ADVANCED TIME SERIES ECONOMETRICS LAB 3

Ping Wu

University of Strathclyde

ping.wu@strath.ac.uk

should I use PCA or factor analysis to obtain my factors?

- Example 1: I want to include a measure of corruption in my regression. But I have 5 correlated measures of corruption for each country.
- Instead of picking one measure, I can use PCA to reduce the observed variables to a small number of uncorrelated factors which can be included in my regression.
- Example 2: I want to examine the effect of shocks to financial conditions. I have 18 correlated measures of financial variables for each country.
- I can use PCA to reduce the observed variables to one factor.
- We can call this factor a "financial conditions index" and include it in a FAVAR (see Koop and Korobilis, 2014, EER).

should I use PCA or factor analysis to obtain my factors?

- Example 3: I am examining business cycle fluctuations and believe that conditions in each country are driven by a world, region and country-specific factor (see Kose et al., 2003, AER).
- PCA has a specific set of identifying assumptions. But I want to make specific identifying assumptions so that my factors have an economic interpretation.
- Put differently, I knew how I want to interpret my factors BEFORE beginning estimation.

Exercise 1

- type "help mvreg".
- We are going through the example on slides 16-18.
- We run the following regression for each company:
  $r_{it} = \alpha_i + \beta_i r_t^m + \varepsilon_t$
- For each company:
- Check whether α_i and β_i are statistically significant.
- IF they are statistically significant, check their magnitude.
- Then consider the economic interpretation of your results in terms of the company's performance and volatility relative to the market.
- Interpret the R_i^2

Moving towards exercises 2 and 3

- We run the following regression for each company: $r_{it} = \alpha_i + \beta_i r_t^m + \varepsilon_t$
- Here, the excess market return (proxied by the S&P 500) represents "the market factor". Our "factor" is not estimated and is just given by one observable variable.
- What happens if you want to assume there is something common to all assets without specifying exactly what it is?
- What happens if the factor(s) are not observable?
- We use $\mathrm{PCA}/$ factor analysis $\rightarrow$ Exercise 2 and 3

Principal Components and the Static Factor model: Checklist

- Explain what the columns "eigenvalues, difference, proportion, cumulative" describe and the information this gives the researcher.
- Explain what the table "eigenvectors" (for PCA) and "factor loadings" (for factor analysis) describes and the information this gives the researcher.
- Hint: we might be able to use this information to interpret our factors economically.
- Explain how many factors should be included in the model and why.

Exercise 2 Code

```
clear
import excel Excess_Returns.xlsx, first
pca AA AGE CAT F FDX GM HPQ KMB MEL NYT PG TRB TXN
screeplot
```

Exercise 2 Output

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	4.62081	3.22198	0.3554	0.3554
Comp2	1.39882	.349711	0.1076	0.4630
Comp3	1.04911	.0681611	0.0807	0.5437
Comp4	.980952	.0726996	0.0755	0.6192
Comp5	.908252	.0748084	0.0699	0.6891
Comp6	.833444	.116074	0.0641	0.7532
Comp7	.71737	.0969486	0.0552	0.8084
Comp8	.620421	.147048	0.0477	0.8561
Comp9	.473373	.0525522	0.0364	0.8925
Comp10	.420821	.0273435	0.0324	0.9249
Comp11	.393477	.0949095	0.0303	0.9551
Comp12	.298568	.013991	0.0230	0.9781
Comp13	.284577	.	0.0219	1.0000

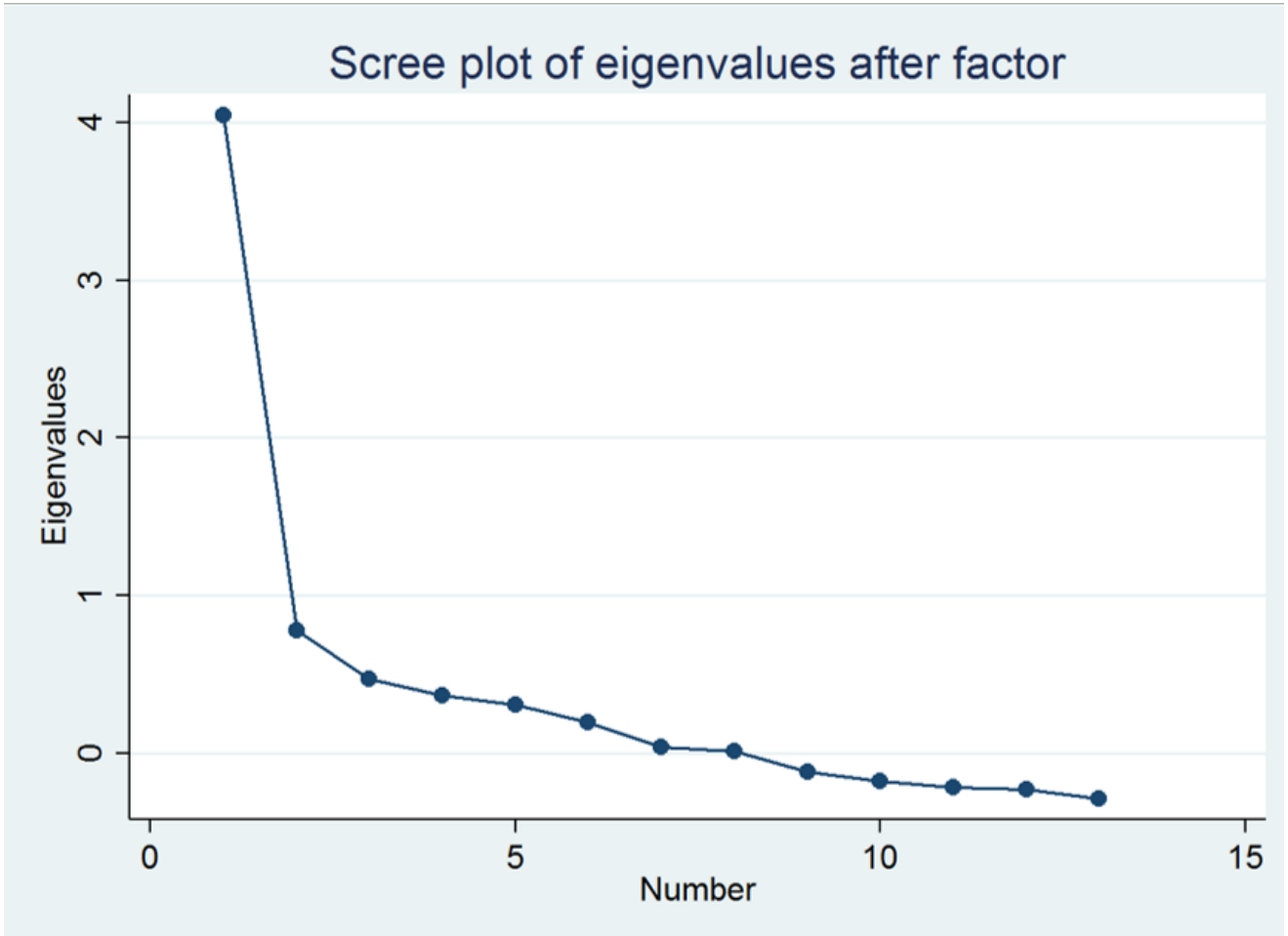
- Eigenvalue = variance of the principal component.
- Difference = difference between the current and subsequent eigenvalue.
- Proportion = proportion of total variability in data explained by component.
- E.g. 35% of variability in the data is explained by principal component 1.
- Cumulative = cumulative proportion of total variability in data explained by current component plus all previous components.

Exercise 2 Optional slide: eigenvalue to proportion

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	4.62081	3.22198	0.3554	0.3554
Comp2	1.39882	.349711	0.1076	0.4630
Comp3	1.04911	.0681611	0.0807	0.5437
Comp4	.980952	.0726996	0.0755	0.6192
Comp5	.908252	.0748084	0.0699	0.6891
Comp6	.833444	.116074	0.0641	0.7532
Comp7	.71737	.0969486	0.0552	0.8084
Comp8	.620421	.147048	0.0477	0.8561
Comp9	.473373	.0525522	0.0364	0.8925
Comp10	.420821	.0273435	0.0324	0.9249
Comp11	.393477	.0949095	0.0303	0.9551
Comp12	.298568	.013991	0.0230	0.9781
Comp13	.284577	.	0.0219	1.0000

- THIS IS OPTIONAL MATERIAL
- Eigenvalue = variance of the principal component.
- Sum of eigenvalues = 13.
- Proportion = Eigenvalue/sum of eigenvalues.
- E.g. Proportion of component 1 = 4.62/13 = 0.3554
- See the STATA manual for more details.

Exercise 2 Output



Exercise 2 Output

- Eigenvectors (output not displayed) = weighting of each company's returns in constructing each component. Please see Gary's slides 28 – 29 for an example of how to interpret eigenvectors.
- Unexplained (output not displayed) is 0 because the 13 principle components explain all variance in all variables.
- Scree plot plots eigenvalue for each component. "Elbow" occurs at 3 so retain first 2 factors.

Exercise 3 TIPs

- I will let you work through exercise 3, but let me pick up on some points from the lab:
- In static factor analysis, factor loadings can be interpreted in a similar way to eigenvectors.
- Factor loadings = weighting of each company's returns in constructing each factor.
- In static factor analysis, uniqueness can be interpreted in a similar way to unexplained.
- Uniqueness = for each variable, the variance that is unexplained by the factors in our model.

Exercise 3: Choosing factors based on information criteria

- Maximum likelihood estimation allows us to obtain IC. These provide an alternative to a scree plot, when trying to determine the number of factors to include.
- For exercise 3:
- The AIC suggests we should have 5 factors.
- The BIC suggests we should have 1 factor.
- The scree plot suggests that we should have 2 factors.
- You could comment on the proportion of variability explained by the factors to help you make a decision. But in this example it's a bit difficult since we have negative proportions. Ideally, you would want to drop factors 9 – 13 and then re-estimate.

Exercise 3: Choosing factors based on information criteria

- Overall, our results suggests that we should include between 1 – 5 factors. In empirical practise, we would analyse how our results changed if we gradually increased the number of factors up to 5. We would also take into account the existing literature. If the factors have an economic interpretation, we may also wish to retain more even if they don't explain much variation in the data.

Exercise 3: More information on unexplained/uniqueness

- Usually, if we have 13 variables and allow STATA to produce 13 components or factors:
- The last row of cumulative proportion $= 1$ and unexplained/uniqueness $= 0$.
- All variability in the data will be explained by the factors.
- Unexplained/uniqueness does not equal 0 when:
- We limit the number of factors.
- E.g. We tell STATA to only keep the first 4 components or factors.
- This means that some of the variation in our 13 variables will not be explained.

Exercise 3: More information on unexplained/uniqueness

- We have negative eigenvalues/proportions.
- If a component or factor has a negative proportion is should be excluded from the model - it is not improving model fit.
- This means that some of the variation in our 13 variables will not be explained.