Factor Exercises

Question 1:

We want to forecast monthly inflation, y_t, and have at our disposal a data set with many regressors $\mathbf{x}_t = (x_{1t}, ..., x_{Nt})'$ for each time $1 \leq t \leq T$. The data has been preprocessed to ensure that $\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t = 0$ and $\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' = \Sigma$. We are aware of the dangers of estimating large regression models and decided to construct a factor model as follows:

$$\mathbf{f}_t = \mathbf{A}'\mathbf{x}_t.$$

The ith column of $\mathbf{A}$ is equal to $\mathbf{a}_i$, such that the i^{th} row of this equation can be written as

$$f_{it} = \mathbf{a}_i'\mathbf{x}_t = a_{i1}x_{1t} + ...a_{iN}x_{Nt}.$$

a) Assume that $\Sigma\mathbf{a}_i = \lambda_i\mathbf{a}_i$ for all $1 \leq i \leq N$, where λ_i denotes the eigenvalue corresponding to the eigenvector $\mathbf{a}_i$. Assume further that we have ordered the eigenvectors such athat $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$. Under suitable conditions on the eigenvectors show that,

$$Cov(f_{it}, f_{jt}) = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}.$$

b) The first r factors explain most of the variance in our predictor variables, so we decide to use only these r factors for predicting y_{T+1}. First, we estimate r univariate regressions by OLS as follows:

$$y_{t+1} = \alpha_i + \beta_i f_{it} + e_{it},\ 1 \leq i \leq r,\ 1 \leq t \leq T.$$

Our output contains $\widehat{\alpha}_i$, $\widehat{\beta}_i$ for all $1 \leq i \leq r$ as well as the prediction error $\widehat{e}_{it}$ for all $1 \leq i \leq r$, $1 \leq t \leq T$. Second we calculate the sample co-variance of the prediction errors, that is

$$\widehat{\mathbf{\Omega}} = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \widehat{e}_{1t} \\ \vdots \\ \widehat{e}_{rt} \end{pmatrix}\begin{pmatrix} \widehat{e}_{1t} & \cdots & \widehat{e}_{rt} \end{pmatrix}.$$

A colleague argues that $\widehat{\mathbf{\Omega}}$ must be diagonal, because the factors are known to be uncorrelated, is this true?

c) The first factor explains a larger part of the variance of the regressors than any other factor. Does this imply that $\widehat{\mathbf{\Omega}}_{11} \leq \widehat{\mathbf{\Omega}}_{22} \leq ... \leq \widehat{\mathbf{\Omega}}_{rr}$? Prove or disprove this hypothesis. (Note that $\widehat{\mathbf{\Omega}}_{jj}$ denotes the $j-th$ diagonal element of $\widehat{\mathbf{\Omega}}$ for $1 \leq j \leq r$.)

Question 2:

Consider the problem of forecasting output growth y_t using a large number N of predictor variables X_t, where N is of comparable magnitude as the number of time series observations T available for model estimation.

Suppose that (y_t, X_t) has a factor representation with k common latent facors $F_t = (f_{1t}, ..., f_{kt})'$:

$$\begin{aligned} X_t &= \Lambda F_t + e_{xt} \\ y_t &= \beta' F_t + e_{yt} \end{aligned}$$

where Λ is an $(N \times k) - matrix$ of factor loadings, β is a (k×1)- vector and e_{xt} and e_{yt} are white noise processes. Assume furthermore that each of the factors f_{jt}, $j = 1, ..., k$ follows a stationary first-order autoregressive process:

$$f_{jt} = \phi_j f_{jt-1} + \varepsilon_{jt}$$

with $|\phi_j| < 1$ and the error processes $\varepsilon_{jt} \sim NID(0, \sigma_j^2)$, for $j = 1, ..., k$ (that is, they are independent cross-sectionally and over time).

Suppose one wants to forecast output growth y_{T+h} at tiem T for a given horizon $h > 0$ using information that is available in the factors F, as follows:

$$y_{T+h} = \beta_h' F_T + \eta_{T+h}.$$

a) Express β_h and η_{T+h} in terms of the parameters and error processes in (4) - (6).

b) What are the properties of β_h (in particular as a function of the forecast horizon h)?

c) What are the properties of η_{T+h}? What do these imply for, for example, the standard error of the OLS estimate of β_h that is obtained from regressing y_{t+h} on (estimates of) F_t using observations $t = 1, ..., T - h$?

Solutions Question 1:

a) The trick is simply to recall that the eigenvectors of a symmetric real-valued positive semi-definite matrix are orthogonal and without loss of generality, normalized. Thus, $\mathbf{a}_i'\mathbf{a}_j = 0$, for $i \neq j$ (orthogonal) and $\mathbf{a}_i'\mathbf{a}_i = 1$ (normalized). Then it follows that

$$Cov(f_{it}, f_{jt}) = E(\mathbf{a}_i'\mathbf{x}_t\mathbf{x}_t'\mathbf{a}_j) = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_j = \lambda_j\mathbf{a}_i'\mathbf{a}_j.$$

and the desired result then follows by the orthogonality and the normalization conditions mentioned above.

b) Suppose for the sake of illustration that all the first r factors are useless for the prediction of y_{t+1}. If we were very unlucky, this would happen when y_t is equal to the (r+1)-th factor. Because the factors are orthogonal by construction, we would find that $\widehat{\beta}_i = 0$ for all $1 \leq i \leq r$ and $\widehat{\alpha}_1 = ... = \widehat{\alpha}_r = \frac{1}{T}\sum_{t=1}^{T} y_{t+1}$. This implies the prediction errors $\widehat{e}_{it} = y_{t+1} - \frac{1}{T}\sum_{t=1}^{T} y_{t+1}$ for all i are identical, hence perfectly correlated, and therefore $\widehat{\mathbf{\Omega}}$ can in general not be diagonal.

A more standard argument is as follows. Notice that the factors f_{it} sum to zero (across time) because x_{it} sum to zero (across time). OLS ensures $\sum_{t=1}^{T} \widehat{e}_{it} = 0$ from which it follows that $\widehat{\alpha}_i = \widehat{\alpha}$ for $1 \leq i \leq r$, where $\widehat{\alpha} = \frac{1}{T}\sum_{t=1}^{T} y_{t+1}$. Consequently we have

$$\begin{aligned}
\sum_{t=1}^{T}[\widehat{e}_{it}\widehat{e}_{jt}] &= \sum_{t=1}^{T}\left[\left(y_{t+1} - \widehat{\alpha} - \widehat{\beta}_i f_{it}\right)\left(y_{t+1} - \widehat{\alpha} - \widehat{\beta}_j f_{jt}\right)\right] \\
&= \sum_{t=1}^{T}(y_{t+1} - \widehat{\alpha})^2 - \widehat{\beta}_i \sum_{t=1}^{T}(y_{t+1} - \widehat{\alpha})f_{it} - \widehat{\beta}_j \sum_{t=1}^{T}(y_{t+1} - \widehat{\alpha})f_{jt}
\end{aligned}$$

where the fourth term disappears because the factors are uncorrelated. The same calculation could equivalently be performed like this:

$$\begin{aligned}
Cov[\widehat{e}_{i,1:T}\widehat{e}_{j,1:T}] &= Cov\left(y_{2:T+1} - \widehat{\alpha} - \widehat{\beta}_i f_{i,1:T}; y_{2:T+1} - \widehat{\alpha} - \widehat{\beta}_j f_{j,1:T}\right) \\
&= Var\left(y_{2:T+1}\right) - \widehat{\beta}_i Cov(y_{2:T+1}; f_{i,1:T}) - \widehat{\beta}_j Cov(y_{2:T+1}; f_{j,1:T}),
\end{aligned}$$

where we have used $Var\left(y_{2:T+1}\right)$ to indicate the sample variance of $y_2, ..., y_{T+1}$ and similarly for the other terms. Proceeding with this notation the use of OLS implies that $\widehat{\beta}_i$ can be written as

$$\widehat{\beta}_i = \frac{Cov(y_{2:T+1}; f_{i,1:T})}{Var(f_{i,1:T})} = \frac{Cov(y_{2:T+1}; f_{i,1:T})}{\lambda_i}, \ 1 \leq i \leq r.$$

Substituting this into the expression above gives

$$Cov[\widehat{e}_{i,1:T}\widehat{e}_{j,1:T}] = Var\left(y_{2:T+1}\right) - \frac{Cov(y_{2:T+1}; f_{i,1:T})^2}{\lambda_i} - \frac{Cov(y_{2:T+1}; f_{j,1:T})^2}{\lambda_j}.$$

This is not in general equal to zero. This expression may be simplified even further by using the definition of the correlation coefficient, defined as $Cor(y_{2:T+1}; f_{i,1:T}) = Cov(y_{2:T+1}; f_{i,1:T})/\sqrt{Var\left(y_{2:T+1}\right)\lambda_i}$. Then we obtain,

$$Cov[\widehat{e}_{i,1:T}\widehat{e}_{j,1:T}] = Var\left(y_{2:T+1}\right)\left[1 - Cov(y_{2:T+1}; f_{i,1:T})^2 - Cov(y_{2:T+1}; f_{j,1:T})^2\right].$$

Indeed, when both factors i and j are uncorrelated with y_{t+1}, we find that $Cov\left(\widehat{e}_{i,1:T}\widehat{e}_{j,1:T}\right) = Var\left(y_{2:T+1}\right)$. On the otherhand, $Cov\left(\widehat{e}_{i,1:T}\widehat{e}_{j,1:T}\right)$ goes to zero when one factor is perfectly correlated with y_{t+1} in which case the other factor must be useless.

c) First, using $Var(\widehat{e}_{i,1:T})$ to denote the sample variance of $\widehat{e}_{i1}$ through $\widehat{e}_{iT}$, we have

$$\begin{aligned}
\Omega_{ii} &= Var(\widehat{e}_{i,1:T}) \\
&= Var\left(y_{2:T+1} - \widehat{\alpha} - \widehat{\beta}_i f_{i,1:T}\right) \\
&= Var\left(y_{2:T+1}\right) + \widehat{\beta}_i^2 Var\left(f_{i,1:T}\right) - 2\widehat{\beta}_i Cov\left(y_{2:T+1}; f_{i,1:T}\right)
\end{aligned}$$

where we have used the simple rule $Var(A+B) = Var(A)+Var(B)+2Cov(A,B)$. Second, recall that using OLS implies that

$$\widehat{\beta}_i = \frac{Cov(y_{2:T+1}; f_{i,1:T})}{Var(f_{i,1:T})} = \frac{Cov(y_{2:T+1}; f_{i,1:T})}{\lambda_i},\ 1 \leq i \leq r.$$

This can be substituted into the above expression to give,

$$\begin{aligned}
\Omega_{ii} &= Var\left(y_{2:T+1}\right) + \frac{Cov(y_{2:T+1}; f_{i,1:T})^2}{\lambda_i} - 2\frac{Cov(y_{2:T+1}; f_{i,1:T})^2}{\lambda_i} \\
&= Var\left(y_{2:T+1}\right) - \frac{Cov(y_{2:T+1}; f_{i,1:T})^2}{\lambda_i}
\end{aligned}$$

By recalling the definition of the correlation coefficient that is $Cor(y_{2:T+1}; f_{i,1:T}) = Cov(y_{2:T+1}; f_{i,1:T})/\sqrt{Var\left(y_{2:T+1}\right)\lambda_i}$ we can conclude that

$$\Omega_{ii} = Var\left(y_{2:T+1}\right)\left[1 - Cor(y_{2:T+1}; f_{i,1:T})^2\right].$$

The conclusion, perhaps unsurprising, is that the factor i with the highest absolute correlation with the data $y_{2:T+1}$ has the lowest diagonal elements Ω_{ii}. Conversely, any facor that is uncorrelated with the $y\acute{s}$ will produce diagonal element in Ω that is equal to the variance of y_t. Clearly, it is not necessary for first factor to have the highest correlation with the $y_t's$ since the factors were constructed of the variable y_t to begin with.

Solution for **Question 2:**

a) To relate the value of output growth y at $t+h$ to the value of the factors F at t, start from (?) and recursively substitute the factors using equation ():

$$\begin{aligned} y_{t+h} &= \beta' F_{t+h} + e_{y,t+h} \\ &= \beta'(\Phi F_{t+h-1} + e_{f,t+h}) + e_{y,t+h} \\ &= \beta'(\Phi(\Phi F_{t+h-2} + e_{f,t+h-1}) + e_{f,t+h}) + e_{y,t+h} \\ &\vdots \\ &= \beta' \Phi^h F_t + \sum_{i=0}^{h-1} \beta' \Phi^i e_{f,t+h-i} + e_{y,t+h} \end{aligned}$$

where Φ is a $(k \times k)$-diagonal matrix with the AR-coefficients ϕ_j in the models for the factors on the diagonal, and $e_{ft} = (\varepsilon_{1t}, ..., \varepsilon_{kt})'$. Hence,

$$\beta_h = \beta' \Phi^h$$

and

$$\eta_{T+h} = \sum_{i=0}^{h-1} \beta' \Phi^i e_{f,t+h-i} + e_{y,t+h}$$

b) To interprete the coefficients β_h take the single-factor case where $k = 1$. Then Φ is just a scalar, where we should have $|\Phi| < 1$ such that the factor is stationary. Then $|\beta_h| < |\beta_{h-1}|$ for all $h > 1$, that is the predictive power of the current factor for future values of y decreases with the forecast horizon h. This generalises directly to the multi-factor case with $k > 1$ given that all factors f_{jt}, $j = 1, 2, ..., k$ follow a univariate $AR(1)$ process.

c) Due to the summation of h 'factor shocks' $e_{f,t+h-i}$, $i = 0, 1, ..., h-1$, η_{t+h} will be serially correlated up to the order $h-1$ for $h > 1$, unless $\beta = 0$ or $\Phi = 0$. Hence, when β_h is estimated using OLS by regressing y_{t+h} on F_t we should preferably use Newey-West standar errors which correct for the rpesence of autocorrelation in the residuals.

Also, the variance of the errors η_{t+h} increases with horizon h.