Package Name: StatFact
Date: 20th October 2014
Add-in Type: Group
Default Proc Name: Statfact
Interface: Dialogue (group proc) and command line

Description:

The common factors are estimated by the method of asymptotic principle components (the maximum number of which is $min\{N,T\}$). The estimated factor matrix ($T \times k$, where k is the number of factors), $\tilde{F}^k$, is $\sqrt{T}$ times the eigenvectors which correspond to the k largest eigenvalues of the $T \times T$ matrix XX' when $T < N$. Alternatively, when $T > N$, the factor matrix is calculated $\bar{F}^k = \frac{X\bar{\Lambda}^k}{N}$, where as the loadings matrix ($\bar{\Lambda}^k$) is calculated as $\sqrt{N}$ multiplied by the eigenvectors corresponding to the k largest eigenvalues of $X'X$. Please see Stock and Watson (2002) for more details. The series which form the group which is then converted into matrix $X'X$ or XX' are also optionally first differenced and/or standardized (mean subtracted and divided by series standard deviation), as per the Gauss files of Igor Maston as per Banerjee et al. (2014). Six criteria proposed by Bai and Ng (2002) are optionally calculated. Their general form is: $PC(k) = V(k,\hat{F}^k) + kg(N,T)$, where $V(k,\hat{F}^k)$ is the sum of squared residuals from estimating k factors (in a dynamic factor model representation), $g(N,T)$ is a necessary penalty function, and we are interested in estimating the most suitable dimension ($r < k\_max$) of $\hat{F}$.

Command Line Syntax:

group_object.statfact(options) *factornames* #

Where *factornames* is the naming pre-fix that will be used to create factor series (e.g. factornames_1, factornames_2, ..., factornames_#) and # is maximum number of factors permitted in the routine. The options are as follows:

- ipc Report Bai and Ng (2002) PC, IPC, AIC3 and BIC3 criteria.
- d Take first differences of the group.
- m Subtract series mean and divide by st. dev.
- l Output loadings from 1 to #.
- n Normalize factors\loadings by element (#,1) of loadings matrix.
- g Output graphs of factors, loadings, scree and % explained.

Example

The 'complete' example, using a group in our workfile called *variablegroup* which has all the variables from which we want to extract static factors:

variablegroup.statfact(ipc, d, m, l, n,g) factor 6

This command line example (akin to inputting the *variablegroup* in the user interface from the group Proc with 6 factors and output of 'factor' then ticking all box) will estimate the factors using all possible options. In the order of the add-in: it will first de-mean and divide by the standard deviation, then take first differences of all series in the group, then calculate the factors (and if checked - also the loadings) as above. It will then calculate the PC1, PC2, PC3, IPC1, IPC2, IPC3, AIC3 and BIC3 criteria from Bai and Ng (2002) as above, and put all values in a table format which will automatically be shown (to suppress output, simply delete or comment out the 'show' commands in the statfact.prg). The example will then plot the factors, loadings, a scree plot of all eigenvalues and then a bar chart which details the percent of the variance explained by each factor (which is simply the eigenvalue associated with the eigenvector used in the creation of that factor divided by the sum of all eigenvalues).

Check with other code

First, to check that this code is correctly estimating the factors, we compare it with the outputs of the 'extract.m' file of Jean Boivin, related to the paper Bernanke et al. (2005). The .m files associated with this are available at:

http://neumann.hec.ca/pages/jean.boivin/mypapers/BBE_Ddisk.zip

In addition to this, we are going to cross-check our Bai and Ng (2002) criteria using the dataset and the MATLAB files provided by Christophe Hurlin for the Bai and Ng (2002) criteria:

http://www.runmycode.org/coder/view/151

Load the dataset from this link (Data_NG.xls) and follow the commands below:

```
group factorgroup *
factorgroup.drop resid
factorgroup.statfact(ipc) factor 6
```

From this we can see that a group of the factor outputs here are the same as the matrix *factors* generated by command $[factors, loadings] = extract(factorgroup, 6)$ using the *extract.m* file. Comparing against the second link, we can also see that the table of all the EViews Bai and Ng (2002) output criteria are the same as the command *[results]=NbFactors(factorgroup,6)* using the Hurlin code (*NbFactors.m*)

References

Bai, J., Ng, S., 2002. Determining the Number of Factors in Approximate Factor Models. Econometrica 70, 191–221.

Banerjee, A., Marcellino, M., Masten, I., 2014. Forecasting with factor-augmented error correction models. International Journal of Forecasting 30, 589–612.

Bernanke, B., Boivin, J., Eliasz, P.S., 2005. Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach. The Quarterly Journal of Economics 120, 387–422.

Stock, J., Watson, M., 2002. Forecasting Using Principal Components From a Large Number of Predictors. Journal of the American Statistical Association 97, 1167–1179.