Factor Models of Asset Returns

Gregory Connor*and Robert Korajczyk†

October 26, 2007

1. Basic Definition of a Factor Model

Factor models of security returns decompose the random return on each of a cross-section of assets into factor-related and asset-specific returns. Letting r denote the vector of random returns on n assets, and assuming k factors, a factor decomposition has the form:

$$r = a + Bf + \varepsilon \tag{1}$$

where B is a nxk-matrix of factor betas, f is a random $k-$vector of factor returns, and ε is an $n-$vector of asset-specific returns. The $n-$vector of coefficients a is set so that $E[\varepsilon] = 0$. By defining B as the least squares projection $B = cov(r, f)C_f^{-1}$, it follows that $cov(f, \varepsilon) = 0^{kxn}$.

The factor decomposition (1) puts no empirical restrictions on returns beyond requiring that the means and variances of r and f exist. So in this sense it is empty of empirical content. To add empirical structure it is commonly assumed that the asset-specific returns ε are cross-sectionally uncorrelated, $E[\varepsilon\varepsilon'] = D$ where D is a diagonal matrix. This implies that the covariance matrix of returns can be written as the sum of a matrix of rank k and a diagonal matrix:

$$cov(r, r') = Bcov(f, f')B' + D. \tag{2}$$

This is called a strict factor model. Without loss of generality one can assume that $cov(f, f')$ has rank k, since otherwise one of the factors can be removed (giving a $k - 1$ factor model) without affecting the fit of the model.

*Department of Finance, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: g.connor@lse.ac.uk.

†Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2001, USA. E-mail address: r-korajczyk@kellogg.northwestern.edu.

An important (and often troublesome) feature of factor models is their rotational indeterminacy. Let L denote any nonsingular kxk-matrix and consider the set of factors $f^* = Lf$ and factor betas $B^* = L^{-1}B$. Note that f^*, B^* can be used in place of f, B since only their matrix product affects returns and the linear "rotation" L disappears from this product. This means that factors f and associated factor betas B are only defined up to a kxk linear transformation. In order to empirically identify the factor model one can set the covariance matrix of the factors equal to an identity matrix, $E[ff'] = I_k$, without loss of generality.

2. Approximate Factor Models

Security market returns have strong comovements, but the assumption that returns obey a strict factor model is easily rejected. In practice, for any reasonable value of k there will at least some discernible positive correlations between the asset–specific returns of at least some assets. An *approximate factor model* (originally developed by Chamberlain and Rothschild (1983)) weakens the strict factor model of exactly zero correlations between all asset specific returns. Instead it assumes that the there is a large number of assets n and the proportion of the correlations which are nonnegligibly different from zero is close to zero. This condition is formalized as a bound on the eigenvalues of the asset-specific return covariance matrix:

$$\lim_{n \to \infty} \max eigval[cov(\varepsilon, \varepsilon')] < c$$

for some fixed $c < \infty$. Crucially, this condition implies that asset-specific returns are *diversifiable risk* in the sense that any well-spread portfolio w will have asset-specific variance near zero:

$$\lim_{n \to \infty} w' cov(\varepsilon, \varepsilon') w = 0 \text{ for any } w \text{ such that } \lim_{n \to \infty} w'w = 0. \tag{3}$$

Note that an approximate factor model uses a "large n" modeling approach: the restrictions on the covariance matrix need only hold approximately as the number of assets n grows large.

Letting $V = cov(\varepsilon, \varepsilon')$ which is no longer diagonal, and choosing the rotation so that $cov(f, f') = I$ we can write the covariance matrix of returns as:

$$cov(r, r') = BB' + V.$$

In addition to (3) it is appropriate to impose the condition that $\lim_{n \to \infty} \min BB' = \infty$. This ensures that each of the k factors represents a pervasive source of risk in the cross-section of returns.

3. Statistical Factor Models

Financial researchers differentiate between characteristic-based, macroeconomic, and statistical factor models. In a characteristic-based model the factor betas of asset are tied to observable characteristics of the securities, such as company size or the book-to-price ratio, or the industry categories to which each security belongs. In macroeconomic factor models, the factors are linked to the innovations in observable economic time series such as inflation and unemployment. In a statistical factor model, neither factors nor betas are tied to any external data sources and the model is identified from the covariances of asset returns alone.

Recall the convenient rotation $E[ff'] = I$ which allows us to write the strict factor model (1) as:

$$cov(r, r') = BB' + D. \tag{4}$$

Assuming that the cross-section of return is multivariate normal and i.i.d. through time, the sample covariance matrix $\widehat{cov}(r, r')$ has a Wishart distribution. Imposing the strict factor model assumption (4) on the true covariance matrix it is possible to estimate the set of parameters B, D by maximum likelihood. This maximum likelihood problem requires high-dimensional nonlinear maximization: there are $nK + n$ parameters to estimate in B, D. There is also an inequality constraint on the maximization problem: the diagonal elements of D must be nonnegative, since they represent variances. The solution to the maximum likelihood problem yields estimates of B and D which correspond to the systematic and unsystematic risk measures. It is often the case that estimates of the time series of factors, f, are of interest. These are called factor scores in the statistical literature. They can be obtained through several approaches, including cross-sectional GLS regressions of r on B

$$\widehat{f_t} = (\widehat{B}'\widehat{D}^{-1}\widehat{B})^{-1}\widehat{B}'\widehat{D}^{-1}r_t.$$

See Basilevsky (1994) for a review of the various iterative algorithms which can be used to numerically solve the maximum likelihood factor analysis problem and estimate the factor scores. See Roll and Ross (1980) for an empirical application to equity returns data.

The first k eigenvectors of the return covariance matrix scaled by the square roots of their respective eigenvalues are called the k *principal components* of the covariance matrix. A restrictive version of the strict factor model is the *scalar factor model*, given by (2) plus the scalar matrix condition $D = \sigma_\varepsilon^2 I$. Under the assumption of a scalar factor model, the maximum likelihood problem simplifies, and the principal components are the maximum likelihood estimates of the factor beta matrix B (the arbitrary choice of rotation is slightly different in this case). This provides a quick and simple alternative to maximum likelihood factor analysis, under the restrictive assumption $D = \sigma_\varepsilon^2 I$.

A. Asymptotic Principal Components

The maximum likelihood method of factor model estimation relies on a strict factor model assumption and a time-series sample which is large relative to the number of assets in the cross-section. Standard principal components requires the even stronger condition of a scalar factor model. Neither method is well-configured for asset returns where the cross-section tends to be very large. Connor and Korajczyk (1986) develop an alternative method called asymptotic principal components, building on the approximate factor model theory of Chamberlain and Rothschild (1983). Connor and Korajczyk analyze the eigenvector decomposition of the TxT cross product matrix of returns rather than of the nxn covariance matrix of returns. They show that given a large cross-section, the first k eigenvectors of this cross-product matrix provide consistent estimates of the kxT matrix of factor returns. Connor and Korajczyk (1988) extend the procedure to account for cross-sectional heteroskedasticity and Jones (2001) extends the procedure to account for time-series heteroskedasticity. Stock and Watson (2002) extend the theory to allow both large time series and large cross-sectional samples, time varying factor betas, and provide a quasi-maximum likelihood interpretation of the technique. Bai (2003) analyzes the large-sample distributions of the factor returns and factor beta matrix estimates in a generalized version of this approach.

4. Macroeconomic Factor Models

The rotational indeterminacy in statistical factor models is unsatisfying for the application of factor models to many research problems. Statistical factor models do not allow the analyst to assign meaningful labels to the factors and betas; one can identify the k pervasive risks in the cross-section of returns, but not what these risks represent in terms of economic and financial theory.

One approach to making the factor decomposition more interpretable is to rotate the statistical factors so that the rotated factors are maximally correlated with pre-specified macroeconomic factors. If f_t is a k-vector of statistical factors and m_t is a k-vector of macroeconomic innovations we can regress the macroeconomic factors on the statistical factors

$$m_t = \Pi f_t + \eta_t.$$

As long as Π has rank k, the span of the rotated factors, Πf_t, is the span of the original statistical factors, f_t. However the rotated factors can now be interpreted as the return factors that are correlated with the specified macroeconomic series. With this rotation the new factors are no longer orthogonal, in general. This approach is described in Connor and Korajczyk (1991).

Alternatively, one can work with the pre-specified macroeconomic series directly. Chan, Chen, and Hsieh (1985) and Chen, Roll and Ross (1986) develop a macroeconomic factor model in which the factor innovations f are observed directly (using innovations in economic time series) and the factor betas are estimated via time-series regression of each asset's return on the time-series of factors. They begin with the standard description of the current price of each asset, p_{it}, as the present discounted value of its expected cash flows:

$$p_{it} = \sum_{s=1}^{\infty} \frac{E[c_{it}]}{(1+\rho_{st})^s}$$

where ρ_{st} is the discount rate at time t for expected cash flows at time $t+s$. Chen, Roll and Ross note that the common factors in returns must be variables which cause pervasive shocks to expected cash flows $E[c_{it}]$ and/or risk-adjusted discount rates ρ_{st}. They propose inflation, interest rate, and business-cycle related variates to capture these common factors. Shanken and Weinstein (2006) find that empirically the model lacks robustness in that small changes in the included factors or the sample period have large effects on the estimates. Connor (1995) argues that although macroeconomic factors models are theoretically attractive, since they provide a deeper explanation for return comovement than statistical factor models, their empirical fit is substantially weaker than statistical and characteristic-based models. Vassalou (2003) argues on the other hand that the ability of the Fama-French model (see below) to explain the cross-section of mean returns can be attributed to the fact that Fama-French factors provide good proxies for macroeconomic factors.

5. Characteristic-based Factor Models

A surprisingly powerful method for factor modeling of security returns is the characteristic-based factor model. Rosenberg (1974) was the first to suggest that suitably scaled versions of standard accounting ratios (book-to-price ratio, market value of equity) could serve as factor betas. Using these predefined betas, he estimates the factor realizations f_t by cross-sectional regression of time-t asset returns on the pre-defined matrix of betas.

In a series of very influential papers, Fama and French (1992, 1993, 1996) propose a two-stage method for estimating characteristic-based factor models. In the first stage they sort assets into fractile portfolios based on book-to-price and market value characteristics. They use the differences between returns on the top and bottom fractile portfolios as proxies for the factor returns. They also include a market factor proxied by the return on a capitalization-weighted market index. In the second stage, the factor betas of portfolios and/or assets are estimated by time-series regression of asset returns on the derived factors. Carhart (1997) and Jagadeesh and Titman (1993, 2001) show that the addition of a momentum factor (proxied by high-twelve-month return minus low twelve-month-return) adds explanatory power to the Fama-French three-factor model, both in terms of explaining comovements and mean returns. Ang, Hodrick, Xing and Zhang (2006a,b) and Goyal and Santa-Clara (2003) also find evidence for a own-volatility-related factor, both for explaining return comovements and mean returns.

A. Industry-Country Components Models

One of the most empirically powerful factor decompositions for equity returns is an error-components model using industry affiliations. This involves setting the factor beta matrix equal to zero/one dummies, with row i containing a one in the j^{th} column if and only if firm i belongs to industry j. This is the simplest type of characteristic-based factor model of equity returns.

The first statistical factor is dominant in equity returns, accounting for 80-90% of the explanatory power in a multi-factor model. The standard specification of a error-components model does not isolate the "first" factor since its influence is spread across the factors. Heston and Rouwenhorst (1994) describe an alternative specification in which this factor is separated from the k industry factors. They add a constant to the model, so that the expanded set of factors $k+1$ is not directly identified (this lack of identification is sometimes called the "dummy variable trap," referring to a model that includes a full set of zero-one dummies plus a constant). Then, Heston and Rouwenhorst, impose an adding-up restriction on the estimated $k+1$ factors: the set of industry factors must sum to zero. This adding-up restriction on the factors restores statistical identification to the model, requiring constrained least squared in place of standard least squares estimation. It also provides a useful interpretation of the estimated factors: the factor associated with the constant term is the "market-wide" or "first" factor, and the factors associated with the industry dummies are the extra-market industry factors. This specification has been widely adopted in the research literature.

Heston and Rouwenhorst's adding-up condition is particularly useful in a multi-country context. It allows one to include an overlapping set of country and industry dummies without encountering the problem of the dummy variable trap. Including a constant, an international industry-country factor model must impose adding-up conditions both on the estimated industry factors and on the estimated country factors. This type of country-industry specification is useful for example in measuring the relative contribution of cross-border and national influences to return comovements, see, for example, Hopkins and Miller (2001).

6. Determining the Number of Factors

In the case of maximum likelihood estimation of a strict factor models, it is possible to test for the correct number of factors by comparing the likelihood of the model with k factors to that of a $k+1$ factor model. Under the standard assumptions, for large time-series samples the ratio of the log likelihoods has an approximate chi-squared distribution. There are drawbacks to this test in the context of asset returns data and its performance has been problematic, see, e.g., Dhrymes, Friend and Gultekin (1984), Conway and Reinganum (1988), and Shanken (1987) who all find the test unreliable. The test may rely too strongly on the assumption of a strict factor model (an exactly diagonal covariance matrix of asset-specific returns) which is at best a convenient fiction in the case of asset returns. Also, it tests the hypothesis that the correct number of factors has been pre-specified, rather than optimally determining the best number of factors to use.

Connor and Korajczyk (1993) derive a test for the number of factors which is robust to having an approximate, rather than strict factor model. It is based on the decline in average idiosyncratic variance as additional factors are added.

Bai and Ng (2002) take a different approach to the factor-number decision. They view the choice of the number of factors as a model selection problem, and build on the Akaike and BIC information criteria-based tests for model selection. Bai and Ng compute the average asset-specific variance (both across time and across securities) in a model with k factors:

$$\sigma_{\varepsilon}^{2}(k) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}^{2}.$$

The Bai-Ng procedure involves choosing k to minimize a degrees-of-freedom adjusted variant of average asset-specific variance:

$$k^{*} = \arg\min \sigma_{\varepsilon}^{2} + kg(n,T). \tag{5}$$

where the penalty function $g(n,T)$ serves to compensate for the lost degrees of freedom when estimating the model with more factors.

7. Factor Beta Pricing Theory

A central concern in asset pricing theory is the determination of asset risk premium and their connection to sources of pervasive risk. Factor models have been central to this research area. In a factor beta pricing model the expected return of each asset equals the risk-free return plus a linear combination of the factor betas of the assets, with the linear weights (factor risk premia) constant across securities:

$$E[r] = 1^n r_0 + B\pi$$

where r_0 is the risk-free return and π is a $k-$vector of factor risk premia. Important special cases include the Capital Asset Pricing Model (Sharpe (1963), Treynor (1961, 1999), the Arbitrage Pricing Theory (Ross (1976)), the Fama-French model (Fama and French (1993,1996)), and Merton's (1973) Intertemporal Capital Asset Pricing Model.

References

[1] Ang, A., R. Hodrick, Y. Xing and X. Zhang, 2006a, The cross-section of volatility and expected returns, *Journal of Finance*, 61, 259-299.

[2] Ang, A., R. Hodrick, Y. Xing and X. Zhang, 2006b, High idiosyncratic risk and low returns: international and further U.S. evidence, working paper, Columbia Business School, Columbia University.

[3] Bai, J., 2003, "Inferential theory for factor models of large dimension," *Econometrica* 71, 135-171.

[4] Bai, J., and Ng, S., 2002, "Determining the number of factors in approximate factor models," *Econometrica* 70, 191-221.

[5] Carhart, M. M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57-82.

[6] Chamberlain, G. and Rothschild, M., 1983, "Arbitrage, factor structure and mean-variance analysis in large asset markets," *Econometrica* 51, 1305-1324.

[7] Chan, K. C., Nai-fu Chen, and David Hsieh, 1985, "An Exploratory Investigation of the Firm Size Effect." *Journal of Financial Economics* 14, 451-471.

[8] Chen, N., Roll, R. and Ross, S., 1986, “Economic forces and the stock market,” *Journal of Business* 59, 383-403.

[9] Connor, G., 1995, "The Three Types of Factor Models: A Comparison of their Explanatory Power," *Financial Analysts Journal*, 15, 42-46.

[10] Connor, G. and Korajczyk, R., 1986, “Performance measurement with the arbitrage pricing theory: A new framework for analysis,” *Journal of Financial Economics* 15, 373-394.

[11] Connor, G. and Korajczyk, R., 1988, “Risk and return in an equilibrium APT: Application of a new test methodology,” *Journal of Financial Economics* 21, 255-289.

[12] Connor, G. and Korajczyk, R., 1991, "The Attributes, Behavior, and Performance of U.S. Mutual Funds." *Review of Quantitative Finance and Accounting* 1, 5-26.

[13] Connor, G. and Korajczyk, R., 1993, "A Test for the Number of Factors in an Approximate Factor Model." *Journal of Finance* 48, 1263-1291.

[14] Daniel, K. and Titman, S., 1997, “Evidence on the characteristics of cross sectional variation in stock returns,” *Journal of Finance* 52, 1-33.

[15] Dhrymes, P.J., I. Friend and N.B. Gultekin, 1984, "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory, *Journal of Finance* 39, 323-246.

[16] Fama, E. and K.R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47, 427-465.

[17] Fama, E. and K.R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.

[18] Fama, E., and K.R. French, 1996, “Multifactor explanations of asset pricing anomalies,” *Journal of Finance* 51, 55-84.

[19] Goyal, A. and P. Santa-Clara, 2003, "Idiosyncratic risk matters!", *Journal of Finance* 58, 975-1007.

[20] Hopkins, P.J.B. and C.H. Miller, 2001, *Country, Sector and Company Factors in Global Equity Models*, The Research Foundation of AIMR and the Blackwell Series in Finance, Charlottesvile VA.

[21] Jagadeesh, N. and S. Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance* 48, 65-91.

[22] Jagadeesh, N. and S. Titman, 2001, "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations," *Journal of Finance* 56, 699-718.

[23] Jones, C. S., 2001, "Extracting factors from heteroskedastic asset returns," *Journal of Financial Economics* 62, 293-325.

[24] Merton, R.C., 1973. "An Intertermporal Capital Asset Pricing Model," *Econometrica* 41, 867-887.

[25] Rosenberg, B., 1974, "Extra-Market Components of Covariance in Security Returns," *Journal of Financial and Quantitative Analysis* 9, 263-274.

[26] Roll, R. and Ross, S., 1980, "An empirical investigation of the arbitrage pricing theory," *Journal of Finance* 35, 1073-1103.

[27] Ross, S., 1976, "The arbitrage theory of capital asset pricing," *Journal of Economic Theory* 13, 341-360.

[28] Shanken, J., 1987, "Nonsynchronous data and the covariance-factor structure of returns," *Journal of Finance* 42, 221-231.

[29] Sharpe, W. F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19, 425-442.

[30] Stock, J.H. and Watson, M.W., 2002, "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics* 20, 147-162.

[31] Treynor, J. L., 1961, "Toward a Theory of Market Value of Risky Assets." Unpublished Working Paper.

[32] Treynor, J. L., 1999, "Towards a Theory of Market Value of Risky Assets." In *Asset Pricing and Portfolio Performance: Models, Strategy, and Performance Metrics,* edited by R. A. Korajczyk. London: Risk Publications.