

Multivariate State Space and
Dynamic Factor Models
Belgische Francqui-leerstoel 2019-2020
Universiteit Antwerpen

Siem Jan Koopman

Vrije Universiteit Amsterdam, Tinbergen Institute

Tuesday 15 December 2020
<http://sjkoopman.net>

Multivariate state space time series models

Program :

- Introduction
- Multivariate Local Level model
- Reduced rank and cointegration
- Kalman filter, equation by equation
- Dynamic factor models
- Collapsed Kalman filter
- Literature : DK book, selected papers

Multivariate State Space Model

The linear Gaussian state space model is defined in three parts:

Observation equation:

$$y_t = c_t + Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, H_t),$$

State equation:

$$\alpha_{t+1} = d_t + T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim \mathcal{NID}(0, Q_t),$$

Initial state distribution $\alpha_1 \sim \mathcal{N}(a_1, P_1)$.

Notice that

- η_t and ε_s independent for all t, s , and independent from α_1 ;
- observation y_t is a **vector**; state vector α_t is unobserved;
- matrices T_t, Z_t, R_t, Q_t, H_t determine structure of model;
- vectors c_t and d_t are also assumed fixed and known.

Multivariate State Space model

So far, we have treated y_t as a scalar and modelled y_t by a univariate time series.

When we need to analyze multiple time series in the **vector** y_t ,

$$y_t = (y_{1t}, \dots, y_{pt})', \quad t = 1, \dots, n,$$

we may want to analyse the p time series simultaneously.

Multivariate time series models become relevant for

- Forecasting many time series
- Cointegration analysis
- Multivariate time series analysis
- Dynamic factor analysis
- Dynamic panel data models

However, multivariate analysis is not straightforward, it requires some work and familiarity with **matrix algebra** !

Multivariate Local Level model

Multivariate Local Level model

Seemingly Unrelated Time Series Equations **SUTSE** model:

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \Sigma_\eta).\end{aligned}$$

- Observations are $p \times 1$ vectors;
- The disturbances ε_t, η_s are independent for all s, t ;
- The p different time series are related through correlations in the disturbances.

Multivariate Local Level model

The multivariate Local Level model is given by

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \Sigma_\eta).\end{aligned}$$

- First difference

$$\Delta y_t = \eta_{t-1} + \Delta \varepsilon_t$$

is stationary;

- Reduced form: Δy_t is vector MA(1) process: similar arguments as in Lecture 1 for the *univariate* Local Level model.

Multivariate Local Level model: variance decomposition

The **multivariate local level model** is given by

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \Sigma_\eta).\end{aligned}$$

Any full-rank (nonsingular) $p \times p$ variance matrix Σ_η can be represented by

$$\Sigma_\eta = A \Sigma_c A',$$

$p \times p$ matrix A is full-rank, Σ_c is positive matrix.

Multivariate Local Level model: variance decomposition

Consider the variance matrix decomposition

$$\Sigma_{\eta} = A \Sigma_c A',$$

To ensure a unique mapping from Σ_{η} to A and Σ_c , and vice-versa, we can impose restrictions on A and Σ_c .

For example, we can consider a Choleski decomposition:

$$A = \text{lower triangular unity matrix, } \begin{bmatrix} 1 & 0 \\ A_{21} & 1 \end{bmatrix},$$

$$\Sigma_c = \text{positive diagonal matrix, } \begin{bmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{22}^2 \end{bmatrix}.$$

Multivariate Local Level model: two representations

Given the **Local Level model**

$$y_t = \mu_t + \varepsilon_t, \quad \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{NID}(0, \Sigma_\eta),$$

with $\Sigma_\eta = A \Sigma_c A'$, we can obtain an **alternative specification**

$$y_t = A \mu_t^c + \varepsilon_t, \quad \mu_{t+1}^c = \mu_t^c + \eta_t^c, \quad \eta_t^c \sim \mathcal{NID}(0, \Sigma_c),$$

with definitions/relations $\mu_t = A \mu_t^c$ and $\eta_t = A \eta_t^c$.

Both models are observationally equivalent, they generate exactly the same properties for y_t , for any matrix A and Σ_c such that

$$\Sigma_\eta = A \Sigma_c A'.$$

Multivariate Local Level Model: two representations

We can regard the multivariate local level model as a system of linear combinations of p independent random walks,

$$y_t = A \mu_t^c + \varepsilon_t, \quad \mu_{t+1}^c = \mu_t^c + \eta_t^c, \quad \eta_t^c \sim \mathcal{NID}(0, \Sigma_c).$$

The i -th equation is

$$y_{i,t} = A_i \mu_t^c + \varepsilon_{i,t} = \sum_{j=1}^p A_{i,j} \mu_{j,t}^c + \varepsilon_{i,t}, \quad i = 1, \dots, p,$$

where $x_{i,t}$ is the i -th element of the vector x_t ($x = \{y, \mu, \varepsilon, \eta\}$), A_i is the i -th row of matrix A , and $A_{i,j}$ is the (i, j) element of A .

Multivariate Local Level Model: Common Levels

For the multivariate Local Level model

$$y_{i,t} = \sum_{j=1}^r A_{i,j} \mu_{j,t}^c + \varepsilon_{i,t}, \quad i = 1, \dots, p,$$

when the sum of r random walks is less than p , the smaller set of level components are shared amongst all p time series.

In other words, the p equations share a smaller number of level components, $r < p$.

We have **Common Levels**.

Common Levels: Bivariate Example

An illustration is the bivariate Local Level model as given by

$$y_{1,t} = \mu_{1,t}^c + \varepsilon_{1,t}, \quad y_{2,t} = A_{2,1} \mu_{1,t}^c + \mu_{2,t}^c + \varepsilon_{2,t},$$

We obtain **Common Level** $\mu_{1,t}^c$ when $\mu_{2,t}^c$ is removed:

$$y_{1,t} = \mu_{1,t}^c + \varepsilon_{1,t}, \quad y_{2,t} = A_{2,1} \mu_{1,t}^c + \varepsilon_{2,t}.$$

Notice that we need to include an intercept in one equation:

$$y_{1,t} = \mu_{1,t}^c + \varepsilon_{1,t}, \quad y_{2,t} = a_2 + A_{2,1} \mu_{1,t}^c + \varepsilon_{2,t},$$

where a_2 is an unknown scalar.

Common Levels

The common local level model remains a system of linear combinations, but of $r < p$ independent random walks,

$$y_t = a + A\mu_t^c + \varepsilon_t, \quad \mu_{t+1}^c = \mu_t^c + \eta_t^c, \quad \eta_t^c \sim \mathcal{NID}(0, \Sigma_c).$$

The i -th equation is

$$y_{i,t} = a_i + \sum_{j=1}^r A_{i,j} \mu_{j,t}^c + \varepsilon_{i,t}, \quad i = 1, \dots, p,$$

where a_i is a constant... to accommodate smaller number in μ_1^c .
We have $p \times r$ matrix A and $r \times r$ matrix Σ_c above, with $r < p$.

Bivariate example : two time series, one common level

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 \\ a_2 \end{bmatrix} + \begin{bmatrix} 1 \\ A_{21} \end{bmatrix} \mu_t^c + \varepsilon_t, \quad \mu_{t+1}^c = \mu_t^c + \eta_t^c.$$

Common Levels

For the common local level model

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \Sigma_\eta),\end{aligned}$$

notice that

- Consider decomposition $\Sigma_\eta = A\Sigma_c A'$ is not unique;
- For example, Choleski decomposition: lower unity triangular matrix A and diagonal matrix Σ_c ;
- Also interesting interpretations can be considered (factor rotations);
- Lower rank matrix Σ_η implies Common Levels.

Cointegration

Common levels imply cointegration :

- Cointegration implies typically that

$$y_t \sim I(1), \quad \text{but} \quad B y_t \sim I(0).$$

with B is $q \times p$ matrix, $1 \leq q < p$, and BB' full rank.

- Assume that y_t is generated by the local level model, we have

$$y_t = \mu_t + \varepsilon_t = a + A \mu_t^c + \varepsilon_t, \quad \mu_{t+1}^c = \mu_t^c + \eta_t^c, \quad \eta_t^c \sim \mathcal{N}(0, \Sigma_\eta^c),$$

and $y_t \sim I(1)$. Notice that $\Sigma_\eta = A \Sigma_\eta^c A'$.

- The cointegration condition implies

$$B y_t = B \mu_t + B \varepsilon_t = B a + B A \mu_t^c + B \varepsilon_t,$$

and to ensure $B y_t \sim I(0)$, we require $BA = 0$.

- Since $\Sigma_\eta = A \Sigma_\eta^c A'$, the condition $BA = 0$ can only be true iff $p \times r$ matrix A has $r < p$, that is iff Σ_η^c has rank $r < p$.
- This last condition implies common levels.

Cointegration

Cointegration and common trends:

- Common trends have consequences for inference and forecasting, see Engle and Granger (1987)
- For example, it leads to dimension reduction of parameter space, we obtain a more parsimonious model.
- The link with cointegration implies that the common local level model (and its extensions) can be generally represented as VAR and VECM models.
- The details can be provided upon request, they are somewhat intricate, ...
- The common local level model can be extended with other (possibly) common components such as slope, seasonal, cycle, ...

Multivariate components

- So far, we have concentrated on the multivariate extension of the local level model.
- Common levels (with levels as $I(1)$ processes) imply cointegration.
- Similar considerations can be applied to other components such as the slope of the trend, seasonal and cycle components and other time-varying features in the multiple time series.
- Harvey and Koopman (1997) review such extensions.
- In particular, they define the similar cycle component.

Common and idiosyncratic factors

Multiple trends can also be decomposed into a one common factor and multiple idiosyncratic factors:

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \Sigma_\varepsilon), \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \Sigma_\eta),\end{aligned}$$

where $\Sigma_\eta = \delta\delta' + D_\eta$ with vector δ and diagonal matrix D_η . This implies that the level can be represented by

$$\mu_t = \delta\mu_t^c + \mu_t^*, \quad \eta_t = \delta\eta_t^c + \eta_t^*$$

with common level (scalar) μ_t^c and "independent" level μ_{it}^* generated by

$$\Delta\mu_{t+1}^c = \eta_t^c \sim \mathcal{NID}(0, 1), \quad \Delta\mu_{t+1}^* = \eta_t^* \sim \mathcal{NID}(0, D_\eta).$$

Multivariate filtering: equation by equation

Multivariate Kalman Filter

Given the state space model $\alpha_{t+1} = T_t\alpha_t + R_t\zeta_t$, $y_t = Z_t\alpha_t + \varepsilon_t$ with known system matrices :

The unobserved state α_t can be “estimated” from the observations with the *Kalman filter*:

$$v_t = y_t - Z_t a_t,$$

$$F_t = Z_t P_t Z_t' + H_t,$$

$$K_t = T_t P_t Z_t' F_t^{-1},$$

$$a_{t+1} = T_t a_t + K_t v_t,$$

$$P_{t+1} = T_t P_t T_t' + R_t Q_t R_t' - K_t F_t K_t',$$

for $t = 1, \dots, n$ and starting with given values for a_1 and P_1 .

- Writing $Y_t = \{y_1, \dots, y_t\}$,

$$a_{t+1} = \mathbb{E}(\alpha_{t+1} | Y_t), \quad P_{t+1} = \text{Var}(\alpha_{t+1} | Y_t).$$

Multivariate Kalman filter

- The Kalman filter is valid for the general multivariate state space model.
- Computationally it is not convenient when p becomes large, very large.
- Each step of the Kalman filter requires the inversion of the $p \times p$ matrix F_t . This is no problem when $p = 1$ (univariate) but when $p > 20$, say, it will slow down the Kalman filter considerably.
- However, we can treat each element in the $p \times 1$ observation vector y_t as a single realisation:

$$y_t = (y_{1t}, y_{2t}, \dots, y_{p,t})'$$

In other words, we can "update" each single element of y_t within the Kalman filter.

- The results below apply to filtering and smoothing.

Disentangle the Multivariate Kalman filter

View multivariate state space model as a system of equations:

$$y_t = Z_t \alpha_t + \varepsilon_t,$$

and assume that $\text{Var}(\varepsilon_t) = H_t$ is diagonal. The single equation is

$$y_{it} = Z_{it} \alpha_t + \varepsilon_{it},$$

where Z_{it} is the i -th row of Z_t , with $i = 1, \dots, p$ or even p_t .

In this case we have the measurement density as

$$p(y_t | \alpha_t) = \prod_{i=1}^{p_t} p(y_{it} | \alpha_t) = \prod_{i=1}^{p_t} p(y_{it} | \alpha_{t,i}),$$

where $\alpha_{t,i} \equiv \alpha_t$, for $i = 1, \dots, p_t$.

Disentangle the Multivariate Kalman filter

Consider standard model: $y_t = Z_t \alpha_t + \varepsilon_t$ and $\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$ where $\text{Var}(\varepsilon_t) = H_t$ is diagonal.

To treat $y_t = (y_{1t}, \dots, y_{p_t,t})'$ element by element, we have introduced the variable

$$\alpha_{t,i} \equiv \alpha_t, \quad i = 1, \dots, p_t.$$

This notation allows us to express the model in recursive (Markovian) form for both indices i and t , that is

$$y_{it} = Z_{it} \alpha_{t,i} + \varepsilon_{it}, \quad \alpha_{t,i+1} = \alpha_{t,i},$$

for $i = 1, \dots, p_t - 1$, BUT for $i = p_t$, we have

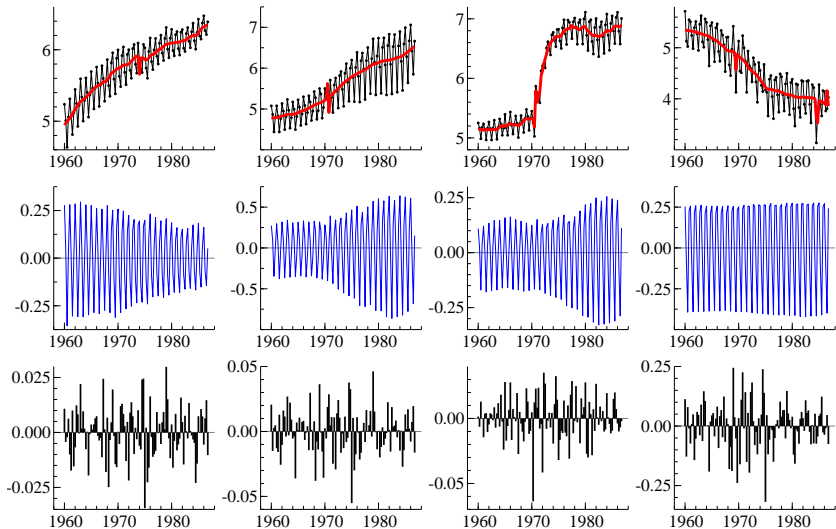
$$\alpha_{t+1,1} = T_t \alpha_{t,p_t} + R_t \eta_t,$$

for $t = 1, \dots, n$.

Univariate treatment of Kalman filter

- Consider standard model: $y_t = Z_t\alpha_t + \varepsilon_t$ and $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$ where $\text{Var}(\varepsilon_t) = H_t$ is diagonal.
- Observation vector $y_t = (y_{1t}, \dots, y_{p_t,t})'$ is treated and we view observation model as a set of p_t separate equations.
- We then have, $y_{it} = Z_{it}\alpha_{t,i} + \varepsilon_{it}$ with $\alpha_{t,i} \equiv \alpha_t$ for $i = 1, \dots, p_t$.
- The associated transition equations become $\alpha_{t,i+1} = \alpha_{t,i}$ for $i = 1, \dots, p_t$ and $\alpha_{t+1,1} = T_t\alpha_{t,p_t} + R_t\eta_t$ for $t = 1, \dots, n$.
- This disentangled model can be treated by the Kalman filter and smoother equations straightforwardly.
- Innovations are now relative to the past and the “previous” observations inside y_{t,p_t} !
- Non-diagonal matrix H_t can be treated by data-transformation or by including ε_t in the state vector α_t .
- More details in DK book §6.4.

Seasonal Adjustment for multiple time series: energy



Seasonal Adjustment for multiple time series: energy

MLEs of variance matrices for Level, Slope, Seasonal and Irregular :

Level disturbance variance/correlation matrix:

| | ofuELI | ofuGASl | oiGASl | ofuCOALl |
|----------|--------------|---------------|-----------|-----------|
| ofuELI | 0.000289 | 0.4111 | 0.5182 | 0.6964 |
| ofuGASl | $7.244e - 5$ | 0.0001075 | -0.5666 | -0.3679 |
| oiGASl | 0.0002409 | -0.0001607 | 0.0007480 | 0.9747 |
| ofuCOALl | 0.0001668 | $-5.373e - 5$ | 0.0003755 | 0.0001985 |

Slope disturbance variance/correlation matrix:

| | ofuELI | ofuGASl | oiGASl | ofuCOALl |
|----------|---------------|---------------|---------------|--------------|
| ofuELI | $1.165e - 6$ | -0.6977 | -0.5429 | 0.4824 |
| ofuGASl | $-2.000e - 6$ | $7.057e - 6$ | 0.9804 | -0.9641 |
| oiGASl | $-5.277e - 6$ | $2.346e - 5$ | $8.113e - 5$ | -0.9975 |
| ofuCOALl | $1.738e - 6$ | $-8.552e - 6$ | $-3.000e - 5$ | $1.115e - 5$ |

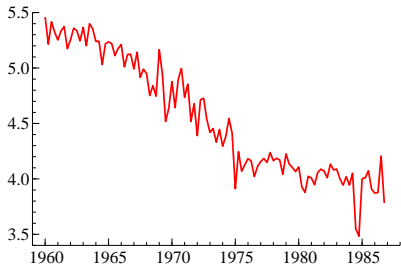
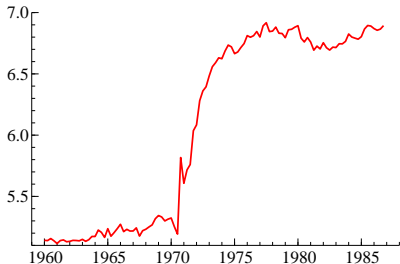
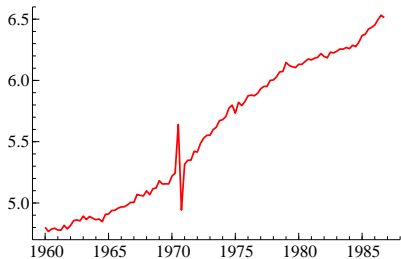
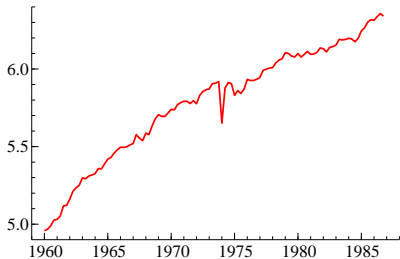
Seasonal disturbance variance/correlation matrix:

| | ofuELI | ofuGASl | oiGASl | ofuCOALl |
|----------|---------------|--------------|--------------|--------------|
| ofuELI | 0.0001246 | -0.05925 | -0.01505 | 0.4935 |
| ofuGASl | $-1.480e - 5$ | 0.0005010 | 0.8069 | 0.8377 |
| oiGASl | $-1.295e - 6$ | 0.0001392 | $5.941e - 5$ | 0.6655 |
| ofuCOALl | $1.390e - 5$ | $4.731e - 5$ | $1.294e - 5$ | $6.367e - 6$ |

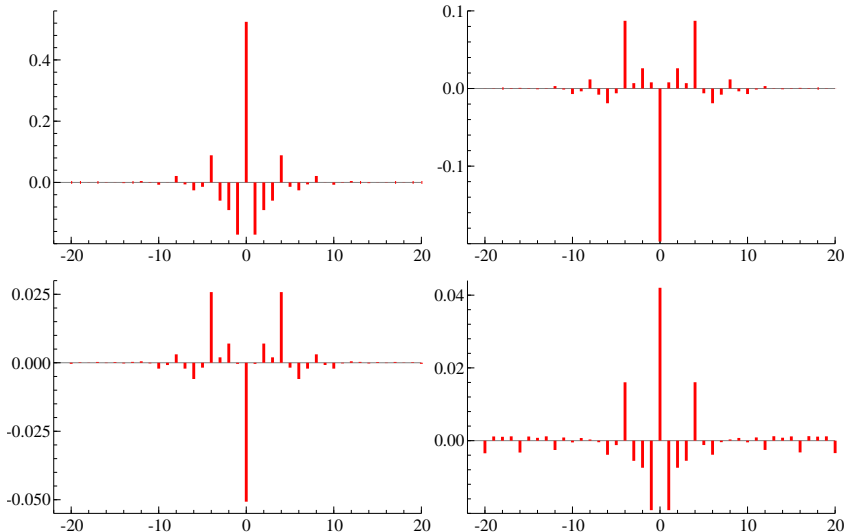
Irregular disturbance variance/correlation matrix:

| | ofuELI | ofuGASl | oiGASl | ofuCOALl |
|----------|-----------|-----------|-----------|----------|
| ofuELI | 0.0003510 | 0.9155 | 0.3332 | 0.4942 |
| ofuGASl | 0.0003780 | 0.0004856 | 0.1956 | 0.7315 |
| oiGASl | 0.0001579 | 0.0001090 | 0.0006401 | 0.4488 |
| ofuCOALl | 0.0009763 | 0.001700 | 0.001197 | 0.01112 |

Multivariate Seasonal Adjustment : energy series



Multivariate Seasonal Component : weight functions



Dynamic factor state space models

Dynamic factor state space models

Program :

- Introduction
- Dynamic factor models
- Collapsed Kalman filter method
- Forecasting and Principal Components
- Collapsed dynamic factor models
- Parameter Estimation in Dynamic Factor Models
- Literature : DK book, selected papers

Multivariate State Space Model

The linear Gaussian state space model is defined in three parts:

Observation equation:

$$y_t = c_t + Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, H_t),$$

State equation:

$$\alpha_{t+1} = d_t + T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim \mathcal{NID}(0, Q_t),$$

Initial state distribution $\alpha_1 \sim \mathcal{N}(a_1, P_1)$.

Notice that

- η_t and ε_s independent for all t, s , and independent from α_1 ;
- observation y_t is a **vector**; state vector α_t is unobserved;
- matrices T_t, Z_t, R_t, Q_t, H_t determine structure of model;
- vectors c_t and d_t are also assumed fixed and known.

Dynamic factor models

Consider a basic example of the dynamic factor model, for $p \times 1$ observation vector y_t and $r \times 1$ latent factor vector f_t , as given by

$$y_t = \Lambda f_t + \varepsilon_t, \quad f_{t+1} = \Phi f_t + \zeta_t, \quad t = 1, \dots, n,$$

where $\varepsilon_t \sim \mathcal{NID}(0, \Sigma_\varepsilon)$ and $\zeta_t \sim \mathcal{NID}(0, \Sigma_\zeta)$.

For identification purposes, we have

$$\text{vec}(\Sigma_\zeta) = (I - \Phi \otimes \Phi) \text{vec}(I).$$

In other words, factors f_t are standardized, $f_t \sim \mathcal{N}(0, I)$.

A basic example

For $p \times 1$ data vector y_t and r factors in f_t , the basic DFM is

$$y_t = \Lambda f_t + \varepsilon_t, \quad f_{t+1} = \Phi f_t + \zeta_t, \quad t = 1, \dots, n.$$

Cross-section dimension p is typically high and time series length n is moderate.

- We are possibly interested in $p \gg n$.
- Estimation concentrates on Λ , Σ_ε and Φ .
- However, first we concentrate on
 - signal extraction of f_t ,
 - likelihood evaluation,for **given** values of Λ , Σ_ε and Φ .

State space formulation

The dynamic factor model for $p \times 1$ observation vector y_t and with $r \times 1$ latent factor vector f_t

$$y_t = \Lambda f_t + \varepsilon_t, \quad f_{t+1} = \Phi f_t + \zeta_t, \quad t = 1, \dots, n,$$

and disturbances

$$\varepsilon_t \sim \mathcal{NID}(0, \Sigma_\varepsilon) \quad \zeta_t \sim \mathcal{NID}(0, \Sigma_\zeta),$$

can be obviously represented in the familiar (partially time-invariant) state space model

$$y_t = Z\alpha_t + \varepsilon_t, \quad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t,$$

with

$$\varepsilon_t \sim \mathcal{NID}(0, H) \quad \eta_t \sim \mathcal{NID}(0, Q_t).$$

Let's adopt the state space representation.

Collapsed Kalman filter

Signal extraction

Consider dynamic factor model

$$y_t = Z\alpha_t + \varepsilon_t, \quad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t,$$

with high-dimensional y_t and low-dimensional α_t .

Likelihood evaluation can be based on prediction error decomposition

$$\ell = p(y_1) \prod_{t=2}^n p(y_t | y_1, \dots, y_{t-1}),$$

and can be routinely computed by the Kalman filter. Evaluation of

$$\tilde{\alpha}_t = E(\alpha_t | y_1, \dots, y_s), \quad \text{Var}(\alpha_t | y_1, \dots, y_s), \quad s = t - 1, \dots, n,$$

for $t = 1, \dots, n$ is then carried out by Kalman filter and related methods.

Kalman filter methods often dismissed as p becomes very large : (

Transformation by regression

However, huge computational gains can be obtained as follows :)

Model

$$y_t = Z\alpha_t + \varepsilon_t, \quad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t,$$

Apply GLS regression lemma for every t :

$$\hat{\alpha}_t = Py_t, \quad \text{where } P = (Z'H^{-1}Z)^{-1}Z'H^{-1}.$$

Then, transform model for y_t to a model for $\hat{\alpha}_t$, that is

$$\hat{\alpha}_t = \alpha_t + e_t,$$

with $e_t = P\varepsilon_t \sim \mathcal{NID}\{0, (Z'H^{-1}Z)^{-1}\}$. It can be shown that

$$\tilde{\alpha}_t = E(\alpha_t | y_1, \dots, y_s) = E(\alpha_t | \hat{\alpha}_1, \dots, \hat{\alpha}_s), \quad t, s = 1, \dots, n.$$

It implies that observation equation dimension N reduces to r .

Two-step method

Model

$$y_t = Z\alpha_t + \varepsilon_t, \quad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t,$$

for known system matrices.

Signal extraction for α_t is carried out in two steps:

1. Cross-section step (GLS)

$$\hat{\alpha}_t = (Z'H^{-1}Z)^{-1}Z'H^{-1}y_t.$$

2. Time series step: use Kalman filter methods to evaluate

$\tilde{\alpha}_t = E(\alpha_t|y_1, \dots, y_s)$ based on low-dimensional model

$$\hat{\alpha}_t = \alpha_t + u_t, \quad u_t \sim \mathcal{NID}(0, C), \quad C = (Z'H^{-1}Z)^{-1}.$$

It turns out that **all** inference can be based on this model for $\hat{\alpha}_t$, including the evaluation of the likelihood function.

For formal derivations and further details, see Jungbacker and Koopman (Ect J, 2015).

Likelihood evaluation

The Gaussian loglikelihood function can be expressed as

$$\ell(y; \psi) = c + \ell(\hat{\alpha}; \psi) - \frac{n}{2} \log \frac{|H|}{|C|} - \frac{1}{2} \sum_{t=1}^n e_t' H^{-1} e_t,$$

where $\hat{\alpha} = (\hat{\alpha}'_1, \dots, \hat{\alpha}'_n)'$, $e_t = y_t - Z\hat{\alpha}_t$ and c is some constant, does not depend on y and ψ .

We evaluate $\ell(\hat{\alpha}; \psi)$ by the Kalman filter applied to $\hat{\alpha}_t = \alpha_t + u_t$.

It follows that for the evaluation of the loglikelihood, large dimensional Kalman filter is not required.

Matrix H is often assumed diagonal or has other strong structure (blocks, bands, spatial). Term $|C|$ is delivered by KFS.

This collapsing approach leads to computationally feasible methods for dynamic factor analysis.

Maximum likelihood estimation

All unknown elements in the state space system matrices are collected in the parameter vector ψ .

Parameter estimation by the method of maximum likelihood (ML) can be carried out in a standard fashion.

However, typically, the parameter vector has a huge dimension (1000 parameters are not an exception). Computer technology provides means to carry out ML in a feasible manner.

Also approximate methods are introduced.

This is another course !!!