IMF INSTITUTE
FOR CAPACITY DEVELOPMENT

State Space Representation of Dynamic Models and the Kalman Filter

Joint Vienna Institute/ IMF ICD
Macro-econometric Forecasting and Analysis
JV16.12, L04, Vienna, Austria May 18, 2016

Presenter
Charis Christofides

Introduction and Motivation

- The dynamics of a time series can be influenced by "unobservable" (sometimes called "latent") variables.
- Examples include:
  - Potential output or the NAIRU
  - A common business-cycle
  - The equilibrium real interest rate
  - Yield curve factors: "level", "slope", "curvature"
- Classical regression analysis is not feasible when unobservable variables are present:
  - If the variables are estimated first and then used for estimation, the estimates are typically biased and inconsistent.

Introduction and Motivation (continued)

- State space representation is a way to describe the law of motion of these latent variables and their linkage with known observations.
- The Kalman filter is a computational algorithm that uses conditional means and expectations to obtain exact (from a statistical point of view) finite sample linear predictions of unobserved latent variables, given observed variables.
- Maximum Likelihood Estimation (MLE) and Bayesian methods are often used to estimate such models and draw statistical inferences.

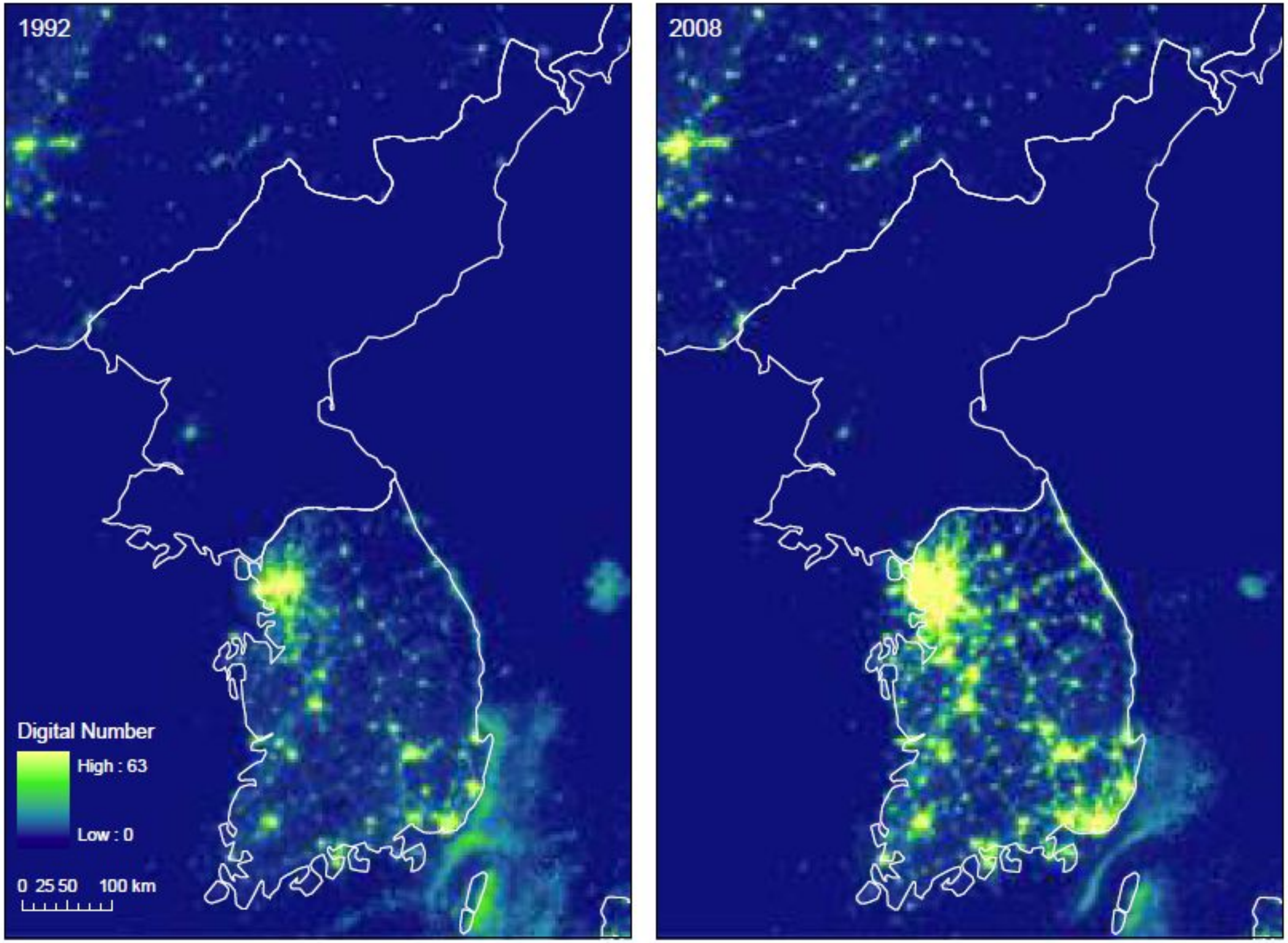
Common Usage of These Techniques

- Macroeconomics, finance, time series models
- Autopilot, radar tracking
- Orbit tracking, satellite navigation (historically important)
- Speech, picture enhancement

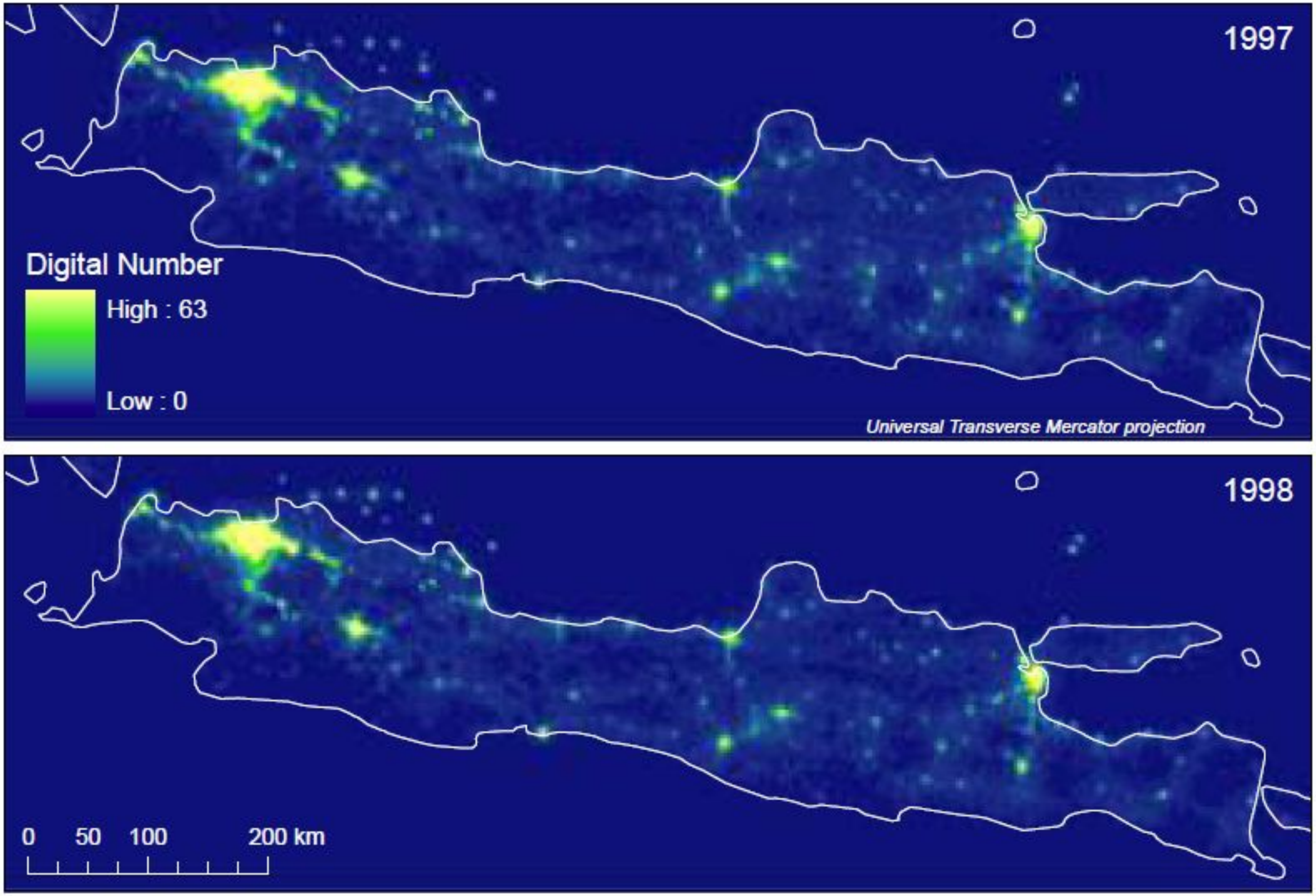
Another example

- Use nightlight data and the Kalman filter to adjust official GDP growth statistics.
- The idea is that economic activity is closely related to nightlight data.
- “Measuring Economic Growth from Outer Space” by Henderson, Storeygard, and Weil AER(2012)

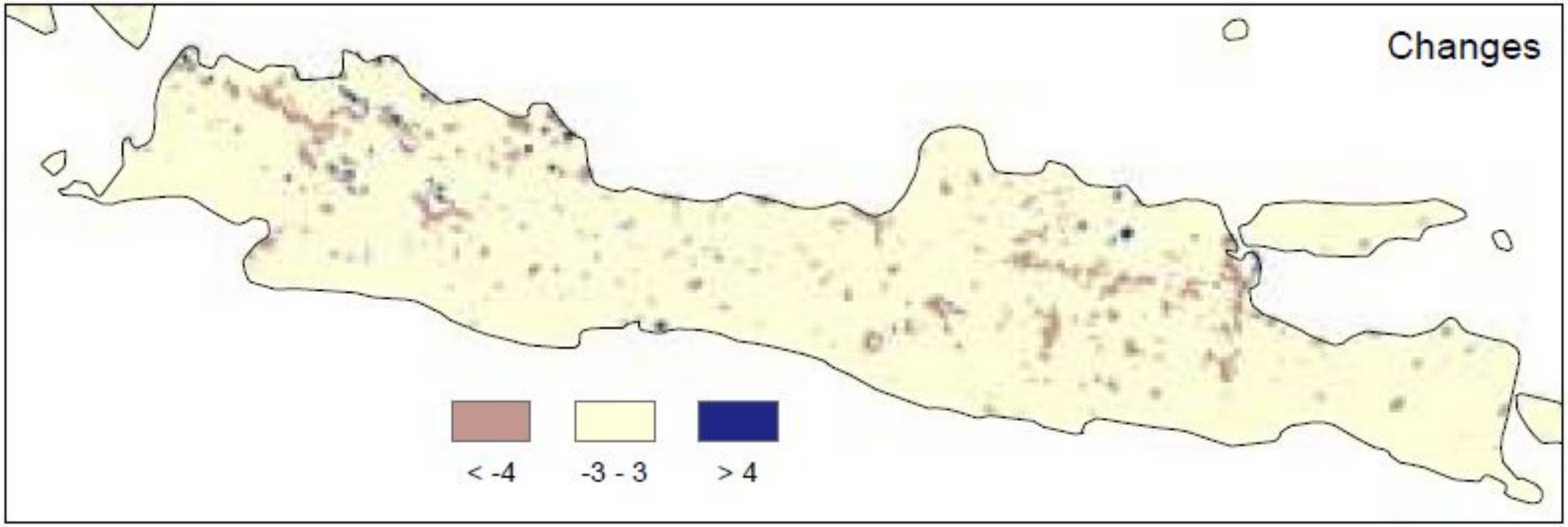
Measuring Long-Term Growth

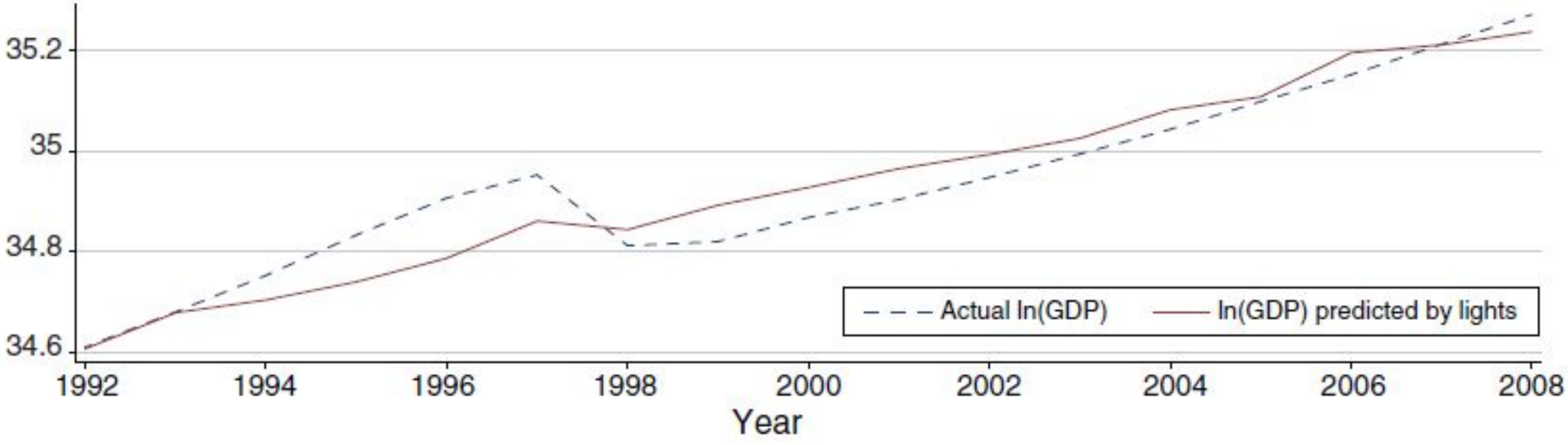


Measuring Short-Term Growth



Measuring Short-Term Growth





Content Outline: Lecture Segments

- State Space Representation
- The Kalman Filter
- Maximum Likelihood Estimation and Kalman Smoothing

Content Outline: Workshops

- Workshops
  - Estimation of equilibrium real interest rate, trend growth rate, and potential output level: Laubach and Williams (ReStat 2003);
  - Estimation of a term structure model of latent factors: Diebold and Li (J. Econometrics 2006);
  - Estimation of output gap (various country examples).

State Space Representation

Basic Setup

Let y_t be an (or a vector) ***observable*** variable(s) at time *t. E.g.,*

- return on asset j
- nominal interest for period from t to $t+j$
- GDP growth

Let x_t be a set of exogenous (pre-determined) variables. E.g.,

- a constant and/or time trend
- the discount rate of the Central Bank
- demand from trading partners

Let s_t be one or a vector of (possibly) ***unobserved*** variable/s: this is the so-called ***state variable***

- Observable variables are assumed to depend on the state variables

Basic Setup

The state-space representation of the dynamics of y_t is given by :

$$s_t = \phi \cdot s_{t-1} + u_t$$ State equation

$$y_t = \alpha \cdot x_t + \beta \cdot s_t + \varepsilon_t$$ Observation equation

We assume that:

- The two equations above represent the true data-generating process for y_t
- All parameters of the process are known
  - Later we will relax this assumption when we discuss estimation
- The unknown (unobserved) variables are s_t, u_t, ε_t for all t, with the last two representing error processes

Basic Setup

The state-space representation of the dynamics of y_t is given by :

$$s_t = \phi \cdot s_{t-1} + u_t \qquad \text{State equation}$$

$$y_t = \alpha \cdot x_t + \beta \cdot s_t + \varepsilon_t \qquad \text{Observation equation}$$

with

ϕ either a constant, or a matrix (if s_t is a vector)

α either a constant, or a matrix (if x_t is a vector)

β either a constant, or a matrix (if s_t is a vector)

The coefficients in β are sometimes called the “loadings”.

Basic Setup

The error terms in the two equations are such that:

$$s_t = \phi \cdot s_{t-1} + u_t$$ State equation

$$y_t = \alpha \cdot x_t + \beta \cdot s_t + \varepsilon_t$$ Observation equation

$E[u_t] = 0$ and $E[\varepsilon_t] = 0$ for every t

$$E[u_t u_{t-j}] = \begin{cases} \mathbf{\Omega} \text{ for every } t \text{ and } j = t \\ 0 \text{ for every } t \text{ and } j \neq t \end{cases}$$

$$\mathbf{\Omega} \text{ is } \begin{cases} \text{a var-cov matrix, if } u_t \text{ is a vector} \\ \text{a variance if } u_t \text{ is one variable} \end{cases}$$

$$E[\varepsilon_t \varepsilon_{t-j}] = \begin{cases} \mathbf{R} \text{ for every } t \text{ and } j = t \\ 0 \text{ for every } t \text{ and } j \neq t \end{cases}$$

$$\mathbf{R} \text{ is } \begin{cases} \text{a var-cov matrix, if } \varepsilon_t \text{ is a vector} \\ \text{a variance if } \varepsilon_t \text{ is one variable} \end{cases}$$

$$E[u_t \varepsilon_{t-j}] = 0 \text{ for every } t \text{ and } j$$

Basic Setup

The error terms in the two equations are such that:

$$s_t = \phi \cdot s_{t-1} + u_t \qquad \text{State equation}$$

$$y_t = \alpha \cdot x_t + \beta \cdot s_t + \varepsilon_t \qquad \text{Observation equation}$$

- What if you know that u_t are serially correlated:
  - $u_t = \rho \cdot u_{t-1} + \eta_t$ and $E[\eta_t] = 0$, $E[\eta_t \eta_j] = \begin{cases} \Sigma, if\ t = j \\ 0, if\ t \neq j \end{cases}$
  - Then $E[u_t u_{t-1}] = \rho \neq 0$ so one of the assumptions is violated!
  - What to do? Can you still apply the model?

The State Space Representation: Examples

<u>*Example #1*</u>: simple version of the CAPM

- s_t one variable, return on all invested wealth
- y_t one variable, return on an asset
- Φ, α, and β constants
- Ω and R constants

$$s_t = \phi \cdot s_{t-1} + u_t$$ State equation

$$y_t = \alpha + \beta \cdot s_t + \varepsilon_t$$ Observation equation

The State Space Representation: Examples

<u>Example #2</u>: growth and real business cycle (small open economy with a large export sector)

- s_t one variable, business cycle
- Y_t vector, GDP growth, unemployment, retail sales
- x_t one variable, demand growth of trading partner
- Φ, and Ω constants
- α, and β vectors
- R matrix

$$s_t = \phi \cdot s_{t-1} + u_t$$

State equation

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{pmatrix} \cdot x_t + \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix} \cdot s_t + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}$$

Observation equation

The State Space Representation: Examples

<u>Example #3</u>: interest rates on zero-coupon bonds of different maturity

- s_t one variable, latent variable
- y_t a vector with interest rates for diff. mat.
- x_t one variable, the Central Bank discount rate
- *Φ, and Ω* constants
- *α* and *β* vectors of constants
- R matrix

$$s_t = \phi \cdot s_{t-1} + u_t$$

State equation

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{pmatrix} = \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \vdots \\ \alpha_{n,t} \end{pmatrix} \cdot x_t + \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \vdots \\ \beta_{n,t} \end{pmatrix} \cdot s_t + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix}$$

Observation equation

The State Space Representation: Examples

__Example #4__: an AR(2) process

$$z_t = \rho_1 \cdot z_{t-1} + \rho_2 \cdot z_{t-2} + v_t \ , \ \ v_t \sim N(0, \sigma_v^2)$$

- Can we still apply the state space representation?
  - Yes!
- Consider the following <u>state</u> equation:

$$\underbrace{\begin{bmatrix} z_t \\ z_{t-1} \end{bmatrix}}_{s_t} = \underbrace{\begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix}}_{\phi} \underbrace{\begin{bmatrix} z_{t-1} \\ z_{t-2} \end{bmatrix}}_{s_{t-1}} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix} v_t}_{u_t}$$

- And the <u>observation</u> equation:

$$\underbrace{z_t}_{y_t} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\beta} \cdot \mathrm{s}_\mathrm{t}$$

The State Space Representation: Examples

Example #4: an AR(2) process

$$z_t = \rho_1 \cdot z_{t-1} + \rho_2 \cdot z_{t-2} + v_t \ , \ \ v_t \sim N(0, \sigma_v^2)$$

- The state equation:

$$\underbrace{\begin{bmatrix} z_t \\ z_{t-1} \end{bmatrix}}_{s_t} = \underbrace{\begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix}}_{\phi} \underbrace{\begin{bmatrix} z_{t-1} \\ z_{t-2} \end{bmatrix}}_{s_{t-1}} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix} v_t}_{u_t}$$

- And the observation equation:

$$\underbrace{z_t}_{y_t} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\beta} \cdot \mathrm{s}_{\mathrm{t}}$$

- What are matrices Ω (var-cov of u_t) and R (var-cov of ε_t) in this case?

The State Space Representation Is Not Unique!

Consider the same AR(2) process

$$z_t = \rho_1 \cdot z_{t-1} + \rho_2 \cdot z_{t-2} + v_t \ , \ \ v_t \sim N(0, \sigma_v^2)$$

- Another possible <u>state</u> equation:

$$\underbrace{\begin{bmatrix} z_t \\ \rho_2 z_{t-1} \end{bmatrix}}_{s_t} = \underbrace{\begin{bmatrix} \rho_1 & 1 \\ \rho_2 & 0 \end{bmatrix}}_{\phi} \underbrace{\begin{bmatrix} z_{t-1} \\ \rho_2 z_{t-2} \end{bmatrix}}_{s_{t-1}} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix} v_t}_{u_t}$$

- And the corresponding <u>observation</u> equation:

$$\underbrace{z_t}_{y_t} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\beta} \cdot \mathrm{s_t}$$

- These two state space representations are equivalent!
- This example can be extended to AR(p) case

The State Space Representation: Examples

Example #5: an MA(2) process

$$z_t = v_t + \theta \cdot v_{t-1}\ ,\ v_t \sim N(0,\sigma_v^2)$$

- Consider the following state equation:

$$\underbrace{\begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix}}_{s_t} = \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}}_{\phi} \underbrace{\begin{bmatrix} v_{t-1} \\ v_{t-2} \end{bmatrix}}_{s_{t-1}} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix} v_t}_{u_t}$$

- And the observation equation:

$$\underbrace{z_t}_{y_t} = \underbrace{\begin{bmatrix} 1 & \theta \end{bmatrix}}_{\beta} \cdot \mathrm{s_t}$$

- What are matrices Ω (var-cov of u_t) and R (var-cov of ε_t) in this case?

The State Space Representation: Examples

Example #5: an MA(2) process

$$z_t = v_t + \theta \cdot v_{t-1} \ , \ v_t \sim N(0, \sigma_v^2)$$

- Consider the following state equation:

$$\underbrace{\begin{bmatrix} y_t \\ \theta \cdot v_t \end{bmatrix}}_{s_t} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\phi} \underbrace{\begin{bmatrix} y_{t-1} \\ \theta \cdot v_{t-1} \end{bmatrix}}_{s_{t-1}} + \underbrace{\begin{bmatrix} 1 \\ \theta \end{bmatrix} v_t}_{u_t}$$

- And the observation equation:

$$\underbrace{z_t}_{y_t} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\beta} \cdot \mathrm{s}_\mathrm{t}$$

- What are matrices Ω (var-cov of u_t) and R (var-cov of ε_t) in this case?

The State Space Representation: Examples

<u>*Example #6*</u>: A random walk plus drift process

$$z_t = z_{t-1} + \gamma + v_t, \; v_t \sim N(0, \sigma_v^2)$$

- <u>State</u> equation? <u>Observation</u> equation?

- What are the loadings ?

- What are matrices Ω (var-cov of u_t) and R (var-cov of ε_t) for your state-space representation?

The State Space Representation: System Stability

- In this course we will deal only with stable systems:
  - Such systems that for any initial state s_0, the state variable (vector) s_t converges to a unique $\bar{s}$ (the steady state)

- The necessary and sufficient condition for the state space representation to be stable is that all eigenvalues of ϕ are less than 1 in absolute value:

$$|\lambda_i(\phi)| < 1 \; for \; \mathrm{all\, i}$$

- Think of a simple univariate AR(1) process $(z_t = \rho_1 \cdot z_{t-1} + v_t)$
  - It is stable as long as $|\rho_1| < 1$

- Why? So that it is possible to be right at least in the "long-run".

The Kalman Filter

Kalman Filter: Introduction

- State Space Representation [univariate case]:

$$s_t = \phi \cdot s_{t-1} + u_t \qquad u_t \sim i.i.d.\, N(0, \sigma_u^2)$$

$$y_t = \alpha \cdot x_t + \beta \cdot s_t + \varepsilon_t \quad \varepsilon_t \sim i.i.d.\, N(0, \sigma_\varepsilon^2)$$

$(\alpha, \beta, \phi, \sigma_u^2, \sigma_\varepsilon^2)$ are known

- Notation:
  - $s_{t|t-1} = E(s_t \mid y_1, \ldots, y_{t-1})$ is the best linear predictor of s_t conditional on the information up to t-1.
  - $y_{t|t-1} = E(y_t \mid y_1, \ldots, y_{t-1})$ is the best linear predictor of y_t conditional on the information up to t-1.
  - $s_{t|t} = E(s_t \mid y_1, \ldots, y_t)$ is the best linear predictor of s_t conditional on the information up to t.

Kalman Filter: Main Idea
Moving from t-1 to t

- Suppose we know $s_{t|t-1}$ and $y_{t|t-1}$ at time t-1.

- When arrive in period t we observe y_t and x_t

- Need to obtain $s_{t|t}$!

Why?

- If we know $s_{t|t}$,
  - using the state equation: $s_{t+1|t} = \phi \cdot s_{t|t}$
  - using the observation equation: $y_{t+1|t} = \alpha x_{t+1} + \beta s_{t+1|t}$

- The key question: how to obtain $s_{t|t}$ from y_t ?

Kalman Filter: Main Idea
How to update $s_{t|t}$?

- Idea: use the observed prediction error $y_t - y_{t|t-1}$ to infer the state at time t, $s_{t|t}$

- It turns out it is optimal to update it using

$$s_{t|t} = s_{t|t-1} + K_t (y_t - y_{t|t-1})$$

- K_t is called Kalman gain
  - It measures how informative is the prediction error about the underlying state vector
    - How do you think it depends on the variance of the observation error?
  - It is chosen so that the new prediction error is orthogonal to all of the previous ones.
    - Thus there is no (linear) predictable component in generated errors.

Kalman Filter: More Notations

- $P_{t|t-1} = E((s_t - s_{t|t-1})^2 \mid y_1, \dots, y_{t-1})$ is the prediction error variance of s_t given the history of observed variables up to t-1.

- $F_{t|t-1} = E((y_t - y_{t|t-1})^2 \mid y_1, \dots, y_{t-1})$ is the prediction error variance of y_t conditional on the information up to t-1.

- $P_{t|t} = E((s_t - s_{t|t})^2 \mid y_1, \dots, y_t)$ is the prediction error variance of s_t conditional on the information up to t.

- Intuitively the Kalman gain is chosen so that $P_{t|t}$ is minimized.
  - Will show this later.

Kalman Gain: Intuition

- Kalman gain is chosen so that $P_{t|t}$ is minimized.

- It can be shown that

$$K_t = \beta P_{t|t-1}(\beta^2 P_{t|t-1} + R)^{-1}$$

- Intuition:
  - If a big mistake is made forecasting $s_{t|t-1}$ ($P_{t|t-1}$ is large), put a lot weight on the new observation (K is large).
  - If the new information is noisy (R is large), put less weight on the new information (K is small).

Kalman Filter: Example

- Kalman gain is $K_t = \beta P_{t|t-1}(\beta^2 P_{t|t-1} + R)^{-1}$
- Consider
  - State equation $s_t = \mu + u_t, u_t \sim N(0, \sigma_u^2)$
  - Observation equation $y_t = s_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$
  - Additionally $\sigma_\varepsilon^2 = \eta \sigma_u^2$, where η is a constant

- Assume that we picked $P_{1|0} = \sigma_u^2$ (we don't know anything about s_1).
- Can you calculate the Kalman gain in the 1st period, K_1?
- What is the interpretation?

Kalman Filter: The last step

- How do we get from $P_{t|t-1}$ to $P_{t+1|t}$ using y_t?
- Recall that for a bivariate normal distribution

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right) \Rightarrow z_2 \mid z_1 \sim N(\mu_2 + \sigma_{12}(\sigma_1^2)^{-1}(z_1 - \mu_1), \sigma_2^2 - \sigma_{12}(\sigma_1^2)^{-1}\sigma_{12})$$

- Using this property and the fact that $E[s_{t|t-1} \mid y_t] = s_{t|t}$

$$\begin{pmatrix} y_t \\ s_t \end{pmatrix} \mid (y_1, \ldots, y_{t-1}) \sim N\left(\begin{pmatrix} y_{t|t-1} \\ s_{t|t-1} \end{pmatrix}, \begin{pmatrix} F_{t|t-1} & \beta P_{t|t-1} \\ \beta P_{t|t-1} & P_{t|t-1} \end{pmatrix} \right)$$

$$\boxed{E\left[(s_t - s_{t|t-1})(y_t - y_{t|t-1}) \mid y_1, \ldots y_{t-1}\right] = E\left[(s_t - s_{t|t-1})(\beta(s_t - s_{t|t-1}) + \varepsilon_t) \mid y_1, \ldots y_{t-1}\right] = \beta P_{t|t-1}}$$

- Thus, $s_{t|t} = s_{t|t-1} + \underline{\beta P_{t|t-1}(F_{t|t-1})^{-1}}(y_t - y_{t|t-1})$ and

Kalman gain

$$P_{t|t} = P_{t|t-1} - \beta P_{t|t-1}(F_{t|t-1})^{-1}\beta P_{t|t-1}$$

Kalman Filter: Finally

- From the previous slide

$$s_{t|t} = s_{t|t-1} + \beta P_{t|t-1}(F_{t|t-1})^{-1}(y_t - y_{t|t-1})$$
$$P_{t|t} = P_{t|t-1} - \beta P_{t|t-1}(F_{t|t-1})^{-1}\beta P_{t|t-1}$$

- Need: from $P_{t|t-1}$ to $P_{t+1|t}$ using y_t

$$F_{t|t-1} = E\left[(y_t - y_{t|t-1})^2 \mid y_1, \ldots y_{t-1}\right] = E\left[(\beta(s_t - s_{t|t-1}) + \varepsilon_t)^2 \mid y_1, \ldots y_{t-1}\right] = \beta^2 P_{t|t-1} + R$$

- Thus, we get the expression for the Kalman gain:

$$K_t = \beta P_{t|t-1}(\beta^2 P_{t|t-1} + R)^{-1}$$

- Similarly

$$P_{t+1|t} = E\left[(s_{t+1} - s_{t+1|t})^2 \mid y_1, \ldots, y_t\right] = E\left[(\phi(s_t - s_{t|t}) + u_t)^2 \mid y_1, \ldots y_t\right] = \phi^2 P_{t|t} + \Omega$$

- And we are done!

Kalman Filter: Review

- We start from $s_{t|t-1}$ and $P_{t|t-1}$.

$$y_{t|t-1} = \alpha x_t + \beta s_{t|t-1}$$

$$F_{t|t-1} = \beta^2 P_{t|t-1} + R$$

- Calculate Kalman gain

$$K_t = \beta P_{t|t-1}(\beta^2 P_{t|t-1} + R)^{-1}$$

- Update using observed y_t

$$s_{t|t} = s_{t|t-1} + K_t(y_t - y_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - \beta P_{t|t-1}(F_{t|t-1})^{-1}\beta P_{t|t-1}$$

- Construct forecasts for the next period

$$s_{t+1|t} = \phi \cdot s_{t|t}$$

$$P_{t+1|t} = \phi^2 P_{t|t} + \Omega$$

- Repeat!

Kalman Filter:
How to choose initial state

- If the sample size is large, the choice of the initial state is not very important
- In short samples can have significant effect
- For stationary models

$$s_{1|0} = s^*$$
$$P_{1|0} = P^*$$

- Where

$$s^* = \phi \cdot s^*$$
$$P^* = \phi P^* \phi' + \Omega$$

- Solution to the last equation is $P^* = [I - \phi \otimes \phi]^{-1} vec(\Omega)$
- Why? Under some very general conditions

$P_{t|t-1} \to P^*$ as $t \to \infty$

Kalman Filter as a Recursive Regression

- Consider a regular regression function

$$E[s \mid y] = a + by$$

where

$$a = E[s] - b \cdot E[y]$$

$$b = Cov(s, y) \cdot (Var(y))^{-1}$$

- Substituting

$$E[s \mid y] = E[s] + Cov(s, y) \cdot Var(y)^{-1} \cdot [y - E[y]]$$

- From one of the previous slides:

$$s_{t|t} = s_{t|t-1} + \beta P_{t|t-1}(F_{t|t-1})^{-1}(y_t - y_{t|t-1})$$

Kalman Filter as a Recursive Regression

- Consider a regular regression function

$$E[s \mid y] = a + by$$

where

$$a = E[s] - b \cdot E[y]$$

$$b = Cov(s, y) \cdot (Var(y))^{-1}$$

- Substituting

$$E[s \mid y] = E[s] + Cov(s, y) \cdot Var(y)^{-1} \cdot [y - E[y]]$$

- From one of the previous slides

$$s_{t|t} = s_{t|t-1} + \beta P_{t|t-1} (F_{t|t-1})^{-1} (y_t - y_{t|t-1})$$

Because

$$Cov(s_t, y_t \mid y_1, \ldots y_{t-1}) = \beta P_{t|t-1}$$

$$Var(y_t \mid y_1, \ldots, y_{t-1}) = F_{t|t-1}$$

$$s_{t|t-1} = E(s_t \mid y_1, \ldots, y_{t-1}) \qquad y_{t|t-1} = E(y_t \mid y_1, \ldots, y_{t-1}) \qquad s_{t|t} = E[E(s_t \mid y_1, \ldots, y_{t-1}) \mid y_t]$$

Kalman Filter as a Recursive Regression

- Thus the Kalman filter can be interpreted as a recursive regression of a type

$$s_t = \sum_{\tau=1}^{t} y_\tau \theta_\tau + v_t$$

where $v_t = s_t - \sum_{\tau=1}^{t} y_\tau \theta_\tau$ is the forecasting error at time t

- The Kalman filter describes how to recursively estimate

θ_t and thus obtain $s_{t|t} = E[s_t \mid y_1,\dots y_t]$

Optimality of the Kalman Filter

- Using the property of OLS estimates that constructed residuals are uncorrelated with regressors

$$E[v_t] = 0 \quad E[v_t y_t] = 0 \text{ for all t}$$

- Using the expression for

$$v_t = s_t - \sum_{\tau=1}^{t} y_\tau \theta_\tau$$

and the state equation, it is easy to show that

$$E[v_t y_{t-k}] = 0$$

for all t and k=0,..t-1

- Thus the errors v_t do not have any (linear) predictable component!

Kalman Filter
Some comments

- Within the class of linear (in observables) predictors the Kalman filter algorithm **minimizes the mean squared prediction error** (i.e., predictions of the state variables based on the Kalman filter are best linear unbiased):

$$\underset{\mathrm{K_t}}{Min}\ \mathrm{E}[\left(\mathrm{s_t} - (s_{t|t-1} + K_t(y_t - y_{t|t-1}))\right)^2]$$

$$\Rightarrow \mathrm{K_t} = \frac{\beta P_{t|t-1}}{\beta^2 P_{t|t-1} + R}$$

- If the model disturbances are normally distributed, predictions based on the Kalman filter are optimal (its MSE is minimal) among all predictors:

$$\mathrm{E}[\left(\mathrm{s_t} - (s_{t|t-1} + K_t(y_t - y_{t|t-1}))\right)^2] \leq \underset{f(y_1,\dots,y_t)}{\mathrm{Min}} E[\left(\mathrm{s_t} - f(y_1,\dots,y_t)\right)^2]$$

- In this sense, the Kalman filter delivers **optimal predictions**.

Kalman Filter - Multivariate Case

- The Kalman Filter algorithm can be easily generalized to the generic multivariate state space representation, including exogenous variables:

$$\mathbf{s}_{t+1} = \mathbf{\Phi}\mathbf{s}_t + \mathbf{u}_{t+1}, \quad E\left[\mathbf{u}_t \mathbf{u}_t'\right] = \mathbf{\Omega}$$

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{s}_t + \mathbf{\varepsilon}_t, \quad E\left[\mathbf{\varepsilon}_t \mathbf{\varepsilon}_t'\right] = \mathbf{R}$$

- Defining similarly as before:

$$\mathbf{s}_{t|t-1} = E\left[\mathbf{s}_t \mid \mathbf{y}_1,\ldots,\mathbf{y}_{t-1}\right], \mathbf{P}_{t|t-1} = E\left[\left(\mathbf{s}_t - \mathbf{s}_{t|t-1}\right)\left(\mathbf{s}_t - \mathbf{s}_{t|t-1}\right)' \mid \mathbf{y}_1,\ldots,\mathbf{y}_{t-1}\right]$$

$$\mathbf{y}_{t|t-1} = E\left[\mathbf{y}_t \mid \mathbf{y}_1,\ldots,\mathbf{y}_{t-1}\right], \mathbf{F}_t = E\left[\left(\mathbf{y}_t - \mathbf{y}_{t|t-1}\right)\left(\mathbf{y}_t - \mathbf{y}_{t|t-1}\right)' \mid \mathbf{y}_1,\ldots,\mathbf{y}_{t-1}\right]$$

- Now we have vectors and matrices

Kalman Filter Algorithm – Multivariate Case

Initialization: $\mathbf{s}_{1|0}, \mathbf{P}_{1|0}$

$$1: \mathbf{y}_{t|t-1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{s}_{t|t-1}$$

$$2: \mathbf{F}_t = \mathbf{B}\mathbf{P}_{t|t-1}\mathbf{B}' + \mathbf{R}$$

$$3: \mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{B}'(\mathbf{F}_t)^{-1}$$

$$4: \mathbf{s}_{t|t} = \mathbf{s}_{t|t} + \mathbf{K}(\mathbf{y}_t - \mathbf{y}_{t|t-1})$$

$$5: \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{B}'(\mathbf{F}_t)^{-1}\mathbf{B}\mathbf{P}_{t|t-1}$$

$$6: \mathbf{s}_{t+1|t} = \mathbf{\Phi}\mathbf{s}_{t|t}$$

$$7: \mathbf{P}_{t+1|t} = \mathbf{\Phi}\mathbf{P}_{t|t}\mathbf{\Phi}' + \mathbf{\Omega}$$

Repeat 1,...,7 from $t = 1$ to $t = T$

Kalman Filter Algorithm – Multivariate Case (cont.)

How do we obtain these expressions?

$$\mathbf{y}_{t|t-1} = E\left[\mathbf{y}_t \mid \mathbf{y}_1,\dots,\mathbf{y}_{t-1}\right] = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{s}_{t|t-1}$$

$$\mathbf{F}_t = E\left[(\mathbf{y}_t - \mathbf{y}_{t|t-1})(\mathbf{y}_t - \mathbf{y}_{t|t-1})' \mid \mathbf{y}_1,\dots,\mathbf{y}_{t-1}\right]$$

$$= E\left[(B(\mathbf{s}_t - \mathbf{s}_{t|t-1}) + \varepsilon_t)(B(\mathbf{s}_t - \mathbf{s}_{t|t-1}) + \varepsilon_t)' \mid \mathbf{y}_1,\dots,\mathbf{y}_{t-1}\right] = \mathbf{B}\mathbf{P}_{t|t-1}\mathbf{B}' + \mathbf{R}$$

Also:

$$E\left[(\mathbf{y}_t - \mathbf{y}_{t|t-1})(\mathbf{s}_t - \mathbf{s}_{t|t-1})' \mid \mathbf{y}_1,\dots,\mathbf{y}_{t-1}\right] = E\left[(\mathbf{B}(\mathbf{s}_t - \mathbf{s}_{t|t-1}) + \varepsilon_t)(\mathbf{s}_t - \mathbf{s}_{t|t-1})' \mid \mathbf{y}_1,\dots,\mathbf{y}_{t-1}\right]$$

$$= \mathbf{B}\mathbf{P}_{t|t-1}$$

Thus:

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{s}_t \end{pmatrix} \mid \mathbf{y}_1,\dots,\mathbf{y}_{t-1} \sim N\left(\begin{pmatrix} \mathbf{y}_{t|t-1} \\ \mathbf{s}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{F}_t & \mathbf{B}\mathbf{P}_{t|t-1} \\ (\mathbf{B}\mathbf{P}_{t|t-1})' & \mathbf{P}_{t|t-1} \end{pmatrix} \right)$$

Kalman Filter Algorithm – Multivariate Case (cont.)

How we obtained these expression? (cont.)

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{s}_t \end{pmatrix} | \mathbf{y}_1,\ldots,\mathbf{y}_{t-1} \sim N\left(\begin{pmatrix} \mathbf{y}_{t|t-1} \\ \mathbf{s}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{F}_t & \mathbf{BP}_{t|t-1} \\ \mathbf{P}_{t|t-1}\mathbf{B'} & \mathbf{P}_{t|t-1} \end{pmatrix} \right)$$

Using the property for a multivariate normal distribution to get conditional disttribution:

$$\mathbf{s}_{t|t} \sim N\left(\underbrace{\mathbf{s}_{t|t-1} + \underbrace{\mathbf{P}_{t|t-1}\mathbf{B'}(\mathbf{F}_t)^{-1}}_{\mathbf{K}_t} \left(\mathbf{y}_t - \mathbf{y}_{t|t-1}\right)}_{\mathbf{s}_{t|t}}, \underbrace{\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{B'}(\mathbf{F}_t)^{-1}\mathbf{BP}_{t|t-1}}_{\mathbf{P}_{t|t}} \right)$$

Also:

$$\mathbf{s}_{t+1|t} = E\left[\mathbf{\Phi s}_t + \mathbf{u}_{t+1} | \mathbf{y}_1,\ldots,\mathbf{y}_t\right] = \mathbf{\Phi s}_{t|t}$$

$$\mathbf{P}_{t+1|t} = E\left[(\mathbf{s}_{t+1} - \mathbf{s}_{t+1|t})(\mathbf{s}_{t+1} - \mathbf{s}_{t+1|t})' | \mathbf{y}_1,\ldots,\mathbf{y}_t\right] = \mathbf{\Phi P}_{t|t}\mathbf{\Phi}' + \mathbf{\Omega}$$

ML Estimation and Kalman Smoothing

Maximum Likelihood Estimation

- The algorithm in the previous section assumes knowledge of the parameters. If these are not known, estimates are needed.
- Consider the univariate case:

$$s_{t+1} = \phi s_t + u_{t+1}, E\left[u_t^2\right] = \sigma_u^2$$

$$y_t = \alpha x_t + \beta s_t + \varepsilon_t, E\left[\varepsilon_t^2\right] = \sigma_\varepsilon^2$$

and using that s_t is normally distributed (u_t is normal) then

$$(y_t \mid y_1, \ldots, y_{t-1}) \sim N(y_{t|t-1}, F_{t|t-1})$$

- Thus we can do maximum likelihood estimation

$$\log l(\phi, \alpha, \beta, \sigma_u, \sigma_\varepsilon) = -\sum_{t=1}^{T}\left[\frac{1}{2}\log 2\pi + \frac{1}{2}\log|F_{t|t-1}| + \frac{1}{2F_{t|t-1}}(y_t - y_{t|t-1})^2\right]$$

- Similarly with the multivariate case:

$$(\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{y}_1, \ldots, \mathbf{y}_{t\text{-}1}) \sim N(\mathbf{y}_{t|t\text{-}1}, \mathbf{F}_t)$$

Maximum Likelihood Estimation

To estimate model parameters through maximizing log-likelihood:

Step 1: For every set of the underlying parameters, θ

Step 2: run the Kalman filter to obtain estimates for the sequence

$$y_{t|t-1}, F_{t|t-1}$$

Step 3: Construct the likelihood function as a function of θ

Step 4: Maximize with respect to the parameters.

Kalman Smoothing

- For each period *t*, the Kalman filter uses only information available up to time *t*:

$$E[\mathbf{s}_{\mathbf{t}} \mid \mathbf{y}_{1},\ldots,\mathbf{y}_{\mathbf{t-1}}] = \mathbf{s}_{\mathbf{t|t-1}}$$

- Is it possible to use all the information available so as to obtain an even better estimate of s_t: $E[\mathbf{s}_{\mathbf{t}} \mid \mathbf{y}_{1},\ldots,\mathbf{y}_{\mathbf{T}}]$?

- This is called smoothed inference of the state and denoted by

$$\mathbf{s}_{\mathbf{t|T}}$$

- In general, we can obtain the smoothed inference

$$\mathbf{s}_{\mathbf{t|}}\ , \tau > t$$

Kalman Smoothing

Using the same principles for normal conditional distribution, it is possible to show that there is a recursive algorithm to compute $\mathbf{s}_{t|T}$ starting from $\mathbf{s}_{T|T}$:

<u>*Step 1*</u>: use Kalman filter to estimate $\mathbf{s}_{1|1}, \ldots, \mathbf{s}_{T|T}$

<u>*Step 2*</u>: use recursive method to obtain, $\mathbf{s}_{t|T}$, the smoothed estimate of s_t:

$$\mathbf{s}_{t|T} = \mathbf{s}_{t|t} + \mathbf{J}_t(\mathbf{s}_{t+1|T} - \mathbf{s}_{t+1|t})$$

where

$$\mathbf{J}_t = \mathbf{P}_{t|t}\Phi'(\mathbf{P}_{t+1|t})^{-1}$$

Conclusion

- Many models require estimations of unobserved variables, either because these are of economic interest, or because one needs them to estimate the model parameters (example, ARMA).

- The Kalman filter is a recursive algorithm that:
  - provides efficient estimates of unobserved variables, and their MSE;
  - can be used for forecasting given estimates of MSE;
  - is used to initialize maximum likelihood estimation of models (for example, of ARMA models) by first producing good estimates of un-observed variables;
  - can also be used to smooth series for unobserved variables.