GRStest august16

Asset Pricing (Erasmus Universiteit Rotterdam)

The GRS-test: Derivation and Interpretation

Erik Kole

August, 2016

1 Introduction

The GRS-test (after Gibbons et al., 1989) tests whether the intercepts in a set of linear time-series regressions are jointly equal to zero. This test can be used for all asset pricing models that have a linear factor structure, and whose factors are excess returns on traded financial assets. This document derives the test from regression theory.

All unconditional asset pricing models with a linear factor structure can be written as

$$r_{i,t} = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{f}_t + \varepsilon_{i,t}, \quad \mathrm{E}[\varepsilon_{i,t}] = 0, \quad i = 1, \ldots, n; \quad t = 1, \ldots, T \tag{1.1}$$

where $r_{i,t}$ is the excess return on asset i, α_i is the intercept, $\boldsymbol{f}_t$ the vector of factor realizations, $\boldsymbol{\beta}_i$ the vector of coefficients for the factors and $\varepsilon_{i,t}$ an error term that has expectation zero. We assume that the researcher considers n assets, and conducts the test based on T observations. The number of factors equals k.

Because the factor realizations are excess returns themselves, we can use a time-series regression to estimate α_i and $\boldsymbol{\beta}_i$ for all assets. When we want to derive the properties of the estimates for α_i and $\boldsymbol{\beta}_i$, we need some extra assumptions. The GRS-test is based on classical regression theory for finite samples, so under the assumption that the errors are independently and identically normally distributed (also referred to as i.i.d. normal)

$$\varepsilon_{i,t} \sim \mathrm{N}(0, \sigma_i^2), \tag{1.2}$$

$$\mathrm{E}[\varepsilon_{i,t}\varepsilon_{j,t}] = \sigma_{ij} \quad \text{for all } i \neq j \text{ and } t, \tag{1.3}$$

$$\mathrm{E}[\varepsilon_{i,t}\varepsilon_{j,t+s}] = 0 \quad \text{for all } i, j, t, \text{ and } s \neq 0, \tag{1.4}$$

I first give a recap of some important results from classical regression theory, and then derive the GRS-test.

2 The univariate linear model

The general form of a linear model is

$$y_t = \boldsymbol{b}'\boldsymbol{x}_t + e_t, \quad \mathrm{E}[e_t] = 0, \quad \text{for } t = 1, \ldots, T \tag{2.1}$$

where y_t is the dependent variable, and $\boldsymbol{x}_t$ is the $m \times 1$ vector of explanatory variables. This model can contain an intercept, in which case the first element of $\boldsymbol{x}_t$ is always equal to 1. The vector $\boldsymbol{b}$ contains the regression coefficients, and e_t is the error term. We gather all observations y_t in the $T \times 1$ vector $\boldsymbol{y}$ and all realizations $\boldsymbol{x}_t$ in the $T \times m$ matrix $\boldsymbol{X}$, and write our model as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}, \quad \mathrm{E}[\boldsymbol{e}] = \boldsymbol{0}_T, \tag{2.2}$$

where $\boldsymbol{0}_T$ denotes the $T \times 1$ vector with zeros.

2.1 The OLS estimator

The estimator $\hat{\boldsymbol{b}}$ result from applying the ordinary least squares (OLS) regression. This means that we minimize the sum of squared error terms

$$\min_{\boldsymbol{b}} \sum_{t=1}^{T} e_t^2 = \min_{\boldsymbol{b}} \boldsymbol{e}'\boldsymbol{e} \tag{2.3}$$

where $\boldsymbol{e}$ is the $T \times 1$ vector of residuals. By substituting $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$, we can write the objective as

$$\min_{\boldsymbol{b}} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}). \tag{2.4}$$

Differentiating the objective with regard to $\boldsymbol{b}$ gives the first order conditions for the optimization problem,

$$-2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{0}. \tag{2.5}$$

Solving them leads to the estimator

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{2.6}$$

For a statistical analysis of the estimator $\hat{\boldsymbol{b}}$, we need further assumptions on the model in eq. (2.1). In the classical regression model, we assume that the error terms are i.i.d. draws from a normal distribution, $e_t \sim \mathrm{N}(0, \sigma^2)$, $\mathrm{E}[e_t e_{t+s}] = 0$ for all $s \neq 0$. In the matrix notation of eq. (2.2), it means that we assume

$$\boldsymbol{e} \sim \mathrm{N}_T\big(\boldsymbol{0}_T, \sigma^2 \boldsymbol{I}_T\big), \tag{2.7}$$

where N_T denotes the multivariate normal distribution with dimension T, and $\boldsymbol{I}_T$ denotes the $T \times T$ identity matrix.

In this setup, we can derive the distribution of $\hat{\boldsymbol{b}}$. Under the assumption that the model specification is correct, we can substitute eq. (2.2),

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}) = \boldsymbol{b} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e} \tag{2.8}$$

This result shows that $\hat{\boldsymbol{b}}$ is a linear transformation of the normally distributed variables $\boldsymbol{e}$. Consequently, $\hat{\boldsymbol{b}}$ has a normal distribution as well. Its expectation follows as

$$\mathrm{E}\big[\hat{\boldsymbol{b}}\big] = \mathrm{E}\big[\boldsymbol{b} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}\big] = \boldsymbol{b}, \tag{2.9}$$

so the estimator is unbiased. Its variance follows as

$$\begin{aligned}
\mathrm{var}\big[\hat{\boldsymbol{b}}\big] &= \mathrm{E}\big[(\hat{\boldsymbol{b}} - \boldsymbol{b})(\hat{\boldsymbol{b}} - \boldsymbol{b})'\big] \\
&= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}\boldsymbol{e}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}[\boldsymbol{e}\boldsymbol{e}']\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{I}_T\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.
\end{aligned} \tag{2.10}$$

We have now established that

$$\hat{\boldsymbol{b}} \sim \mathrm{N}_m\left(\boldsymbol{b}, \sigma^2\boldsymbol{Q}\right), \tag{2.11}$$

with $\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$.

We often also need an estimator for the true residual variance σ^2. It can be estimated by

$$\hat{\sigma}^2 = \frac{1}{T-m}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}, \tag{2.12}$$

where $\hat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}$. The estimator $\hat{\sigma}^2$ has the property

$$(T-m)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{T-m}, \tag{2.13}$$

where χ^2_{T-m} denotes a χ^2-distribution with $T-m$ degrees of freedom. The expectation of a χ^2_ν distribution is equal to ν, so $\mathrm{E}[(T-m)\hat{\sigma}^2/\sigma^2] = T-m$, and hence $\mathrm{E}[\hat{\sigma}^2] = \sigma^2$. This result shows that $\hat{\sigma}^2$ is an unbiased estimator of the variance σ^2. Though the estimator

$$\tilde{\sigma}^2 = \frac{1}{T}\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \frac{T-m}{T}\hat{\sigma}^2, \tag{2.14}$$

is a biased estimator of σ^2, we will see later that this estimator can be useful.

2.2 Tests of a single hypothesis

Suppose that we want to test whether the true value of a specific coefficient b_j is equal to a. Under the null hypothesis that $b_j = a$, we have that

$$\hat{b}_j \sim \mathrm{N}\left(a, q_{jj}\sigma^2\right), \tag{2.15}$$

where q_{jj} gives the j-th diagonal element of the matrix $\boldsymbol{Q}$. Consequently,

$$\frac{\hat{b}_j - a}{\sqrt{q_{jj}\sigma^2}} \sim \mathrm{N}(0,1). \tag{2.16}$$

However, we cannot calculate the value of this statistic, because the true variance σ^2 is unknown. Instead, we can use the estimator $\hat{\sigma}^2$, which produces the t-statistic

$$t = \frac{\hat{b}_j - a}{\sqrt{q_{jj}\hat{\sigma}^2}}. \tag{2.17}$$

This statistic does not follow a normal distribution anymore, because it is a ratio of two estimators $\hat{b}_j$ and $\hat{\sigma}^2$. Using the properties of $\hat{\sigma}^2$, we can write

$$t = \frac{(\hat{b}_j - a)/\sqrt{q_{jj}\sigma^2}}{\sqrt{\frac{(T-m)\hat{\sigma}^2}{\sigma^2}/(T-m)}} \sim \mathrm{t}_{T-m}. \tag{2.18}$$

The numerator in this expression equals eq. (2.16) and follows a standard normal distribution. The denominator is the square root of a χ^2 distributed random variable divided by its degrees of freedom. Because of this structure, the t-statistic follows a Student's t distribution with $T-m$ degrees of freedom. Compared to the standard normal distribution, it accounts for the additional estimation uncertainty in $\hat{\sigma}^2$.

2.3 Tests of joint hypotheses

It is also possible to test a set of l linear coefficient restrictions. Suppose that $\boldsymbol{H}$ is an $l \times m$ matrix with rank l. For example, when we want to test whether all coefficients are equal to zero, $\boldsymbol{H}$ equals the $m \times m$ identity matrix $\boldsymbol{I}_m$; when we want to test whether the first two coefficients are equal to zero, $H_{11} = H_{22} = 1$ and all other elements are zero. We write our hypothesis as $\boldsymbol{Hb} = \boldsymbol{a}$ with $\boldsymbol{a}$ a vector of size l. Under the null hypothesis, we have

$$\boldsymbol{H}\hat{\boldsymbol{b}} \sim \mathrm{N}_l\left(\boldsymbol{a}, \boldsymbol{HQH}'\sigma^2\right). \tag{2.19}$$

We can show that the quadratic form

$$(\boldsymbol{H}\hat{\boldsymbol{b}} - \boldsymbol{a})' \left(\boldsymbol{HQH}'\sigma^2\right)^{-1} (\boldsymbol{H}\hat{\boldsymbol{b}} - \boldsymbol{a}) \sim \chi^2_l. \tag{2.20}$$

Again, we cannot calculate the value of this statistic, because the true variance σ^2 is unknown. We can again substitute the estimator $\hat{\sigma}^2$. After proper scaling, we construct the F-statistic

$$\begin{aligned} F &= \frac{1}{l}\left(\boldsymbol{H}\hat{\boldsymbol{b}} - \boldsymbol{a}\right)' \left(\boldsymbol{HQH}'\hat{\sigma}^2\right)^{-1} \left(\boldsymbol{H}\hat{\boldsymbol{b}} - \boldsymbol{a}\right) & (2.21) \\ &= \frac{\left(\boldsymbol{H}\hat{\boldsymbol{b}} - \boldsymbol{a}\right)' (\boldsymbol{HQH}'\hat{\sigma}^2)^{-1} \left(\boldsymbol{H}\hat{\boldsymbol{b}} - \boldsymbol{a}\right)/l}{\dfrac{(T-m)\hat{\sigma}^2}{\sigma^2}/(T-m)} \sim \mathrm{F}_{l,T-m}. & (2.22) \end{aligned}$$

Both the numerator and denominator in this expression are χ^2 distributed random variables divided by their degrees of freedom. Because of this structure, the F-statistic follows an F-distribution with l and $T-m$ degrees of freedom.

3 The multivariate linear model

The standard model in eq. (2.1) can be extended to a setting where a set of n dependent variables $\boldsymbol{y}_t$ are considered,

$$\boldsymbol{y}_t = \boldsymbol{B}'\boldsymbol{x}_t + \boldsymbol{e}_t, \quad \mathrm{E}[\boldsymbol{e}_t] = 0, \tag{3.1}$$

where $\boldsymbol{B} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$ is the $m \times n$ matrix of regression coefficients, and $\boldsymbol{e}_t$ is the $n \times 1$-vector of error terms. In the classical setting, these error terms are i.i.d. draws from a multivariate normal distribution,

$$\boldsymbol{e}_t \sim \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{3.2}$$

where $\boldsymbol{\Sigma}$ is an $n \times n$ variance-covariance matrix.

3.1 The SUR estimator

The model in eq. (3.1) is called a Seemingly Unrelated Regression (SUR) model, because the different variables in $\boldsymbol{y}_t$ are linked only via the variance matrix $\boldsymbol{\Sigma}$.[1] Because the explanatory variables are the same for each dependent variable, the estimators $\hat{\boldsymbol{b}}_i$ can be estimated by a

[1]See Heij et al. (2004, Ch. 7.7.2) or Greene (2003, Ch. 14.2) for a general discussion of the SUR model.

set of regressions that are separate for each dependent variable, as in eq. (2.6). The matrix of estimators $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{b}}_1, \cdots, \hat{\boldsymbol{b}}_n)$ satisfies

$$\hat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \tag{3.3}$$

where $\boldsymbol{Y}$ is the $T \times n$ matrix with all observations of the dependent variables.

To derive the properties of the estimators, we write the model again in matrix form such that all observations are included,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \tag{3.4}$$

where $\boldsymbol{E}$ is the $T \times n$ matrix containing the error terms. The expectation follows directly from substitution of eq. (3.4) in eq. (3.3),

$$\mathrm{E}[\hat{\boldsymbol{B}}] = \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\right] = \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{B} + \boldsymbol{E})\right] = \boldsymbol{B}. \tag{3.5}$$

Similar to eq. (2.9), this result shows that $\hat{\boldsymbol{B}}$ unbiased.

We cannot find an expression for the variance of $\hat{\boldsymbol{B}}$, because the variance of a matrix is not defined. Instead, we can turn $\hat{\boldsymbol{B}}$ into a vector by stacking its columns. I show the result in appendix A.1, which uses some more advanced linear algebra. Two results are particularly useful. The first pertains to the distribution of column i of $\hat{\boldsymbol{B}}$. We denote this column by $\hat{\boldsymbol{B}}_{\bullet i}$. It contains all coefficient for dependent variable y_i. Its distribution is given by

$$\hat{\boldsymbol{B}}_{\bullet i} = \hat{\boldsymbol{b}}_i \sim \mathrm{N}_m\big(\boldsymbol{b}_i, \sigma_{ii}^2 \boldsymbol{Q}\big), \tag{3.6}$$

where σ_{ii}^2 denotes the i^{th} diagonal element of $\boldsymbol{\Sigma}$. This is essentially the same as we derived in eq. (2.11).

The second result pertains to the distribution of row j of $\hat{\boldsymbol{B}}$, denoted by $\hat{\boldsymbol{B}}_{j\bullet}$. It contains all coefficients for explanatory variable x_j. Its distribution is given by

$$\hat{\boldsymbol{B}}'_{j\bullet} \sim \mathrm{N}_n\big(\boldsymbol{B}'_{j\bullet}, q_{jj}\boldsymbol{\Sigma}\big). \tag{3.7}$$

I derive this result in appendix A.2.

For testing, the variance matrix of the error terms is needed. It can can be estimated as

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-m}\hat{\boldsymbol{E}}'\hat{\boldsymbol{E}}, \tag{3.8}$$

where $\hat{\boldsymbol{E}} = \boldsymbol{Y} - \hat{\boldsymbol{B}}\boldsymbol{X}$ is the $T \times n$ matrix with the residuals. This estimator has the property

$$(T-m)\hat{\boldsymbol{\Sigma}} \sim \mathrm{W}_n(\boldsymbol{\Sigma}, T-m), \tag{3.9}$$

where $\mathrm{W}_n(\boldsymbol{\Sigma}, T-m)$ denotes a Wishart distribution with scale matrix $\boldsymbol{\Sigma}$ and $T-m$ degrees of freedom. The Wishart distribution is the multivariate generalization of the χ^2-distribution. The expectation of a Wishart distribution $\mathrm{W}_n(\boldsymbol{V}, \nu)$ equals $\nu\boldsymbol{V}$, so $\mathrm{E}[(T-m)\hat{\boldsymbol{\Sigma}}] = (T-m)\boldsymbol{\Sigma}$, and $\mathrm{E}[\hat{\boldsymbol{\Sigma}}] = \boldsymbol{\Sigma}$ which establishes that $\hat{\boldsymbol{\Sigma}}$ is an unbiased estimator. As in the univariate case, the estimator

$$\widetilde{\boldsymbol{\Sigma}} = \frac{1}{T}\hat{\boldsymbol{E}}'\hat{\boldsymbol{E}} = \frac{T-m}{T}\hat{\boldsymbol{\Sigma}} \tag{3.10}$$

is biased, but will turn out useful later.

3.2 Tests in a SUR model

When we want to test on a single coefficient b_{ji}, or coefficients that belong to a specific asset i, we can follow the procedures of the previous subsection. The multivariate setting does not influence the result. A special case arises, when we want to test on a set of coefficients that corresponds with the same explanatory variable j, so on $\boldsymbol{B}_{j\bullet}$. Its distribution is given in eq. (3.7). We can base a test of the hypothesis $\boldsymbol{B}'_{j\bullet} = \boldsymbol{a}$ on the quadratic form

$$(\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{a}')\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{B}}'_{j\bullet} - \boldsymbol{a})/q_{jj} \sim \chi^2_n, \tag{3.11}$$

but this form still depends on the unknown variance matrix $\boldsymbol{\Sigma}$. Replacing it by the estimate $\hat{\boldsymbol{\Sigma}}$ gives the statistic

$$t^2 = (\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{a}')\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{B}}'_{j\bullet} - \boldsymbol{a})/q_{jj} \sim \mathrm{T}^2_{n,T-m}, \tag{3.12}$$

where $\mathrm{T}^2_{n,T-m}$ denotes Hotelling's T^2-distribution with parameters n and $T-m$.[2] I show in appendix A.2 that the t^2-statistic is a quadratic form of an n-variate variable with a standard normal distribution and the inverse of an independent n-variate standard Wishart distribution with $T-m$ degrees of freedom. The T^2 distribution is proportional to the F-distribution. The F-statistic follows by scaling t^2 to

$$F = \frac{T-m-n+1}{n(T-m)}t^2 \sim \mathrm{F}_{n,T-m-n+1}, \tag{3.13}$$

which follows an F-distribution with n and $T-m-n+1$ degrees of freedom.

4 Tests in a one-factor model

We first look at testing when a single factor is considered. The factor can be the excess return on the market portfolio, but the results apply to any return factor. Therefore, we write the

[2] Hotelling (1931) derives it as a generalization of the t-statistic.

model in eq. (1.1) as

$$r_{i,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t}, \tag{4.1}$$

for n assets. To apply regression theory, we gather the time series of excess returns in a T-vector $\boldsymbol{r}_i$, and the time-series of factor returns in a T-vector $\boldsymbol{f}$. We construct the auxiliary matrix $\boldsymbol{X} = (\boldsymbol{\iota}_T \ \boldsymbol{f})$, where $\boldsymbol{\iota}_T$ denotes a T-vector of ones.

4.1 Estimator and test statistics

The OLS estimator for α_i and β_i follow as

$$\begin{pmatrix} \hat{\alpha}_i \\ \hat{\beta}_i \end{pmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{r}_i. \tag{4.2}$$

When we test the factor model), our interest is on $\hat{\alpha}_i$. If the factor model is correct, it means that $\alpha_i = 0$. Based on eq. (2.11), we find for a single $\hat{\alpha}_i$,

$$\hat{\alpha}_i \sim \mathrm{N}\left(\alpha_i, q_{11}\sigma_i^2\right). \tag{4.3}$$

Because all coefficients $\hat{\alpha}_i$ correspond with the intercept, their joint distribution results from eq. (3.7),

$$\hat{\boldsymbol{\alpha}} \sim \mathrm{N}\left(\boldsymbol{\alpha}, q_{11}\boldsymbol{\Sigma}\right). \tag{4.4}$$

These two results are sufficient to derive tests on the intercepts. To test the hypothesis $\alpha_i = 0$, we construct the t-statistic as in eq. (2.17)

$$t_i = \frac{\hat{\alpha}_i}{\sqrt{q_{11}\hat{\sigma}_i^2}} \sim \mathrm{t}_{T-2}. \tag{4.5}$$

So, for an individual asset we conduct a t-test with $T-2$ degrees of freedom to determine whether the pricing error α_i is significant. Mostly, we conduct a two-sided test and calculate a p-value as $1 - \Pr[-|t_i| \leq t \leq |t_i|]$. If the probability is smaller than typical values of 0.05 or 0.01, we reject $\alpha_i = 0$.

The GRS-test with one factor

To test the joint hypothesis $\boldsymbol{\alpha} = 0$, we use the t^2-statistic of eq. (3.12) combined with the scaling in eq. (3.13) (and using $m = 2$),

$$F = \frac{T-n-1}{n(T-2)}\frac{1}{q_{11}}\boldsymbol{\alpha}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\alpha} \sim \mathrm{F}_{n,T-n-1}. \tag{4.6}$$

This statistic is the GRS-statistic. It follows F-distribution with n and $T-n-1$ degrees of freedom. Because $z>0$, a test with z is always one-sided. We reject $\boldsymbol{\alpha}=\mathbf{0}$ if $\Pr[Z \geq z]$ falls below typical values of 0.05 or 0.01. In that case, the asset pricing model does not pass the test.

4.2 Economic interpretation

To obtain better insight in what an asset pricing test means in economic terms, we take a closer look at the estimators and the test statistics. Therefore, we derive the elements of the 2×2 matrix $\boldsymbol{X}'\boldsymbol{X}$, its inverse $\boldsymbol{Q}$, and the vector that results from $\boldsymbol{X}'\boldsymbol{r}_i$. For $\boldsymbol{X}'\boldsymbol{X}$ we find

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} \boldsymbol{\imath}_T'\boldsymbol{\imath}_T & \boldsymbol{\imath}_T'\boldsymbol{f} \\ \boldsymbol{f}'\boldsymbol{\imath}_T & \boldsymbol{f}'\boldsymbol{f} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T 1 & \sum_{t=1}^T f_t \\ \sum_{t=1}^T f_t & \sum_{t=1}^T f_t f_t \end{pmatrix} = T \begin{pmatrix} 1 & \overline{f} \\ \overline{f} & \overline{f^2} \end{pmatrix}, \tag{4.7}$$

with sample average $\overline{f} = \frac{1}{T}\sum_{t=1}^T f_t$ and sample second moment $\overline{f^2} = \frac{1}{T}\sum_{t=1}^T f_t^2$. Inverting this matrix[3] yields

$$\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{T\left(\overline{f^2} - \overline{f}^2\right)} \begin{pmatrix} \overline{f^2} & -\overline{f} \\ -\overline{f} & 1 \end{pmatrix} = \frac{1}{T\widetilde{\text{var}}[f]} \begin{pmatrix} \overline{f^2} & -\overline{f} \\ -\overline{f} & 1 \end{pmatrix}. \tag{4.8}$$

In this expression $\widetilde{\text{var}}[f]$ denotes the biased variance estimator,

$$\widetilde{\text{var}}[f] = \frac{1}{T}\sum_{t=1}^T (f_t - \bar{f})^2 = \overline{f^2} - \overline{f}^2, \tag{4.9}$$

which uses the divisor T instead of $T-1$.[4] Next, we consider the 2×1 vector $\boldsymbol{X}'\boldsymbol{r}_i$

$$\boldsymbol{X}'\boldsymbol{r}_i = \begin{pmatrix} \boldsymbol{\imath}'\boldsymbol{r}_i \\ \boldsymbol{f}'\boldsymbol{r}_i \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T r_{i,t} \\ \sum_{t=1}^T f_t r_{i,t} \end{pmatrix} = T \begin{pmatrix} \overline{r_i} \\ \overline{f r_i} \end{pmatrix}, \tag{4.10}$$

with sample second co-moment $\overline{f r_i} = \frac{1}{T}\sum_{t=1}^T f_t r_{i,t}$.

Combining these expressions yields

$$\begin{aligned} \begin{pmatrix} \hat{\alpha}_i \\ \hat{\beta}_i \end{pmatrix} &= \boldsymbol{Q}\boldsymbol{X}'\boldsymbol{r}_i = \frac{1}{T\widetilde{\text{var}}[f]} \begin{pmatrix} \overline{f^2} & -\overline{f} \\ -\overline{f} & 1 \end{pmatrix} T \begin{pmatrix} \overline{r_i} \\ \overline{f r_i} \end{pmatrix} \\ &= \frac{1}{\widetilde{\text{var}}[f]} \begin{pmatrix} \overline{f^2} \cdot \overline{r_i} - \overline{f} \cdot \overline{f r_i} \\ -\overline{f} \cdot \overline{r_i} + \overline{f r_i} \end{pmatrix}. \end{aligned} \tag{4.11}$$

---

[3]The inverse of a symmetric definite matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is given by $\frac{1}{ac-b^2}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}$.

[4]The second step corresponds with the relation $\text{var}[X] = \text{E}\left[(X - \text{E}[X])^2\right] = \text{E}\left[X^2\right] - \text{E}\left[X\right]^2$.

The second element in the matrix-part is the (biased) covariance estimator[5]

$$\widetilde{\text{cov}}[f, r_i] = \frac{1}{T}\sum_{t=1}^{T}(f - \bar{f})(r_i - \bar{r}_i) = \overline{fr_i} - \bar{f}\bar{r}_i. \tag{4.12}$$

This means that

$$\hat{\beta}_i = \widetilde{\text{cov}}[f, r_i]/\widetilde{\text{var}}[f], \tag{4.13}$$

the ratio of factor-return covariance to factor variance. To find the expression for $\hat{\alpha}_i$, we apply

$$\begin{aligned}\hat{\alpha}_i =& \frac{\overline{f^2} \cdot \overline{r_i} - \overline{f} \cdot \overline{fr_i}}{\widetilde{\text{var}}[f]} = \frac{\overline{f^2} \cdot \overline{r_i} - \overline{f}^2\overline{r_i} + \overline{f}^2\overline{r_i} - \overline{f} \cdot \overline{fr_i}}{\widetilde{\text{var}}[f]} \\ =& \frac{\widetilde{\text{var}}[f]\overline{r_i} - \overline{f}\widetilde{\text{cov}}[f, r_i]}{\widetilde{\text{var}}[f]} = \overline{r_i} - \hat{\beta}_i\overline{f}\end{aligned} \tag{4.14}$$

The estimate $\hat{\alpha}_i$ is the difference between the average return, $\overline{r_i}$, and the predicted average return, $\hat{\beta}_i\overline{f}$. If the model is correct, this difference is not significantly different from zero, and therefore we test whether all coefficients $\alpha_i = 0$.

Finally, we consider the test statistics in more detail. For the top-left element of q_{11} we have

$$q_{11} = \frac{\overline{f^2}}{T\left(\overline{f^2} - \overline{f}^2\right)} = \frac{\overline{f^2} - \overline{f}^2 + \overline{f}^2}{T\left(\overline{f^2} - \overline{f}^2\right)} = \frac{1}{T}\left(1 + \frac{\bar{f}^2}{\widetilde{\text{var}}[f]}\right) = \frac{1}{T}(1 + \text{Sh}[f]^2) \tag{4.15}$$

The last fraction in this expression is the squared average factor return divided by the variance of the factor, so it is the squared Sharpe ratio of the factor, $\text{Sh}[f]$. The variance in the denominator uses the biased variance estimator. Substitution of eq. (4.15) in eq. (4.5) yields

$$t_i = \sqrt{\frac{T}{1 + \text{Sh}[f]^2}}\frac{\hat{\alpha}_i}{\hat{\sigma}_i} \tag{4.16}$$

Substitution in eq. (4.6) leads to

$$F = \frac{T(T - n - 1)}{n(T - 2)}\frac{1}{1 + \text{Sh}[f]^2}\boldsymbol{\alpha}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\alpha} = \frac{T - n - 1}{n}\frac{1}{1 + \text{Sh}[f]^2}\boldsymbol{\alpha}'\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\alpha} \tag{4.17}$$

The last expression uses the biased estimator for the residual variance of eq. (3.10).

[5] The second step corresponds with the relation $\text{cov}[X, Y] = \text{E}\left[(X - \text{E}[X])(Y - \text{E}[Y])\right] = \text{E}[XY] - \text{E}[X]\,\text{E}[Y]$.

Gibbons et al. (1989) show that the term $\boldsymbol{\alpha}'\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\alpha}$ is related to Sharpe ratios. Let Sh[$\boldsymbol{r}$] denote the Sharpe ratio of the efficient portfolio constructed from a set of assets with excess returns $\boldsymbol{r}$ when a riskless asset is present. Then

$$\boldsymbol{\alpha}'\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\alpha} = \mathrm{Sh}[(\boldsymbol{r}', f')']^2 - \mathrm{Sh}[f]^2, \tag{4.18}$$

(see appendix B for a derivation). It gives the difference between the square of the maximum Sharpe ratio that can be obtained by a portfolio of the tests assets AND the market, and the squared Sharpe ratio of the market portfolio alone. So eq. (4.17) can be expressed completely in terms of Sharpe ratios, so as tests of efficient portfolios.

5 Tests in a multiple factor model

The model in eq. (2.1) and the approach so far can be easily extended to a setting with multiple factors. We replace the single factor f_t by a vector of factors $\boldsymbol{f}_t$,

$$r_{i,t} = \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{f}_t + \varepsilon_{i,t}. \tag{5.1}$$

When n assets are available, this model is still a SUR, and hence the results of the previous sections apply. One change is that the expression for q_{11} does not longer satisfy eq. (4.15). Gibbons et al. (1989) show that for k-factors, this term becomes

$$q_{11} = \frac{1}{T}\big(1 + \bar{\boldsymbol{f}}'\widetilde{\boldsymbol{\Omega}}^{-1}\bar{\boldsymbol{f}}\big), \tag{5.2}$$

where $\widetilde{\boldsymbol{\Omega}}$ is the sample variance matrix of the risk factors

$$\widetilde{\boldsymbol{\Omega}} = \widetilde{\mathrm{var}}[\boldsymbol{f}] = \frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})' = \overline{ff'} - \bar{f}\bar{f}' \tag{5.3}$$

The expression $\bar{\boldsymbol{f}}'\widetilde{\boldsymbol{\Omega}}^{-1}\bar{\boldsymbol{f}}$ corresponds with the square of the Sharpe ratio of the efficient portfolio constructed from the risk factors and the riskless asset.

The GRS-test with multiple factors

Together this means that the GRS-test statistic with n assets and k factors takes the form

$$z = \frac{T-n-k}{n}\frac{1}{1 + \overline{\boldsymbol{f}}'\widetilde{\boldsymbol{\Omega}}^{-1}\overline{\boldsymbol{f}}}\hat{\boldsymbol{\alpha}}'\widetilde{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}} \sim \mathrm{F}_{n,T-n-k}. \tag{5.4}$$

This statistic has an F-distribution with n and $T-n-k$ degrees of freedom. Also for this multifactor case Gibbons et al. (1989) show that the term $\boldsymbol{\alpha}'\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\alpha}$ is the difference of the squared Sharpe ratios of the efficient portfolio of the risk factors with and without the test assets.

A The OLS estimator in a SUR

A.1 The variance of $\hat{B}$

To circumvent the variance of a matrix, the model in eq. (3.4) is vectorized.[6] We derive

$$\begin{aligned}\operatorname{vec}(\boldsymbol{Y}) &= \operatorname{vec}(\boldsymbol{XB}) + \operatorname{vec}(\boldsymbol{E}) \\ &= \operatorname{vec}(\boldsymbol{XBI}_n) + \operatorname{vec}(\boldsymbol{E}) \\ &= (\boldsymbol{I}_n \otimes \boldsymbol{X}) \operatorname{vec}(\boldsymbol{B}) + \operatorname{vec}(\boldsymbol{E}),\end{aligned} \tag{A.1}$$

where $\otimes$ denotes the Kronecker product. In this derivation, we use that for conformable matrix $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$, $\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{B})$.[7] $\operatorname{vec}(\boldsymbol{E})$ is the $nT \times 1$ vector with all stacked error terms. In the classical setting, their joint distribution is given by

$$\operatorname{vec}(\boldsymbol{E}) \sim \mathrm{N}_{nT}\big(\boldsymbol{0}_{nT}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T\big). \tag{A.2}$$

Vectorizing the estimator $\hat{\boldsymbol{B}}$ and substituting eq. (3.4) yields

$$\begin{aligned}\operatorname{vec}(\hat{\boldsymbol{B}}) &= \operatorname{vec}\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\big) \\ &= \operatorname{vec}\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{XB} + \boldsymbol{E})\big) \\ &= \operatorname{vec}\big(\boldsymbol{B} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E}\big) \\ &= \operatorname{vec}\big(\boldsymbol{B}\big) + \operatorname{vec}\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E}\big) \\ &= \operatorname{vec}(\boldsymbol{B}) + \big(\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big) \operatorname{vec}(\boldsymbol{E})\end{aligned} \tag{A.3}$$

This result shows that the estimator $\operatorname{vec}(\hat{\boldsymbol{B}})$ is a linear transformation of $\operatorname{vec}(\boldsymbol{E})$, so it follows a normal distribution as well. Its variance follows as

$$\begin{aligned}\operatorname{var}\big[\operatorname{vec}(\hat{\boldsymbol{B}})\big] &= \mathrm{E}\big[\big(\operatorname{vec}(\hat{\boldsymbol{B}}) - \operatorname{vec}(\boldsymbol{B})\big)\big(\operatorname{vec}(\hat{\boldsymbol{B}}) - \operatorname{vec}(\boldsymbol{B})\big)'\big] \\ &= \mathrm{E}\big[\big(\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big) \operatorname{vec}(\boldsymbol{E}) \operatorname{vec}(\boldsymbol{E})'\big(\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big)'\big] \\ &= \big(\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big) \mathrm{E}\big[\operatorname{vec}(\boldsymbol{E}) \operatorname{vec}(\boldsymbol{E})'\big] \big(\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big)' \\ &= \big(\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big)\big(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T\big)\big(\boldsymbol{I}_n \otimes \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\big) \\ &= \boldsymbol{\Sigma} \otimes \big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{I}_T\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\big) \\ &= \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}.\end{aligned} \tag{A.4}$$

This establishes that

$$\operatorname{vec}(\hat{\boldsymbol{B}}) \sim \mathrm{N}_{kn}(\operatorname{vec}(\boldsymbol{B}), \boldsymbol{\Sigma} \otimes \boldsymbol{Q}). \tag{A.5}$$

[6] Vectorizing the $m \times n$ matrix $\boldsymbol{A}$ means that the columns of $\boldsymbol{A}$ are stacked. The operation is written as $\operatorname{vec}(\boldsymbol{A})$.

[7] See `www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html` for an overview of matrix operations.

A.2 Testing on $\boldsymbol{B}_{j\bullet}$

We first derive the distribution of row j of $\boldsymbol{B}$ which we denote by $\boldsymbol{B}_{j\bullet}$. This row results from

$$\boldsymbol{B}_{j\bullet} = \boldsymbol{e}_j'\boldsymbol{B}, \tag{A.6}$$

where $\boldsymbol{e}_j$ is the unit vector of direction j. The relation with $\operatorname{vec}(\boldsymbol{B})$ follows from

$$\boldsymbol{B}_{j\bullet}' = \operatorname{vec}(\boldsymbol{B}_{j\bullet}) = \operatorname{vec}(\boldsymbol{e}_j'\boldsymbol{B}) = \operatorname{vec}(\boldsymbol{e}_j'\boldsymbol{B}\boldsymbol{I}_n) = (\boldsymbol{I}_n \otimes \boldsymbol{e}_j')\operatorname{vec}(\boldsymbol{B}).,\tag{A.7}$$

The expectation of $\hat{\boldsymbol{B}}_{j\bullet}'$ follows directly from this relation,

$$\mathrm{E}[\hat{\boldsymbol{B}}_{j\bullet}'] = \mathrm{E}[(\boldsymbol{I}_n \otimes \boldsymbol{e}_j')\operatorname{vec}(\hat{\boldsymbol{B}})] = (\boldsymbol{I}_n \otimes \boldsymbol{e}_j')\operatorname{vec}(\boldsymbol{B}) = (\boldsymbol{B}_{j\bullet})'. \tag{A.8}$$

For the variance, we have

$$\begin{aligned}
\operatorname{var}\big[\hat{\boldsymbol{B}}_{j\bullet}'\big] &= \operatorname{var}\big[(\boldsymbol{I}_n \otimes \boldsymbol{e}_j')\operatorname{vec}(\hat{\boldsymbol{B}})\big] \\
&= (\boldsymbol{I}_n \otimes \boldsymbol{e}_j')\operatorname{var}[\operatorname{vec}(\hat{\boldsymbol{B}})](\boldsymbol{I}_n \otimes \boldsymbol{e}_j')' \\
&= (\boldsymbol{I}_n \otimes \boldsymbol{e}_j')(\boldsymbol{\Sigma} \otimes \boldsymbol{Q})(\boldsymbol{I}_n \otimes \boldsymbol{e}_j) \\
&= \boldsymbol{\Sigma} \otimes \boldsymbol{e}_j'\boldsymbol{Q}\boldsymbol{e}_j = q_{jj}\boldsymbol{\Sigma}.
\end{aligned}\tag{A.9}$$

The distribution of $\hat{\boldsymbol{B}}_{j\bullet}'$ then follows as

$$\hat{\boldsymbol{B}}_{j\bullet}' \sim \mathrm{N}_n\big(\boldsymbol{B}_{j\bullet}', q_{jj}\boldsymbol{\Sigma}\big) \tag{A.10}$$

To derive the result that

$$t^2 = (\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})'/q_{jj} \sim \mathrm{T}^2_{n,T-m},$$

I take the following steps. First I define the lower triangular matrix $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}$. I use it to construct the n-variate random variable

$$\boldsymbol{z} = \frac{1}{\sqrt{q_{jj}}}\boldsymbol{\Lambda}^{-1}(\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})' \sim \mathrm{N}_n(\boldsymbol{0}_n, \boldsymbol{I}_n), \tag{A.11}$$

which follows a standard normal distribution. Next I use that when $\boldsymbol{W} \sim \mathrm{W}_n(\boldsymbol{V}, \nu)$ and $\boldsymbol{C}$ is a $k \times n$ matrix of rank k that $\boldsymbol{C}\boldsymbol{W}\boldsymbol{C}' \sim \mathrm{W}_k(\boldsymbol{C}\boldsymbol{V}\boldsymbol{C}', \nu)$ to construct the random variable

$$\boldsymbol{U} = \boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Lambda}^{-1})' \sim \mathrm{W}_n(\boldsymbol{I}_n, T-m), \tag{A.12}$$

which follows a standard Wishart distribution with $T-m$ degrees of freedom, and where the scale matrix results from $\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Lambda}^{-1})' = \boldsymbol{I}_n$. Consequently, I can write t^2 as

$$\begin{aligned}
t^2 &= (\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})'/q_{jj} \\
&= (\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})(\boldsymbol{\Lambda}^{-1})'\boldsymbol{\Lambda}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}(\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})'/q_{jj} \\
&= \frac{1}{\sqrt{q_{jj}}}\left(\boldsymbol{\Lambda}^{-1}(\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})'\right)'\left(\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Lambda}^{-1})'\right)^{-1}\frac{1}{\sqrt{q_{jj}}}\left(\boldsymbol{\Lambda}^{-1}(\hat{\boldsymbol{B}}_{j\bullet} - \boldsymbol{B}_{j\bullet})'\right) \\
&= \boldsymbol{z}'\boldsymbol{U}^{-1}\boldsymbol{z},
\end{aligned}$$

which is a quadratic form of a standard normal variable $\boldsymbol{z}$ and the inverse of a standard Wishart distributed variable $\boldsymbol{U}$. The result follows Hotelling's T^2-distribution with parameters n and $T-m$.

B The relation between the GRS-statistic and Sharpe ratios

Gibbons et al. (1989) show that the GRS-test can be interpreted as a test on the difference between the squared Sharpe ratios of the efficient portfolios based on the factor portfolios alone, and on the combination of test assets and factor portfolios. To find this result, we need again some matrix algebra.

First, we formulate the Sharpe ratio of an efficient portfolio of n assets in terms of their means and variances. We assume that a riskless asset is present. We denote the excess returns by the n-vector $\boldsymbol{r}$, the expected excess returns by the n-vector $\boldsymbol{\mu}$ and their variance by the $n \times n$ matrix $\boldsymbol{S}$. Back (2010, H5.5) shows that the Sharpe ratio of any efficient portfolio is given by

$$\mathrm{Sh}[\boldsymbol{r}] = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}. \tag{B.1}$$

For the factor returns $\boldsymbol{f}$ with mean $\bar{\boldsymbol{f}}$ and variance $\boldsymbol{\Omega}$, this means $\mathrm{Sh}[\boldsymbol{f}] = \sqrt{\bar{\boldsymbol{f}}'\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{f}}}$.

Next, we need the Sharpe ratio of the efficient portfolio of the n test assets AND the factor portfolios. We gather the excess returns of the assets and the factor returns in the $(n+k)$-vector

$$\boldsymbol{r}^* = \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{f} \end{pmatrix}. \tag{B.2}$$

We use the factor model in eq. (1.1) to find expressions for the mean $\boldsymbol{\mu}^*$ and variance $\boldsymbol{\Sigma}^*$ of $\boldsymbol{r}^*$. For the mean we find

$$\boldsymbol{\mu}^* = \begin{pmatrix} \boldsymbol{\alpha} + \boldsymbol{B}'\bar{\boldsymbol{f}} \\ \bar{\boldsymbol{f}}, \end{pmatrix} \tag{B.3}$$

and for the variance

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} \boldsymbol{B}'\boldsymbol{\Omega}\boldsymbol{B} + \boldsymbol{\Sigma} & \boldsymbol{B}'\boldsymbol{\Omega} \\ \boldsymbol{\Omega}\boldsymbol{B} & \boldsymbol{\Omega} \end{pmatrix} \tag{B.4}$$

For the expression of $\mathrm{Sh}[\boldsymbol{r}^*]$ we need the inverse of $\boldsymbol{S}^*$. This matrix is a partitioned matrix. Its inverse is given by

$$\boldsymbol{\Sigma}^{*-1} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & -\boldsymbol{\Sigma}^{-1}\boldsymbol{B}' \\ -\boldsymbol{B}\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Omega}^{-1} + \boldsymbol{B}\boldsymbol{\Sigma}^{-1}\boldsymbol{B}' \end{pmatrix} \tag{B.5}$$

Consequently, we derive

$$\begin{aligned}
\mathrm{Sh}[\boldsymbol{r}^*]^2 =& \boldsymbol{\mu}^{*\prime}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\mu}^* \\
=& \begin{pmatrix} \boldsymbol{\alpha} + \boldsymbol{B}'\bar{\boldsymbol{f}} & \bar{\boldsymbol{f}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & -\boldsymbol{\Sigma}^{-1}\boldsymbol{B}' \\ -\boldsymbol{B}\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Omega}^{-1} + \boldsymbol{B}\boldsymbol{\Sigma}^{-1}\boldsymbol{B}' \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} + \boldsymbol{B}'\bar{\boldsymbol{f}} \\ \bar{\boldsymbol{f}}, \end{pmatrix} \\
=& (\boldsymbol{\alpha} + \boldsymbol{B}'\bar{\boldsymbol{f}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha} + \boldsymbol{B}'\bar{\boldsymbol{f}}) - (\boldsymbol{\alpha} + \boldsymbol{B}'\bar{\boldsymbol{f}})'\boldsymbol{\Sigma}^{-1}\boldsymbol{B}'\bar{\boldsymbol{f}} - \\
& \boldsymbol{f}'\boldsymbol{B}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha} + \boldsymbol{B}'\bar{\boldsymbol{f}}) + \bar{\boldsymbol{f}}'(\boldsymbol{\Omega}^{-1} + \boldsymbol{B}\boldsymbol{\Sigma}^{-1}\boldsymbol{B}')\bar{\boldsymbol{f}} \\
=& \boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} + \bar{\boldsymbol{f}}'\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{f}}.
\end{aligned}$$

From the last equality, the result follows,

$$\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} = \mathrm{Sh}[\boldsymbol{r}^*]^2 - \mathrm{Sh}[\boldsymbol{f}]^2. \tag{B.6}$$

References

Back, K. E. (2010). *Asset Pricing and Portfolio Choice Theory*. Oxford University Press, 1 edition.

Gibbons, M., Ross, S., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57:1121–1151.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education, Upper Saddle River, NJ, USA, 5th edition.

Heij, C., De Boer, P., Franses, P. H., Kloek, T., and Van Dijk, H. K. (2004). *Econometric methods with applications in business and economics*. Oxford University Press, Oxford, UK.

Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2(3):360–378.