Factor-Augmented VAR (FAVAR): Simulation and Econometric Details

May 15, 2025

Overview

- Motivation for Factor-Augmented VAR (FAVAR)
- Simulation setup with artificial data
- Econometric methodology
- Estimation steps
- Impulse response computation for excluded variables

Why FAVAR?

- Traditional VAR models are limited in dimensionality
- Many macro variables are observed but not included
- FAVAR leverages latent factors to summarize large datasets
- Improves estimation and inference in macroeconomic settings

Simulation Setup

- Generate $N = 15$ macro variables over $T = 200$ time periods
- Underlying structure: $X_t = \Lambda F_t + e_t$
- F_t are 3 latent factors, Λ is a loading matrix
- Add idiosyncratic noise e_t to each variable

Estimation Steps

1. Standardize 14 macro variables (exclude 15th)
2. Apply PCA to extract $\hat{F}_t$ (first 3 factors)
3. Construct VAR with $[\hat{F}_t, X_t^{(15)}]$
4. Estimate VAR model using least squares
5. Compute IRFs for $\hat{F}_t$ and $X_t^{(15)}$

IRFs for Excluded Variables

- Original data: $X_t = \Lambda \hat{F}_t + e_t$
- IRFs for excluded variables: $IRF_X(h) = \Lambda \cdot IRF_F(h)$
- Use PCA loadings (Λ) and factor IRFs to project responses
- Include direct response of $X_t^{(15)}$ from VAR

Econometric Details: Constructing IRFs for Excluded Variables

- After estimating the VAR for $[\hat{F}_t, X_t^{(15)}]$, extract $IRF_{\hat{F}}(h)$ and $IRF_{X^{(15)}}(h)$.
- Let $\hat{\Lambda}$ be the PCA loading matrix from the 14 variables.
- Then, for each horizon h and each shock s:

$$IRF_{X_i}(h, s) = \hat{\lambda}_i' \cdot IRF_{\hat{F}}(h, s)$$

  where $\hat{\lambda}_i$ is the i-th row of the loading matrix.
- This reconstructs the IRFs for each excluded variable X_i $(i = 1, ..., 14)$.
- $X^{(15)}$ is included directly in the VAR, so its IRF is taken as estimated.

PCA vs. EIG in FAVAR

- **pca()**: MATLAB's high-level function, based on SVD, numerically stable and recommended for FAVAR.
- **eig()**: Lower-level function applied to covariance matrix, requires manual sorting and standardization.
- **Key distinction**: pca() ensures orthonormality of factor loadings and orders components by explained variance.
- For high-dimensional or noisy macro data, pca() is more robust and preferred.

When to Use $N^{-1}\Lambda'\Lambda$

- When PCA is applied to **unstandardized** data: the loading matrix Λ may not be orthonormal.
- In **asymptotic factor models** (e.g., Stock & Watson): $\frac{1}{N}\Lambda'\Lambda$ converges to a finite matrix.
- When projecting X to estimate standardized factors: $\hat{F}_t = \frac{1}{N}\Lambda' X_t$.
- Helps recover consistent IRFs across high-dimensional data.
- Not needed if using MATLAB's `pca()` on standardized data — orthonormality is already ensured.

Summary

- FAVAR effectively captures dynamics from high-dimensional data
- Simulation demonstrates estimation with latent and observed components
- IRFs of excluded variables recovered via factor structure
- Approach scalable to large datasets (e.g., FRED-MD)

Example: Using Factors as Instruments (2SLS)

Goal: Estimate the causal effect of an endogenous variable x_t on an outcome y_t.

Problem: x_t may be correlated with the error term in the structural equation:

$$y_t = \beta x_t + u_t$$

Solution: Use latent factors $\hat{F}_t$ as instruments for x_t:

1. Estimate factors $\hat{F}_t$ using PCA on a large panel Z_t.
2. First stage: regress x_t on $\hat{F}_t$ to get $\hat{x}_t$.
3. Second stage: regress y_t on $\hat{x}_t$.

Assumption: $\hat{F}_t$ are correlated with x_t but uncorrelated with u_t.

Interpretation and Applications

- Factors act as proxies for latent economic forces (e.g. monetary policy stance, business cycle).
- Especially useful in settings where traditional instruments are weak or unavailable.
- Common in monetary policy analysis, where policy shocks are hard to isolate.
- This method links well to the "proxy SVAR" and "external instruments" literature.

Quiz: Understanding Factor Models and FAVARs (Part 1)

1. What is the main purpose of extracting factors from macroeconomic datasets?
2. How does PCA help summarize high-dimensional data?
3. Why is it important to standardize variables before applying PCA?
4. How are extracted factors used in forecasting?
5. How can factors be used to construct instruments for 2SLS estimation?

Quiz: Understanding Factor Models and FAVARs (Part 2)

6. What variables are included in a FAVAR model?
7. How do you compute IRFs for variables not directly included in the VAR?
8. What conditions must be satisfied for the PCA loadings to be orthonormal?

Quiz: Advanced Concepts in Factor Models (Part 3)

1. How can factor models help in nowcasting?
2. What is the role of idiosyncratic errors in factor models?
3. How can factor rotation affect interpretability?
4. What is the difference between static and dynamic factor models?

Quiz: FAVAR Models (Part 4)

1. What distinguishes a FAVAR from a traditional VAR model?
2. Why are latent factors included in a FAVAR, and how are they estimated?
3. What assumptions are required for factors to adequately capture omitted variable effects in a VAR?
4. How does including factors improve the identification of structural shocks?
5. How can FAVARs be estimated in high-dimensional settings?
6. What is the effect of the number of factors on FAVAR performance?
7. How do factor estimation errors affect impulse response interpretation?

Quiz Answers (1–5)

1. To reduce dimensionality while preserving common variation across many economic indicators.
2. PCA finds orthogonal components (factors) that explain the maximum variance in the data.
3. To ensure all variables contribute equally to the factor extraction process.
4. The extracted factors can be used as predictors in regression models.
5. Factors can proxy for latent variables, serving as strong instruments when exogenous variation is hard to isolate.

Quiz Answers (6–10)

6. Observed variables of interest (e.g. inflation, interest rate) and latent factors.
7. Use the PCA loading matrix to project IRFs of factors back onto the macro variables.
8. When PCA is applied to standardized data, the loading matrix columns are orthonormal.

Quiz Answers (11–15)

11. By capturing real-time updates from high-frequency indicators (e.g. industrial production, surveys).
12. Idiosyncratic errors capture variable-specific shocks not explained by common factors.
13. Rotating factors can improve interpretability but does not change the information content.
14. Static factor models assume contemporaneous relationships; dynamic ones allow lag structures.
15. A Bayesian approach can incorporate prior beliefs and shrinkage, especially useful with small samples.

Quiz Answers (16–23)

16. A FAVAR includes both latent factors and observed macro variables in the VAR system.
17. Latent factors summarize broad macroeconomic information; estimated using PCA on a large panel.
18. Factors should explain the common variation and be uncorrelated with the VAR residuals.
19. They help capture broader economic shocks and reduce omitted variable bias.
20. Informational variables provide data richness for factor extraction beyond the observed VAR variables.
21. By using PCA and dimensionality reduction techniques to avoid overfitting.
22. Too few factors miss key variation; too many introduce noise and overfitting.
23. Estimation error in factors can attenuate IRFs and affect inference.

References

- Bernanke, Boivin, Eliasz (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach
- Stock & Watson (2002). Forecasting using principal components