Practical Guide to Factor Covariance Matrix Construction

Grant Holtes

March 20, 2024

1 Introduction

In the area of investment portfolio construction, a factor covariance matrix is a useful alternative to the sample covariance matrix. This method operates under the assumption that the covariance between assets can be predominantly explained by a limited number of factors, while any remaining variance is deemed idiosyncratic and is excluded from the covariance[1].

The factors used are chosen based on various criteria, although the factor design and selection process falls beyond the scope of this note. Factor covariance matrices are particularly beneficial when estimating correlation matrices for high-dimensional datasets or when dealing with data histories of varying lengths and frequencies.

This note provides a simple introduction to factor covariance matrix estimation and some suggestions for further reading, as cited throughout.

2 Estimate Factor Loadings

The initial step entails computing a matrix of factor loadings, B, wherein each row corresponds to a factor, F and each column represents an asset, X. These loadings are derived from linear regressions between factor returns and asset returns, with the coefficients in these regressions constituting the factor loadings and the residual term representing the idiosyncratic component (and C providing the constant term).

$$X = C + BF + \epsilon \tag{1}$$

The regression between asset returns can be specified for all assets at once, or be done on a per-asset basis. The latter can be helpful if priors on the coefficients need to be accounted for, such as setting the coefficient on a hedging cost factor to zero for domestic asset classes, or when the length or periodicity of returns data differs by asset. Either way, what matters is that a well estimated matrix of factor loadings, B, is produced.

This relation between factors and assets, and the majority of the material for this note is based on the work of Jianqing Fan and Mincheva, 2011.

[1] This is assumed to be a desired property - if it is not, Jianqing Fan and Mincheva, 2011 provide a methodology to re-introduce correlations between assets that are not explained by the factors in a controlled manner

3 Estimate covariance Between Factors

Next, the covariance matrix between factors, $cov(F)$, must be estimated. This can be achieved simply through calculating the sample covariance matrix, $c\hat{o}v(F)$

It is worth considering if other estimators can provide a better estimate of the factor covariance, as the final asset covariance matrix will be heavily influence by this result. Other estimators to consider include the Ledoit-Wolf estimator as outlined in Olivier Ledoit, 2003, or using Random Matrix theory, as in Dmitri Mossessian, 2014.

4 Compute Pure Factor Covariance Matrix

The factor covariance matrix is then obtained through a simple matrix multiplication between the factor loadings matrix and the factor covariance matrix.

$$c\hat{o}v(X) = B^T c\hat{o}v(F) B \tag{2}$$

5 Compute Asset Variances

The diagonal of the computed factor covariance matrix provides the variance of each asset class that is attributable to the factors, which exclude the idiosyncratic component of variance. To incorporate the idiosyncratic component, the variances of all assets must be computed separately. The diagonal components of the raw factor covariance matrix should replaced with the respective computed variances.

$$diag(c\hat{o}v(X) = v\hat{a}r(X) \tag{3}$$

6 Completed Covariance Matrix

Once the diagonal substitution is done the covariance matrix is complete. The factor correlation matrix can now be computed from the factor covariance matrix.

7 Python Example

The below code gives a simple implementation of the documented steps in Python. Naturally the random values for the assets, X, and factors, F, should be replaced with your asset and factor return series.

```
import numpy as np

X = np.random.random((100,10)) # Asset return array, cols are assets, rows are observations
F = np.random.random((100,5)) # Factor return array, cols are factors, rows are observations, plus a

B = np.linalg.lstsq(F, X)[0] # Regress Assets on Factors, extract coeff
C_f = np.cov(F.T) # Get rowwise factor covarience
V = np.var(X, axis=0) # Get asset variance
C_x = B.T @ C_f @ B # Get raw factor covariance
np.fill_diagonal(C_x, V) # replace diagonal with var
```

Note:

Views expressed are the author's, and may differ from those of JANA investments. This material does not constitute investment advice and should not be relied upon as such. Investors should seek independent advice before making investment decisions. Past performance cannot guarantee future results. The charts and tables are shown for illustrative purposes only.

References

Dmitri Mossessian, V. V. (2014). Robust estimation of risk factor model covariance matrix. *FactSet Research Systems Inc.*

Jianqing Fan, Y. L., & Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics.*

Olivier Ledoit, M. W. (2003). Honey, i shrunk the sample covariance matrix.