

Lecture: State-Space Form

222061-1617: Time Series Econometrics

Spring 2019

Jacek Suda

Outline

Outline:

- 1 Prediction Error Decomposition
- 2 State-Space Form
- 3 Kalman Filter

PREDICTION ERROR DECOMPOSITION

Likelihood

- Consider a covariance-stationary $\{Y_t\}$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \tilde{y}_T \sim N(\mu_{T \times 1}, \Omega_{T \times T}),$$

- Since it is covariance stationary process, each Y_t has the same mean and variance, $\omega_{11} = \sigma^2 = \omega_{22} = \omega_{TT}$.

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} & & \omega_{1T} \\ \omega_{21} & \omega_{22} & & \vdots \\ \vdots & & \ddots & \\ \omega_{T1} & \dots & & \omega_{TT} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & & \vdots \\ \vdots & & \ddots & \\ \gamma_{T-1} & \dots & & \gamma_0 \end{bmatrix}, \text{ as } \begin{cases} \omega_{jj} = \gamma_0, \\ \omega_{ij} = \gamma_{i-j}, \\ i > j. \end{cases}$$

- The likelihood function:

$$L(\tilde{\theta} | \tilde{y}_T) = (2\pi)^{-\frac{T}{2}} \det(\Omega)^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{y}_T - \mu)' \Omega^{-1} (\tilde{y}_T - \mu)}.$$

Factorization

- For large T , Ω might be large and difficult to invert.
- Since Ω is positive definite symmetric matrix, there exists a unique, triangular factorization of Ω ,

$$\Omega = AfA'$$

where

$$f_{T \times T} = \begin{bmatrix} f_1 & 0 & \cdots & 0 \\ 0 & f_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & f_T \end{bmatrix}, \quad f_t > 0 \forall t \quad \text{diagonal matrix}$$

$$A_{T \times T} = \begin{bmatrix} 1 & & & 0 \\ a_{21} & 1 & & \\ \vdots & & \ddots & \\ a_{T1} & a_{T2} & \cdots & 1 \end{bmatrix}, \quad \text{lower triangular matrix}$$

Prediction errors

- Define $\eta = A^{-1}(\tilde{y}_T - \mu)$,

$$E(\eta) = 0, \quad \text{var}(\eta) = \text{var}(A^{-1}(\tilde{y}_T - \mu)) = A^{-1}\Omega[A']^{-1} = A^{-1}AfA'[A']^{-1} = f.$$

- Since f is diagonal, η is a series of random variables that are uncorrelated with each other, i.e. $E(\eta_t\eta_\tau) = 0$, $\forall t \neq \tau$.
- Since $A\eta = (\tilde{y}_T - \mu)$ and A is a lower-triangular matrix with 1s along the principal diagonal,

$$\eta_1 = y_1 - \mu$$

$$\eta_2 = y_2 - \mu - a_{11}^*\eta_1$$

$$\eta_3 = y_3 - \mu - a_{21}^*\eta_1 - a_{22}^*\eta_2$$

$$\vdots$$

$$\eta_T = y_T - \mu - \sum_{i=1}^{T-1} a_{T-1,i}^*\eta_i$$

where $a_{i,j}^* = A_{i+1,j}$

Prediction errors

- Since $E(\eta_t \eta_\tau) = 0$,

$$E(\eta_2 \eta_1) = E((y_2 - \mu - a_{11}^*(y_1 - \mu))(y_1 - \mu)) = 0$$

so the a_{11}^* is the coefficient of the linear projection of $(y_2 - \mu)$ on $(y_1 - \mu)$ and $\eta_2 = y_2 - y_{2|1}$ is a prediction error .

- Similarly, $E(\eta_3 \eta_2) = E(\eta_3 \eta_1) = 0$ imply that a_{21}^* and a_{22}^* are linear projection coefficients of $(y_3 - \mu)$ on $(y_2 - \mu)$ and $(y_1 - \mu)$, with $\eta_3 = y_3 - y_{3|2}$
- Therefore, η_t is t^{th} element of $\eta_{T \times 1} =$ prediction error $y_t - \hat{y}_{t|t-1}$.

Likelihood

- The likelihood function can be rewritten as:

$$L(\tilde{\theta}|\tilde{y}_T) = (2\pi)^{-\frac{T}{2}} \det(AfA')^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{y}_T - \mu)'(AfA')^{-1}(\tilde{y}_T - \mu)}$$

- As A is lower triangular with 1s along the principal diagonal, $\det(A) = 1$ and

$$\det(AfA) = \det(A) \cdot \det(f) \cdot \det(A') = \det(f).$$

- Then,

$$\begin{aligned} L(\tilde{\theta}|\tilde{y}_T) &= (2\pi)^{-\frac{T}{2}} \det(f^{-1})^{-\frac{1}{2}} e^{-\frac{1}{2}\eta'(f^{-1})^{-1}\eta} \\ &= \prod_{t=1}^T \left(\frac{1}{\sqrt{2\pi f_t}} e^{-\frac{1}{2} \frac{\eta_t^2}{f_t}} \right), \end{aligned}$$

where η_t is t^{th} element of $\eta_{T \times 1} =$ prediction error $y_t - \hat{y}_{t|t-1}$,

$$\hat{y}_{t|t-1} = \sum_{i=1}^{t-1} a_{t,i}^* y_i, \quad i = 2, 3, \dots, T.$$

Kalman Filter

Note: Given $y_t \sim N(\mu, \Omega)$,

$$\eta_t | \Omega_{t-1} \sim N(0, f_t),$$

where f_t is an (t, t) diagonal element of f matrix,

$$\ln L = -\frac{1}{2} \sum_{t=1}^T \ln(2\pi f_t) - \frac{1}{2} \sum_{t=1}^T \frac{\eta_t^2}{f_t},$$

since $\eta_t \sim N$ and independent of each other.

- The Kalman filter recursively calculates linear projection of y_t on past information Ω_{t-1} for any model that can be cast in state-space form.
- *Kalman filter: for any structure it solves for linear prediction.*

STATE-SPACE

Measurement (Observation) Equation

General form that encompasses a wide variety of models.

1 Measurement (Observation) Equation

- Represent the static relationship between observed variables (data) and unobserved state variables.

$$y_t = H_t \beta_t + A z_t + e_t,$$

where y_t denotes observed data, β_t is a state vector that captures the dynamics, z_t is exogenous, observed variables *for example, lagged values of y_t but also other data*, and e_t is an error term,

$$e_t \sim N(0, R).$$

The existence of the state vector makes this representation not a simple linear model.

Transition (State) Equation

② Transition (State) Equation

- *Captures the dynamics in the system, causes the system to go on and on.*

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + v_t,$$

where $\tilde{\mu}$ is a vector of constants, F is the transition matrix, and v_t is an error vector,

$$v_t \sim N(0, Q).$$

- *Like AR(1) but in vector/matrix form.*

Transition (State) Equation

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + v_t,$$

- The state vector has an AR(1) kind of representation.
- Describes evolution of state vector.
- These state vectors can be unobservable.
- Transition equation can be used to get information about the unobservable, conditioning on data which is observable (Bayesian).

Error terms

Error terms:

$$e_t \sim N(0, R), \quad v_t \sim N(0, Q),$$

where R, Q are var-cov matrices and

$$E[e_t v_\tau'] = 0, \quad \forall t, \tau$$

- Restrictive assumption
- The model can be represented in a way that is not very restrictive.
- Even with $E[e_t v_\tau'] \neq 0$ we can estimate the model with (modified) Kalman Filter but it becomes more complicated.
- The normality assumption might not be always good...
...but it allows to use MLE.

Examples: AR(2)

Consider an AR(2) process

$$\begin{aligned}y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \\ \varepsilon_t &= WN(0, \sigma^2).\end{aligned}$$

State equation

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} c \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

Observation equation

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}.$$

Examples: AR(2) again

Consider again an AR(2) process

$$\begin{aligned}y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \\ \varepsilon_t &= WN(0, \sigma^2).\end{aligned}$$

State equation

$$\beta_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \phi_2 y_{t-2} \end{bmatrix} + \begin{bmatrix} c \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

Observation equation

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix}.$$

Examples: MA(1)

- Consider an MA(1) process

$$\begin{aligned}y_t &= \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \\ \varepsilon_t &= WN(0, \sigma^2).\end{aligned}$$

- Define

$$\beta_t = \begin{bmatrix} y_t - \mu \\ \theta\varepsilon_t \end{bmatrix}$$

- Then

State equation

$$\beta_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \beta_{t-1} + \begin{bmatrix} 1 \\ \theta \end{bmatrix} \varepsilon_t$$

Observation equation

$$y_t = \mu + \begin{bmatrix} 1 & 0 \end{bmatrix} \beta_t.$$

Examples: ARMA(1,1)

ARMA(1,1):

Set $\mu = 0$,

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma^2).$$

- *There might be more than one way to represent a model in a state-space form.*
- *There might be differences in efficiency between different ways.*

Examples: ARMA(1,1)

State equation:

- The general form

$$\beta_t = F\beta_{t-1} + v_t.$$

- Put

$$\beta_t = \begin{bmatrix} y_t \\ \varepsilon_t \end{bmatrix} \Rightarrow \beta_{t-1} = \begin{bmatrix} y_{t-1} \\ \varepsilon_{t-1} \end{bmatrix}$$

- Put $y_t = \phi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$ in a matrix notation:

$$\begin{bmatrix} y_t \\ \varepsilon_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \phi & \theta \\ 0 & 0 \\ F \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \varepsilon_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \varepsilon_t \\ v_t \end{bmatrix},$$

and $v_t \sim N(0, Q)$, $Q = \begin{bmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{bmatrix}$.

y_t – observable, ε_t – unobservable, forecast error

Examples: ARMA(1,1)

Observation equations:

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \varepsilon_t \\ \beta_t \end{bmatrix}$$
$$y_t \quad \quad \quad H$$

no exogenous variables: $A = 0$, also $R = 0$.

$y_t = H\beta_t$ for this case (ARMA(1,1)).

The parameters ϕ, θ, σ^2 are captured in F, Q matrices. The Kalman Filter will estimate them.

- *For KF what goes in β_t doesn't matter.*
- *Only parameters F, Q, H, R will matter.*
- *The state vector is now defined by F, Q, H , and the observations.*

ARMA(1,1): Alternative Representation

A more “elegant” (i.e. easier for computation) representation.

Lag notation (alternative representation for ARMA(1,1))

$$\begin{aligned}(1 - \phi L)y_t &= (1 + \theta L)\varepsilon_t \\ y_t &= (1 - \phi L)^{-1}(1 + \theta L)\varepsilon_t \\ y_t &= (1 + \theta L)(1 - \phi L)^{-1}\varepsilon_t.\end{aligned}$$

Define $x_t = (1 - \phi L)^{-1}\varepsilon_t$

$$\begin{aligned}(1 - \phi L)x_t &= \varepsilon_t, & (x_t \text{ is AR(1), not observed}) \\ x_t - \phi x_{t-1} &= \varepsilon_t\end{aligned}$$

Then,

$$\begin{aligned}y_t &= (1 + \theta L)x_t \\ y_t &= x_t + \theta x_{t-1}.\end{aligned}$$

So y_t is a linear combination of 2 unobservable AR(1) processes, x_t and x_{t-1} .

ARMA(1,1): State-Space

Observation equation (*all randomness in the state equation*)

$$y_t = H\beta_t,$$

where

$$y_t = \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}$$

- *Inside H there are parameters to be estimated.*
- *$A = 0$, no exogenous, $R = 0$ as the observable equation is just the identity (no randomness of e_t).*

ARMA(1,1): State-Space

State equation

$$\begin{bmatrix} x_t \\ x_{t-1} \\ \beta_t \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \\ & F \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ v_t \end{bmatrix},$$

so

$$v_t \sim N(0, Q), \quad Q = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

- So ϕ is in F , θ in H , and σ^2 in Q .

Given F, Q, H, A, R and data (y_t 's), use Kalman Filter to find prediction error decomposition of joint likelihood for $\tilde{y}_T = (y_1, \dots, y_T)$, given by $L(\theta, \phi, \sigma^2 | \tilde{y}_T)$. (exact likelihood)

KALMAN FILTER

Kalman Filter

Kalman filter:

- *purpose: to make inference about unobservable given the observable,*
- *application: signal extraction in engineering,*
- *economics: don't know the parameters F , Q , H and want to estimate them.*

State-space form

ME: Measurement (Observation) equation:

$$y_t = H\beta_t + e_t, \quad e_t \sim N(0, R)$$

SE: Transition (State) equation:

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + v_t, \quad v_t \sim N(0, Q),$$

$$E[e_t v_\tau'] = 0.$$

Mean of β

- ① β_t is a random variable
 - it might be unobservable and no data for it,
 - it is normal random variable as it is sum of normal variables, $v_t \sim N$.

Conditional mean

$$\begin{aligned} \beta_t | \Omega_{t-1} &\sim N(E[\beta_t | \Omega_{t-1}], \text{var}(\beta_t | \Omega_{t-1})) \\ E[\beta_t | \Omega_{t-1}] &= \beta_{t|t-1}, \quad \text{conditional expectations.} \end{aligned}$$

- We may not know what β 's are.
- If we have information about its distribution, we can calculate mean, variance, etc.
- β_{t-1} may be not observable: take expectations of it

$$\begin{aligned} E[\beta_t | \Omega_{t-1}] \equiv \beta_{t|t-1} &= \tilde{\mu} + FE[\beta_{t-1} | \Omega_{t-1}] + 0 \\ \beta_{t|t-1} &= \tilde{\mu} + F\beta_{t-1|t-1}, \end{aligned}$$

- In AR(1): $E[y_t] = \mu + \phi E[y_{t-1}]$, last term is observable.

Variance of β

Conditional variance

$$\text{Var}(\beta_t | \Omega_{t-1}) \equiv P_{t|t-1} = E[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})'].$$

Recall

$$\text{var}(ax) = a^2 \text{var}(x), \quad a - \text{scalar}, x - \text{random vector}.$$

Two sources of randomness (variation) for β_t :

- ① v_t is a random variable,
- ② β_{t-1} is also random so there might be difference between β_{t-1} and $\beta_{t|t-1}$, *there may not be equal to each other.*

$$P_{t|t-1} = F P_{t-1|t-1} F' + Q,$$

where $P_{t|t-1}$, uncertainty about β_t equals sum of uncertainty about β_{t-1} , $P_{t-1|t-1}$, and uncertainty about v_t .

Note: $\text{cov}(\beta_{t-1}, v_t) = 0$.

y_t

- ② y_t is a random variable.
 - Now, we have data on y_t .
 - We have some joint density of y_t, β_t and some prior.
 - Using data we get posterior of β_t .

- *We want to make inference for β_t which we don't observe.*
- *We see y_t which is related to β_t .*
- *We make inferences on β_t by observing joint density (distribution) of y_t and β_t (Bayesian view).*

Distribution of y_t

Distribution of y_t given state-space

$$y_t | \Omega_{t-1} \sim N(E[y_t | \Omega_{t-1}], \text{var}(y_t | \Omega_{t-1})),$$

- Conditional mean

$$E[y_t | \Omega_{t-1}] \equiv y_{t|t-1} = H\beta_{t|t-1} + 0$$

- Conditional variance

$$\text{var}(y_t | \Omega_{t-1}) \equiv f_{t|t-1} = HP_{t|t-1}H' + R,$$

since we don't know β_t .

- **Note:** $\text{cov}(H\beta_t, e_t) = 0$ because $E[v_t e_t] = 0$.

If $E[v_t e_t] \neq 0$ we will add another term in the $\text{var}(y_t | \Omega_{t-1})$ capturing that.

Joint Distribution

- Covariance between β_t and y_t :

$$\text{cov}(y_t, \beta_t | \Omega_{t-1}) = P_{t|t-1} H',$$

$$\text{as } \text{cov}(H\beta_t + e_t, \beta_t) = \text{cov}(H\beta_t, \beta_t) + \text{cov}(e_t, \beta_t) = \text{cov}(\beta_t, \beta_t) H' + 0.$$

Then, the joint distribution for y_t and β_t is joint normal:

$$\begin{pmatrix} \beta_t \\ y_t \end{pmatrix} \bigg| \Omega_{t-1} \sim N \left(\begin{bmatrix} \beta_{t|t-1} \\ H\beta_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1} H' \\ P_{t|t-1} H' & f_{t|t-1} \end{bmatrix} \right).$$

Kalman Filter

Two steps of Kalman Filter :

- (a) Prediction,
- (b) Given y_t updating inference on β_t .

Definition

Given $\beta_{0|0}, P_{0|0}$, Kalman Filter solves the following six equations for $i = 1, \dots, T$

Prediction of y_t, β_t

$$(1) \quad \beta_{t|t-1} = \tilde{\mu} + F\beta_{t-1|t-1},$$

$$(2) \quad P_{t|t-1} = F P_{t-1|t-1} F' + Q,$$

Forecast error:

$$(3) \quad \eta_{t|t-1} \equiv y_t - y_{t|t-1} = y_t - H\beta_{t|t-1},$$

Variance of forecast error:

$$(4) \quad f_{t|t-1} = H P_{t|t-1} H' + R$$

Updating of y_t, β_t

$$(5) \quad \beta_{t|t} = \beta_{t|t-1} + \kappa_t \eta_{t|t-1},$$

$$(6) \quad P_{t|t} = P_{t|t-1} - \kappa_t H P_{t|t-1},$$

$$\kappa_t \equiv P_{t|t-1} H' f_{t|t-1}^{-1} \quad \text{“Kalman gain”}.$$

Kalman Filter

- $\beta_{0|0}, P_{0|0}$, are equal to unconditional mean and variance, and reflect prior beliefs.
- If the state space model is covariance stationary,

$$\begin{aligned}
 E[\beta] = \beta_{0|0} &= (\mathbb{I} - \mu)^{-1} \tilde{\mu} \\
 \text{var}(\beta) = P_{0|0} &= FP_{0|0}F' + Q \\
 \text{vec}(P_{0|0}) &= \text{vec}(FP_{0|0}F') + \text{vec}(Q) \\
 \text{vec}(P_{0|0}) &= (F \otimes F)\text{vec}(P_{0|0}) + \text{vec}(Q) \\
 \text{vec}(P_{0|0}) &= (\mathbb{I} - F \otimes F)^{-1}\text{vec}(Q).
 \end{aligned}$$

since $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$.

- Equation (5) is a linear combination of previous guess and forecast error.

$$(5) \quad \beta_{t|t} = \beta_{t|t-1} + \kappa_t \eta_{t|t-1},$$

$$(6) \quad P_{t|t} = P_{t|t-1} - \kappa_t H P_{t|t-1},$$

$$\kappa_t \equiv P_{t|t-1} H' f_{t|t-1}^{-1} \quad \text{“Kalman gain”}.$$

Kalman Gain

- The stronger the covariance between y_t and β_t , the more we will update when we see high forecast error.
- If the relationship is weaker, we don't put much weight as probably it is not driven by β_t .
- The weight depends on the variance of forecast error: if f^{-1} big, put high weight on that observations.
- Once we have $\eta_{t|t-1}, f_{t|t-1}$, we can do MLE after constructing the joint likelihood of prediction error decomposition.
 - *The Kalman gain depends on the relationship between y_t and β_t since $P_{t|t-1}H' = cov(\beta_t, y_t)$ and $f_{t|t-1}^{-1}$ is the precision of the forecast error.*
 - *The bigger the variance of forecast error the smaller the Kalman gain and less weight put to updating.*
 - *Equation (6) measures conditional variance.*
 - *Since we observe y_t the uncertainty declines.*