

Factor-augmented VARs

Tyler Sotomayor

2024

A two-variable example makes the factor normalization and reconstruction explicit. The remaining solutions derive an observable recursive rotation, recover level responses from differenced variables, and show exactly how a metadata row can contaminate the estimation sample.

Based on the 2024 BSE course taught by Luca Gambetti and Konstantin Boss. I have revised the classroom code and added the accompanying explanations. The notes identify the papers, data, and assumptions used in each application. These are my course notes, not an official BSE publication. A classroom replication should not be read as a replication of every result in its underlying paper.

1 Normalization and reconstruction

The sample covariance is

$$\frac{Z'Z}{3-1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Its eigenvalues are two and zero. Choose $v_1 = (1, 1)'/\sqrt{2}$, whose Euclidean norm is one. The first factor is $F = Zv_1 = (-\sqrt{2}, 0, \sqrt{2})'$. Its mean is zero and its sample variance is $(2 + 0 + 2)/2 = 2$, not one.

The common component Fv_1' equals Z exactly, so the idiosyncratic component is the zero matrix. If instead the factor were divided by $\sqrt{2}$ to have unit sample variance, its loading vector would need to be multiplied by $\sqrt{2}$. Normalization changes coordinates, not the fitted common component, provided both sides of the decomposition change together.

2 An observable rotation

First,

$$B_0B_0' = \begin{pmatrix} 5 & 5 \\ 5 & 10 \end{pmatrix}, \quad L = \begin{pmatrix} \sqrt{5} & 0 \\ \sqrt{5} & \sqrt{5} \end{pmatrix}.$$

Solving the two linear systems $B_0h_j = L_{.j}$ gives

$$H = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}.$$

Indeed $H'H = \frac{1}{5} \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = I$ and $B_0H = L$. The first observable has zero impact response to the second shock, because $L_{12} = 0$. The second observable may respond immediately to the first shock. Reversing the observable order would impose a different economic restriction, even though each ordering reproduces the same reduced-form covariance.

3 Cumulating responses

The first cumulative sum of $(1, 0, 0)$ is $(1, 1, 1)$, the response of the first difference. The second cumulative sum is $(1, 2, 3)$, the level response. A one-time acceleration therefore produces a persistent change in growth and a growing level effect under this simple response path.

For the first difference of a log series, one cumulative sum gives $(0.01, 0.008, 0.008)$. Multiplying by 100 yields approximate percentage responses $(1, 0.8, 0.8)$. The exact proportional level change is $100[\exp(b_h^{log}) - 1]$, approximately $(1.0050, 0.8032, 0.8032)$ percent. The figure uses the explicitly labelled log approximation, not the exact nonlinear transformation.

4 A metadata row

The preferred procedure first removes the metadata row, leaving 756 observations. Dropping two unavailable initial differences then gives $756 - 2 = 754$ usable months. Counting metadata as an observation creates 757 rows, after which dropping two leaves 755.

For real values x_1, x_2 and metadata value m , the erroneous second difference attached to x_2 is $x_2 - 2x_1 + m$. It survives the deletion of the first two rows because it is in the third row of the augmented array. The valid second difference first exists at x_3 , where it equals $x_3 - 2x_2 + x_1$. Thus the wrong rule introduces both an extra observation and an artificial initial transformed value for second-differenced series. The preferred and compatibility samples must be stored separately.