result$fac.ret <- t(z)
   result <- as.list(result)
   if(calc.inv)
```

29

```r
        result$inv.cov <- inv
    return(result)
}

qrm.gen.het <- function (ret, ind, mkt.fac = F, rm.sing.tkr = F, p = 0,
                            append.style = F, k.style = 0, style = 0)
{
    print("Running qrm.gen.het()...")

    null.elem <- function(z, prec = 1e-10)
    {
        bad <- sum(z == 0 | z < max(z) * prec)
        return(bad)
    }

    s <- 0
    if(length(style) > 1)
        s <- ncol(style)
    else
        append.style <- F

    if(append.style | p > 0)
        k.style <- 0

    if(mkt.fac)
        ind[[length(ind) + 1]] <- matrix(1, nrow(ind[[1]]), 1)

    if(rm.sing.tkr)
    {
        bad <- colSums(ind[[1]]) == 1
        ind[[1]] <- ind[[1]][, !bad]
        bad <- rowSums(ind[[1]]) == 0
        for(lvl in 1:length(ind))
            ind[[lvl]] <- ind[[lvl]][!bad, ]
        ret <- ret[!bad, ]

        if(length(style) > 1)
            if(ncol(style) > 1)
                style <- style[!bad, ]
            else
                style <- matrix(style[!bad, ], length(style[!bad, ]), 1)
    }
```

30

```r
flm <- ind
fac.ret <- list()
spec.risk <- list()
fac.cov <- list()
fac.ret[[1]] <- ret

if(k.style > 0)
{
   x <- y <- rep(NA, nrow(ind[[1]]))
   for(a in 1:ncol(ind[[1]]))
   {
      take <- ind[[1]][, a] > 0
      k <- sum(take)
      for(j in 1:ncol(style))
         if(sum(style[take, j] == 0) == k)
            style[take, j] <- 1

      chk.mat <- k > 1
      if(k < k.style)
         k <- 1
      if(k > 1)
         k <- min(k - 1, ncol(style))
      if(chk.mat & k > 1)
         if(null.elem(eigen(t(style[take, 1:k]) %*%
                  style[take, 1:k])$values)> 0)
            k <- 1

      while(chk.mat)
      {
         z <- residuals(lm(fac.ret[[1]][take, ] ~ -1 +
               style[take, 1:k]))
         z <- apply(z, 1, sd)
         bad <- null.elem(z)
         if(chk.mat <- bad > 0)
            if(bad < k)
               k <- k - bad
            else
            {
               k <- 1
               style[take, 1] <- 1
               chk.mat <- F
            }
      }
```

```
      x <- cbind(x, matrix(ind[[1]][, a] * style[, 1:k],
            nrow(ind[[1]]), k))
      y <- cbind(y, matrix(ind[[1]][, a], nrow(ind[[1]]), k))
   }
   flm[[1]] <- x[, -1]
   ind[[1]] <- y[, -1]
}

calc.load <- function(load, load1)
{
   x <- colSums(load1)
   load <- (t(load1) %*% load) / x
   return(load)
}

for(lvl in 1:length(ind))
{
   if(lvl > 1)
      flm[[lvl]] <- calc.load(ind[[lvl]], ind[[lvl - 1]])

   if(k.style == 0 | lvl > 1)
      for(a in 1:ncol(flm[[lvl]]))
      {
         take <- as.logical(flm[[lvl]][, a])

         if(lvl == 1)
            k <- nrow(fac.ret[[lvl]])
         else
            k <- ncol(ind[[lvl - 1]])

         x <- matrix(fac.ret[[lvl]][1:k, ][take, ], sum(take),
               ncol(fac.ret[[lvl]]))
         y <- eigen(cor(t(x)))$vectors
         y1 <- y[, 1]
         flm[[lvl]][take, a] <- y1 * flm[[lvl]][take, a]
      }

   if(p > 0 | append.style)
      if(lvl == 1)
      {
         s <- 0
         if(length(style) > 1)
            s <- ncol(style)
```

32

```
        tmp <- matrix(0, nrow(flm[[lvl]]), p + s)
        if(p > 0)
            tmp[, 1:p] <- eigen(cor(t(ret)))$vectors[, 1:p]
        if(s > 0)
            tmp[, (p+1):(p+s)] <- style
        flm[[lvl]] <- cbind(flm[[lvl]], tmp)
        p <- p + s
    }
    else
    {
        tmp <- matrix(0, nrow(flm[[lvl]]) + p, ncol(flm[[lvl]]) + p)
        tmp[1:nrow(flm[[lvl]]), 1:ncol(flm[[lvl]])] <- flm[[lvl]]
        tmp[(nrow(flm[[lvl]]) + 1):(nrow(flm[[lvl]]) + p),
            (ncol(flm[[lvl]]) + 1):(ncol(flm[[lvl]]) + p)]
                <- diag(1, p)
        flm[[lvl]] <- tmp
    }

    res <- qrm.cov.gen(fac.ret[[lvl]], flm[[lvl]], calc.inv = F)
    spec.risk[[lvl]] <- res$spec.risk
    fac.cov[[lvl]] <- res$fac.cov
    flm[[lvl]] <- res$fac.load
    fac.ret[[lvl + 1]] <- res$fac.ret
}

for(lvl in length(ind):1)
    if(lvl > 1)
        fac.cov[[lvl - 1]] <- diag(spec.risk[[lvl]]^2) +
                flm[[lvl]] %*% fac.cov[[lvl]] %*% t(flm[[lvl]])
    else
    {
        spec.risk <- spec.risk[[1]]
        fac.cov <- fac.cov[[1]]
        flm <- flm[[1]]
        mod.mat <- diag(spec.risk^2) + flm %*% fac.cov %*% t(flm)
    }

sv <- spec.risk^2

if(!rm.sing.tkr)
{
    k <- ncol(ind[[1]])
    sv1 <- colSums((t(flm[, 1:k]))^2 * diag(fac.cov[1:k, 1:k]))
```

```
      take <- colSums(ind[[1]]) == 1
      x <- diag(fac.cov)
      y <- x[1:k]
      y[take] <- 0
      x[1:k] <- y
      diag(fac.cov) <- x
      take <- rowSums(ind[[1]][, take]) == 1
      sv[take] <- sv1[take]
      spec.risk <- sqrt(sv)
   }

   v <- flm / sv
   d <- solve(fac.cov) + t(flm) %*% v
   inv <- diag(1 / sv) - v %*% solve(d) %*% t(v)

   result <- new.env()
   result$spec.risk <- spec.risk
   result$fac.load <- flm
   result$fac.cov <- fac.cov
   result$cov.mat <- mod.mat
   result$inv.cov <- inv
   result <- as.list(result)
   return(result)
}

qrm.normalize <- function(x, center = mean(x), sdev = sd(x))
   qnorm(ppoints(x)[sort.list(sort.list(x), method = "radix")],
         center, sdev)

qrm.style <- function (prc, do.norm = T)
{
   open <- prc$open
   close <- prc$close
   high <- prc$high
   low <- prc$low
   vol <- prc$vol

   ### prc factor
   st1 <- log(close[, 1])
   ### mom factor
   st2 <- log(close[, 1]/open[, 1])
   ### hlv factor
   st3 <- .5 * log(apply(((high - low)/close)^2, 1, mean))
```

```r
   if(do.norm)
      st3 <- qrm.normalize(st3)
   ### vol factor
   st4 <- log(apply(vol, 1, mean))
   if(do.norm)
      st4 <- qrm.normalize(st4)

   prc$style <- cbind(st1, st2, st3, st4)
}

qrm.fr <- function(fr, tv, low = (.1)^2, high = (.9)^2)
{
   y <- log(fr)
   y <- y - .5 * log(tv)
   y <- qrm.normalize(y, median(y), mad(y))
   min.y <- .5 * log(low)
   max.y <- .5 * log(high)
   center <- median(y)
      sdev <- mad(y)
      x <- (y - center) / sdev
   min.x <- max((min.y - center) / sdev, min(x))
   max.x <- min((max.y - center) / sdev, max(x))
   if(min.x < 0)
      x[x < 0] <- min.x * (1 - exp(x[x < 0]))
   else
      x[x < min.x] <- min.x
   if(max.x > 0)
      x[x > 0] <- max.x * (1 - exp(-x[x > 0]))
   else
      x[x > max.x] <- max.x
   y <- center + x * sdev
   y <- exp(y) * sqrt(tv)
   return(y)
}

qrm.cov.gen.alt <- function (ret, load, ind, calc.inv = T)
{
   print("Running qrm.cov.gen.alt()...")

   tr <- sqrt(tv <- apply(ret, 1, var))
   r1 <- ret / tr

   reg <- lm(r1 ~ -1 + load)
```

35

```
    z <- t(coef(reg))
    g <- var(z, z)
    load <- tr * load
    x.f <- load %*% g %*% t(load)
    fr <- sqrt(diag(x.f))
    y <- qrm.fr(fr, tv)
    take <- fr > 0
    y[!take] <- 0
    sv <- tv - y^2
    y[take] <- y[take] / fr[take]
    load <- load * y
    cov.mat <- diag(sv) + t(x.f * y) * y

    if(calc.inv)
    {
        v <- load / sv
        d <- solve(g) + t(load) %*% v
        inv <- diag(1 / sv) - v %*% solve(d) %*% t(v)
    }

    result <- new.env()
    result$spec.risk <- sqrt(sv)
    result$fac.load <- load
    result$fac.cov <- g
    result$cov.mat <- cov.mat
    result$fac.ret <- t(z)
    result <- as.list(result)
    if(calc.inv)
        result$inv.cov <- inv
    return(result)
}
```

B   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 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 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.

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.

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.

Ederington, L. and Guan, W. (2002) Is implied volatility an informationally efficient and effective predictor of future volatility? *The Journal of Risk* 4(3): 29-46.

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.

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) *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. (2015e) 101 Formulaic Alphas. *Wilmott Magazine* (forthcoming). Available online: http://ssrn.com/abstract=2701346.

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.

41

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.

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. SSRN Working Papers Series, http://ssrn.com/abstract=1969689 (December 7, 2011).

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): 254-286.

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.

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.

Table 1: Simulation results for the optimized alphas discussed in Subsection 5.4.4 using the heterotic risk model with various style and/or principal component risk factors appended to the industry factors, without any bounds on the dollar holdings. All quantities are rounded to 2 digits. The result in the third line is the same as in (Kakushadze, 2015d) with the slight difference in CPS due to rounding down employed therein. See Subsection 5.4.4 for details.

Risk Model	ROC	SR	CPS
Heterotic, BICS sectors	50.88%	13.04	2.41
Heterotic, BICS industries	54.26%	14.45	2.56
Heterotic, BICS sub-industries	55.90%	15.41	2.68
Heterotic, BICS sub-industries + prc	55.92%	15.47	2.68
Heterotic, BICS sub-industries + hlv	55.88%	15.50	2.68
Heterotic, BICS sub-industries + vol	56.02%	15.46	2.68
Heterotic, BICS sub-industries + mom	56.09%	14.71	2.65
Heterotic, BICS sub-industries + prc + hlv + vol	55.85%	15.80	2.67
Heterotic, BICS sub-industries + 1st prin.comp	55.92%	15.47	2.67
Heterotic, BICS sub-industries + 1st 10 prin.comps	55.19%	15.15	2.66
Heterotic, BICS sub-industries + 1st prin.com + prc	55.97%	15.55	2.68

Table 2: Simulation results for the optimized alphas using the heterotic CAPM risk model based on various style factors, without any bounds on the dollar holdings. All quantities are rounded to 2 digits. The result in the first line is the same as in (Kakushadze, 2015d) with the slight difference in CPS due to rounding down employed therein. See Subsection 6.1 for details.

Risk Model	ROC	SR	CPS
Heterotic, BICS sub-industries	55.90%	15.41	2.68
Heterotic CAPM, int	55.99%	15.04	2.60
Heterotic CAPM, prc	55.06%	15.24	2.99
Heterotic CAPM, hlv	55.21%	16.07	2.66
Heterotic CAPM, vol	54.82%	14.59	2.50
Heterotic CAPM, mom	45.86%	9.92	1.98

Table 3: Simulation results for the heterotic CAPM model based on the "prc" style factor for various values of the weight q defined in (110) in Subsection 6.1. ROC (return-on-capital) decreases with increasing q, SR (Sharpe ratio) peaks around $q = 0.4$, while CPS (cents-per-share) peaks around $q = 0.6$.

q	ROC	SR	CPS
0	55.99%	15.04	2.60
0.1	55.99%	15.11	2.65
0.2	55.94%	15.19	2.72
0.3	55.83%	15.25	2.80
0.4	55.58%	15.29	2.89
0.5	55.06%	15.24	2.99
0.6	54.01%	14.97	3.03
0.7	51.99%	14.21	2.96
0.8	48.84%	12.81	2.73
0.9	45.36%	11.04	2.31
1.0	42.44%	9.91	1.86

Table 4: Simulation results for the heterotic CAPM model based on the "hlv" style factor for various values of the weight q defined in (110) in Subsection 6.1. ROC (return-on-capital) decreases with increasing q, while both SR (Sharpe ratio) and CPS (cents-per-share) peak around $q = 0.7$.

q	ROC	SR	CPS
0	55.99%	15.04	2.60
0.1	55.91%	15.16	2.61
0.2	55.81%	15.32	2.62
0.3	55.68%	15.51	2.63
0.4	55.49%	15.76	2.64
0.5	55.21%	16.07	2.66
0.6	54.73%	16.46	2.68
0.7	53.76%	16.82	2.70
0.8	51.23%	16.43	2.64
0.9	46.31%	13.88	2.39
1.0	43.31%	11.02	1.99

Table 5: Simulation results for the vanilla heterotic and heterotic CAPM models with the bounds (111). The result in the first line is the same as in (Kakushadze, 2015d) with the slight difference in CPS due to rounding down employed therein.

Risk Model	ROC	SR	CPS
Heterotic, BICS sub-industries	49.00%	19.23	2.37
Heterotic CAPM, int	49.24%	19.04	2.31
Heterotic CAPM, prc	48.47%	18.89	2.64

Table 6: Simulation results for the heterotic model with the first principal component for each sub-industry replaced by the k_{pc}-th principal component (or the number of tickers in said sub-industry if this number is smaller than k_{pc}; see Subsection 6.3). The first row ($k_{pc} = 1$) corresponds to the original heterotic risk model.

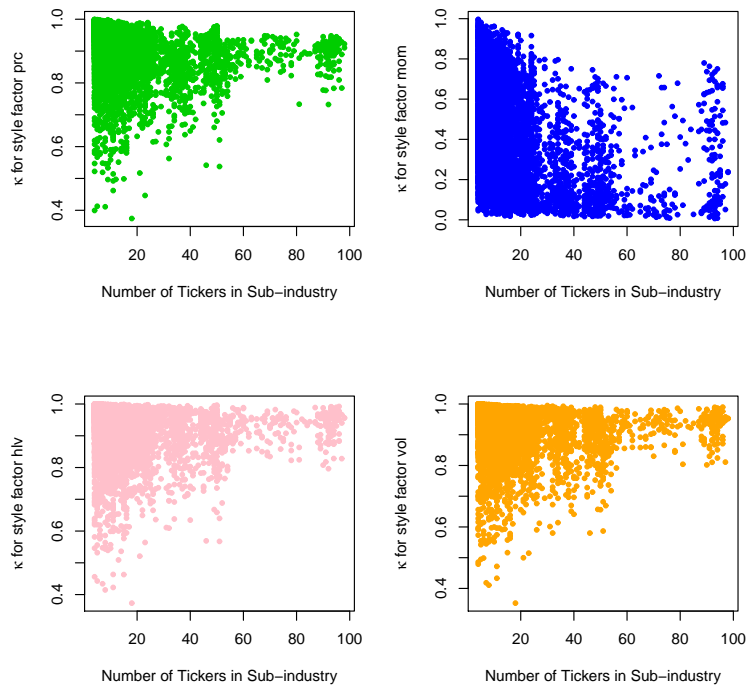| k_{pc}	ROC	SR	CPS
1	55.90%	15.41	2.68
2	37.97%	8.22	1.66
3	38.25%	8.19	1.68
4	38.21%	8.09	1.68
5	38.31%	8.11	1.68

Figure 1. The values of the coefficient $\kappa(A)$ defined in (115) vs. the number $N(A)$ of tickers in sub-industries for the style factors "prc", "mom", "hlv" and "vol". See Subsection 6.3 for details.

48