Seminarpresentation

November 20, 2018

0.1   1. Exploring political differences through latent factor analysis

- Motivation
  - Political spectrum is usually considered 1-dimensional
  - Latent factor analysis has in other instances been used to understand human actions and motivatation

- Overview of project
  - Survey data
  - Naive Bayesian latent factor analysis
  - Extension with ordered probit to accomadate the discrete nature of the data

0.2   2. Introduktion til Faktormodeller

$$\underset{(p\times 1)}{y_i} = \underset{(p\times k)}{\beta}\ \underset{(k\times 1)}{F_i} + \underset{(p\times 1)}{\epsilon_i} \tag{1}$$

for $i = 1, ..., T$ individuals. β denotes the factor loadings and F are the factor scores

$$\epsilon \sim I.I.D.N(0, \Sigma) \tag{2}$$

$$\underset{(p\times p)}{var(y)} = \underset{(p\times k)(k\times k)(k\times p)}{\beta\psi\beta'} + \underset{(p\times p)}{\Sigma} = \underset{(p\times p)}{\Omega} \tag{3}$$

where a normal assumption is...

$$\psi = I_k \tag{4}$$

0.3   3. Naive Factor model (Jonas)

0.4   4. Naive Factor model 2 (Jonas)

0.5   5. Ordered probit Factor Model

$$Y_i = l \text{ if } \tau_{l-1} \le Y_i^* \le \tau_l \text{ for l=1,2,3,4} \tag{5}$$

$$Y_i^* = \beta F_i + \epsilon_i \tag{6}$$

Posterior:

$$P(\beta|Y^*, Y, \tau, F) \propto P(Y|Y^*, \tau)P(Y^*|\beta, F)P(\beta)P(F) \tag{7}$$

0.6 6. Probit Factor Model

In order to properly identify the model we do the following:

$$\Sigma = I_p \tag{8}$$

If we include a constant term, we would have to fix one of our $\tau = 0$, but we don't

$$Y^* \sim TN_{[\tau_l, \tau_{l+1}]}(\beta F; 1) \tag{9}$$

$$\tau_l \sim U[\underline{\tau_l}; \bar{\tau}_l] \tag{10}$$

0.7 7. Data, simulations and factor loading (Jeppe)

- Data

  - Survey data from municipality elections 2017 (KV17)
  - Politicians answered questionaire of 15 questions by the danish broadcasting service
  - 1200 answers by the largest danish parties used in analysis
  - dataset of size 15 variables and 1200 rows

- Simulations

  - To test if our algorithm converges 2 test data strategies utilized:
  - Strategy 1:
    * simulate from 3 underlying factors
  - Strategy 2:
    * Use cholesky decomposition to create dataset with correct covariance matrix
  - Both had fast convergence

- Factor loading

  - Upper triangular matrix require we think about the ordering of the questions
  - Which question loads to which matrix
  - 2 first questions should be different in nature (which latent factor the questions represent)
    * loads only to factor 0: *Municipality tax should be reduced*
    * loads only to factor 0 & 1: *Institutions run by local authorities takes to much into consideration the concerns of religious minorities.*

0.8 8. Results Naive - Trace and distribution of β

```
<td> <img src="../figs/estimation_trace_beta_plot.png", width = 400> </td>
<td> <img src="../figs/estimation_dist_beta_plot.png", width = 400> </td>
```

0.9 9. Results Naive - scatter plots 1

```
<td> <img src="../figs/estimation_scatter_la_el.png", width = 400> </td>
<td> <img src="../figs/estimation_scatter_rv_df.png", width = 400> </td>
```

0.10 9. Results Naive - scatter plots 2

```
<td> <img src="../figs/estimation_scatter_v_c.png", width = 400> </td>
<td> <img src="../figs/estimation_scatter_v_df.png", width = 400> </td>
```

0.11 10. Feedback/issues

- Computation issues
- Factor loading matrix - not what expected
- Unknown unknowns