Time Series Data Analysis and Forecasting

MSDS

Module 5

Introduction to State Space Models

Topic Covered

- What are state space models
- Linear gaussian models
- Filtering, smoothing and forecasting
- Kalman filter and Kalman smoother
- Maximum likelihood estimation
- Missing data modifications
- Structural models
- Signal extraction and forecasting
- State space model with correlated errors

What are State space models? (1)

- A general model that subsumes a whole class of special cases of interest in much the same way that linear regression does is the state-space model or the dynamic linear model
- it was introduced in Kalman (1960) and Kalman and Bucy (1961)
- Introduced as a method primarily for use in aerospace-related research, the model has been applied to modeling data from economics, medicine and soil sciences.
- A state space model (SSM) is a time series model in which the time series **Yt** is interpreted as the result of a noisy observation of a stochastic process **Xt** . The values of the variables **Xt** and **Yt** can be continuous (scalar or vector) or discrete.

Characteristics of State space models?

State space model is characterized by two principles.

1. there is a hidden or latent process x_t called the state process. State process is a Markov process , meaning that future $\{x_s: s > t\}$, and past $\{x_s: s < t\}$ are independent conditioned on present x_t
2. observations , y_t are independent given states x_t , means that dependence among observations is generated by states.

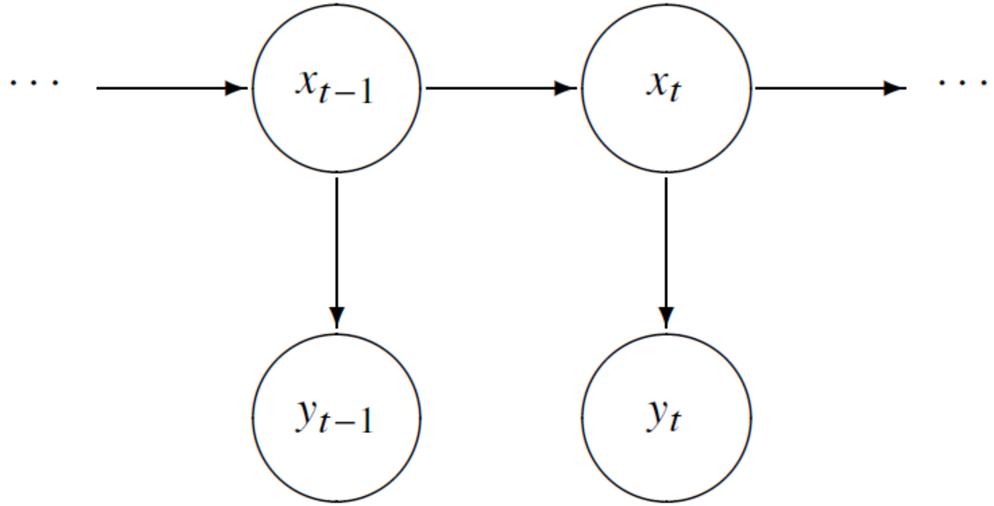
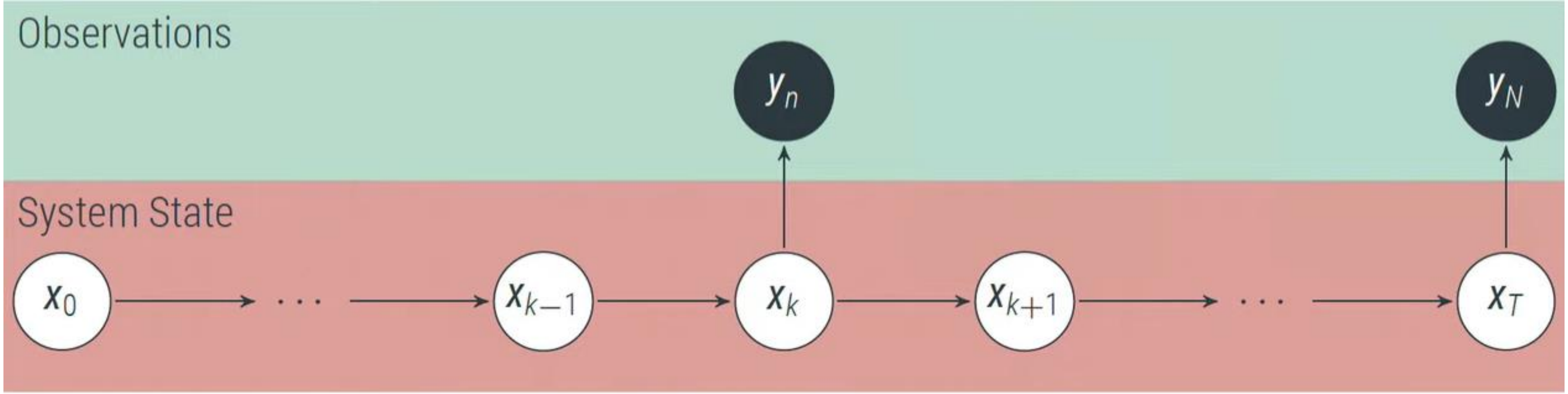


Fig. 6.1. Diagram of a state space model.

$$p(x_k \mid x_{1:k-1}, y_{1:k-1}) = p(x_k \mid x_{k-1}).$$

$$p(y_k \mid x_{1:k}, y_{1:k-1}) = p(y_k \mid x_k)$$

Representation of a Dynamic System



Definition

The *state of a dynamic system* is a **set of physical quantities**, the **specification** of which [...] **completely determines the evolution** of the system.

Definition

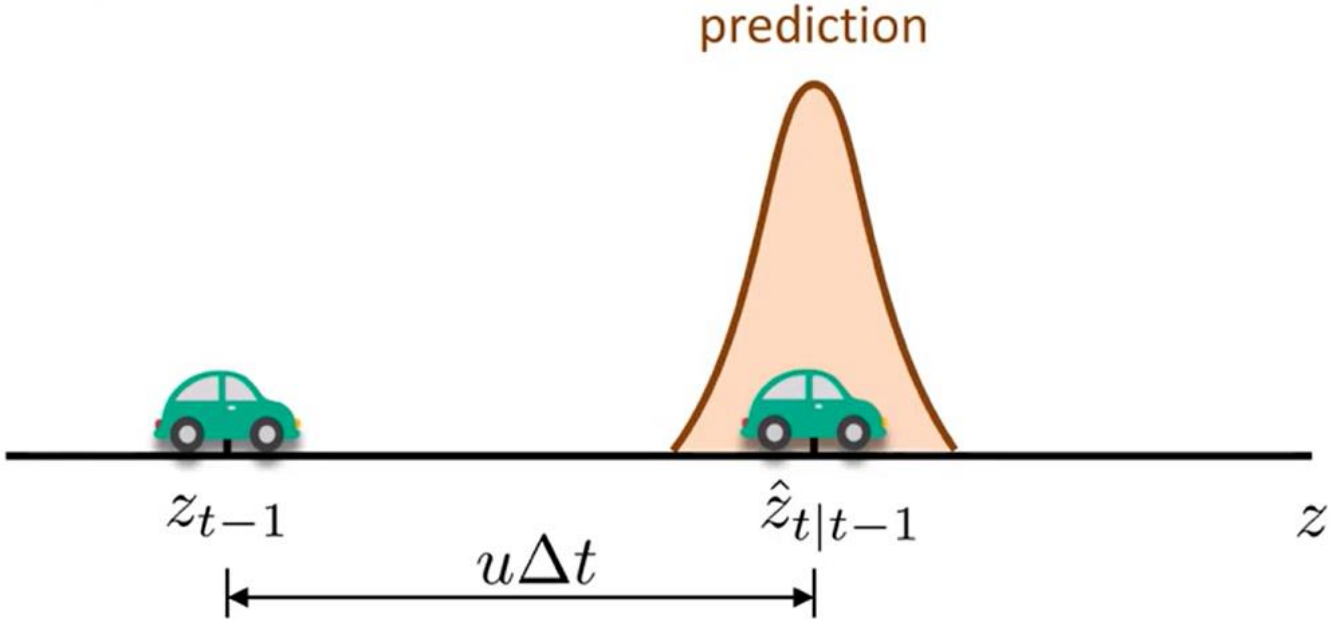
A *time series* is a sequence $[y(t_i)]_{i \in \mathbb{N}}$ of observations $y_i \in \mathbb{Y}$, indexed by a scalar variable $t \in \mathbb{R}$.

Example: 1D Tracking of Constant velocity of car

- A car is driving along the z-axis and its real position (state) at time t is denoted as z_t.
- It has a odometer showing that the car is driving at a 'constant' speed u. It brings in a process noise $w_t \sim \mathcal{N}(0, q)$.
- State equation:

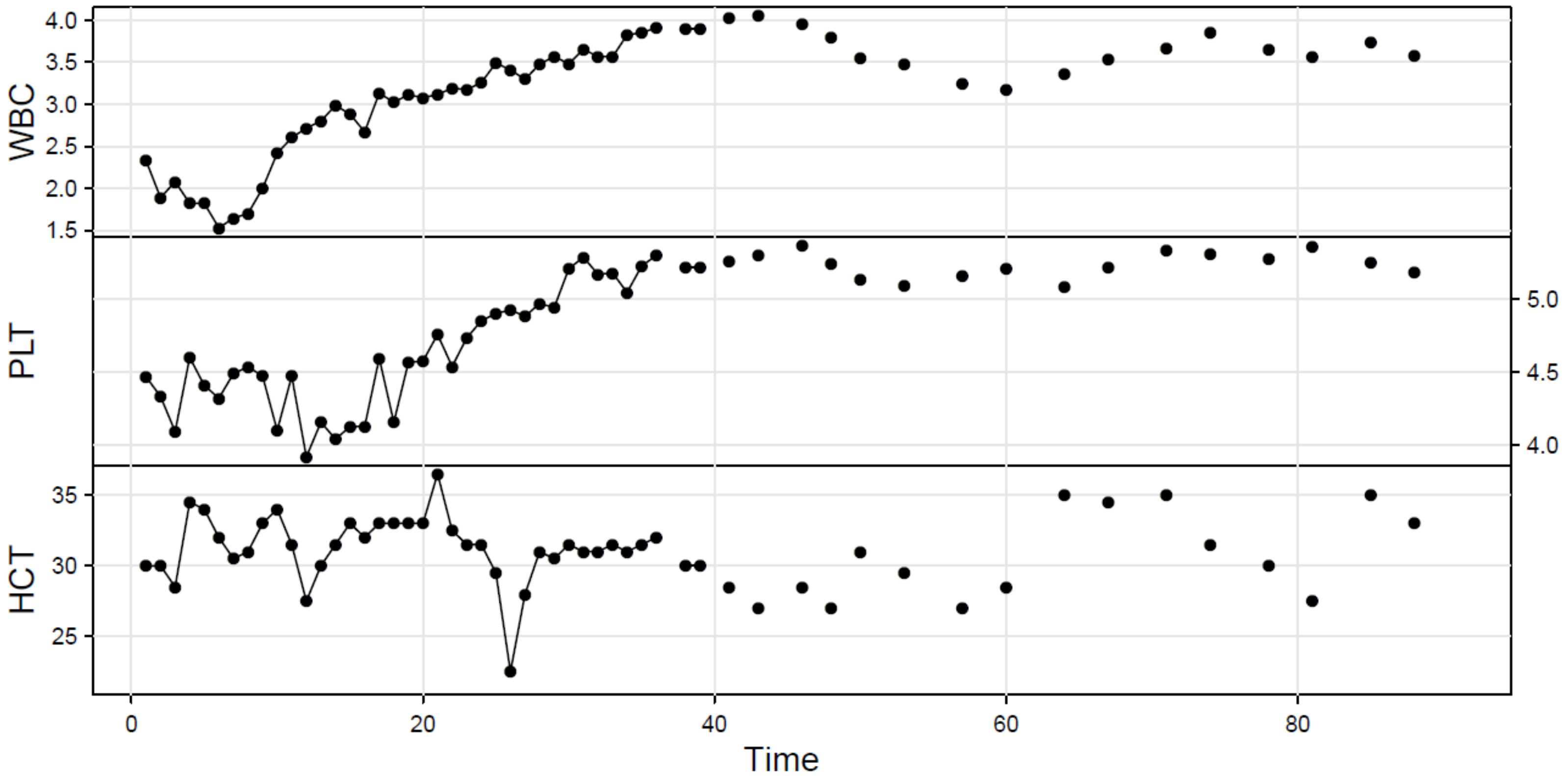
$$z_t = z_{t-1} + u\Delta t + w_t$$

- Given z_{t-1}, the prediction of z_t:



A Biomedical Example (1)

- Consider the problem of monitoring the level of several biomedical markers after a cancer patient undergoes a bone marrow transplant. The data in Figure used by Jones (1984), are measurements made for 91 days on three variables, log(white blood count) [WBC], log(platelet) [PLT], and hematocrit [HCT].



A Biomedical Example (2)

- Approximately 40% of the values are missing, with missing values occurring primarily after the 35th day. The main objectives are to model the three variables using the state-space approach, and to estimate the missing values.
- According to Jones, “Platelet count at about 100 days post transplant has previously been shown to be a good indicator of subsequent long term survival
- Model three variable using state space equation.

$$\begin{pmatrix} x_{t1} \\ x_{t2} \\ x_{t3} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \\ x_{t-1,3} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \\ w_{t3} \end{pmatrix}$$

Filtering, smoothing and forecasting

- Primary aim of analysis using State space model is to produce estimators for underlying **unobserved signal x_t** , given the data ,

$y_{1:s} = \{y_1, \ldots., y_s\}$ to time s.

- As seen state estimation is essential component of parameter estimations.
- When **s<t** , problem is called **forecasting or prediction**
- When **s=t** , problem is called **filtering**
- When **s>t** , problem is called **smoothing**

Quantities of interests

$$x_0 \sim p(x_0) \qquad \text{Initial distribution}$$
$$x_k \mid x_{k-1} \sim p(x_k \mid x_{k-1}) \qquad \text{Markovian dynamics}$$
$$y_k \mid x_k \sim p(y_k \mid x_k) \qquad \text{Measurements}$$

Analyze: Computing different kinds of distributions

Prediction	Filtering	Data likelihood	Smoothing
$p(x_k \mid y_{1:k-1})$	$p(x_k \mid y_{1:k})$	$p(y_k \mid y_{1:k-1})$	$p(x_k \mid y_{1:T})$

We are interested in computing those for every step k in a sequence.

Filtering – Part I: Prediction

- Prediction and filtering distributions are computed from each other

→ **Recursion** starting at x_0

Let us start the recursion from a filtering estimate

$$p(x_k \mid y_{1:k})$$

By basic rules of probability, we compute the **joint distribution**

$$p(x_{k+1}, x_k \mid y_{1:k}) = p(x_{k+1} \mid x_k) p(x_k \mid y_{1:k})$$

Marginalizing over x_k yields the *Chapman-Kolmogorov Equation*

$$p(x_{k+1} \mid y_{1:k}) = \int p(x_{k+1} \mid x_k) p(x_k \mid y_{1:k}) \, \mathrm{d}x_k$$

Filtering – Part II: Correction

Given the prediction (from before)

$$p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:k}) = \int p(\mathbf{x}_{k+1} \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \, \mathrm{d}\mathbf{x}_k$$

we want to include the data point $\mathbf{y}_{k+1}$ into our estimate of $\mathbf{x}_{k+1}$.
How?

Bayes' theorem.

$$p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:k+1}) \propto p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:k}) \, p(\mathbf{y}_{k+1} \mid \mathbf{x}_{k+1})$$

which has to be normalized by dividing by the constant

$$\int p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:k}) \, p(\mathbf{y}_{k+1} \mid \mathbf{x}_{k+1}) \, \mathrm{d}\mathbf{x}_{k+1} = p(\mathbf{y}_{k+1} \mid \mathbf{y}_{1:k})$$

Bayesian Filtering Algorithm

Algorithm 1 Bayesian filtering

procedure BAYESIAN FILTER$(p(x_0), p(x_k \mid x_{k-1}), p(y_k \mid x_k), (y_{k\ldots}))$

  Initialize $k \leftarrow 0$

  while not finished **do**

    Predict: $p(x_{k+1} \mid y_{1:k}) = \int p(x_{k+1} \mid x_k) p(x_k \mid y_{1:k}) \, \mathrm{d}x_k$

    Correct: $p(x_{k+1} \mid y_{1:k+1}) \propto p(x_{k+1} \mid y_{1:k}) \, p(y_{k+1} \mid x_{k+1})$

    $k \leftarrow k + 1$

  end while

  return $\left(p(x_k \mid y_{1:k})\right)_{k=1}^{T}$

end procedure

Linear Gaussian State Space Model

$$x_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$ Initial distribution

$$x_k \mid x_{k-1} \sim \mathcal{N}(A_{k-1}x_{k-1} + b_{k-1}, Q_{k-1})$$ Markovian dynamics

$$y_k \mid x_k \sim \mathcal{N}(H_k x_k + c_k, R_k)$$ Measurements

Analyze: Computing different kinds of distributions

Prediction

$$p(x_k \mid y_{1:k-1}) = \mathcal{N}(\boldsymbol{\mu}_k^-, \boldsymbol{\Sigma}_k^-)$$

Filtering

$$p(x_k \mid y_{1:k}) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Data likelihood

$$p(y_k \mid y_{1:k-1}) = \mathcal{N}(\hat{y}_k, S_k)$$

Smoothing

$$p(x_k \mid y_{1:T}) = \mathcal{N}(\boldsymbol{\xi}_k, \boldsymbol{\Lambda}_k)$$

Gaussian Inference

Theorem

$$
\begin{aligned}
\textit{If} \quad & p(x) = \mathcal{N}(x; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\textit{and} \quad & p(y \mid x) = \mathcal{N}(y; Ax + b, \boldsymbol{\Lambda}), \\
\textit{then} \quad & p(y) = \mathcal{N}(y; A\boldsymbol{\mu} + b, \boldsymbol{\Lambda} + A\boldsymbol{\Sigma}A^\top) \\
\textit{and} \quad & p(x \mid y) = \mathcal{N}(x; \boldsymbol{\mu} + \underbrace{\boldsymbol{\Sigma}A^\top(A\boldsymbol{\Sigma}A^\top + \boldsymbol{\Lambda})^{-1}}_{\textit{gain}} \underbrace{(y - (A\boldsymbol{\mu} + b))}_{\textit{residual}}, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}A^\top(\underbrace{A\boldsymbol{\Sigma}A^\top + \boldsymbol{\Lambda}}_{\textit{Gram matrix}})^{-1}A\boldsymbol{\Sigma})
\end{aligned}
$$

Kalman filtering

- It is an effective and versatile mathematical procedure for combining **noisy sensor** outputs to estimate the state of a system with uncertain dynamics.
- Kalman filtering is a relatively recent (1960) development in filtering.
- Kalman filtering has been applied in areas as diverse as aerospace, tracking missiles, navigation, nuclear power plant instrumentation, demographic modeling, manufacturing, computer vision applications.
- For Kalman filter the problem is formulated is state space and is time varying.

What is Kalman filter?

- The Kalman filter is a linear, recursive estimator that produces the minimum variance estimate in a least squares sense under the assumption of white, Gaussian noise processes.

Delay is the price for filtering

What is Kalman filter?

- Filtering in linear Gaussian state-space models $\Rightarrow$ ***Kalman filter***.

$\rightarrow$ Compute prediction- and filtering distribution + data likelihood in *closed form*.

1. **Prediction:** $p(x_k \mid y_{1:k-1}) = \mathcal{N}(\boldsymbol{\mu}_k^-, \boldsymbol{\Sigma}_k^-)$

$$\boldsymbol{\mu}_k^- = A_{k-1}\boldsymbol{\mu}_{k-1} + b_{k-1} \qquad \boldsymbol{\Sigma}_k^- = A_{k-1}\boldsymbol{\Sigma}_{k-1}A_{k-1}^\top + Q_{k-1}$$

2. **Correction:** $p(x_k \mid y_{1:k}) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$

$$\hat{y}_k = H_k\boldsymbol{\mu}_k^- + c_k \qquad S_k = H_k\boldsymbol{\Sigma}_k^- H_k^\top + R_k$$

$$K_k = \boldsymbol{\Sigma}_k^- H_k^\top S_k^{-1}$$

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}_k^- + K_k(y_k - \hat{y}_k) \qquad \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_k^- - K_k S_k K_k^\top$$

Note: $p(y_k \mid y_{1:k-1}) = \mathcal{N}(y_k; \hat{y}_k, S_k)$

Kalman Filtering Algorithm

Algorithm 2 The Kalman filter

procedure KALMAN FILTER($\boldsymbol{\mu}_0, \Sigma_0, A_k \ldots, Q_k \ldots, b_k \ldots H_k \ldots, R_k \ldots, c_k \ldots, y_k \ldots$)

  Initialize $k \leftarrow 0$

  while not finished **do**

    $\boldsymbol{\mu}^- \leftarrow A_{k-1} \boldsymbol{\mu}_{k-1} + b_{k-1}$ // predict mean

    $\Sigma^- \leftarrow A_{k-1} \Sigma_{k-1} A_{k-1}^\top + Q_{k-1}$ // predict covariance

    ---

    $S_k \leftarrow H_k \Sigma_k^- H_k^\top + R_k$

    $K_k \leftarrow \Sigma_k^- H_k^\top S_k^{-1}$

    $\boldsymbol{\mu}_k \leftarrow \boldsymbol{\mu}_k^- + K_k(y_k - (H_k \boldsymbol{\mu}_k^- + c_k))$ // correct mean

    $\Sigma_k \leftarrow \Sigma_k^- - K_k S_k K_k^\top$ // correct covariance

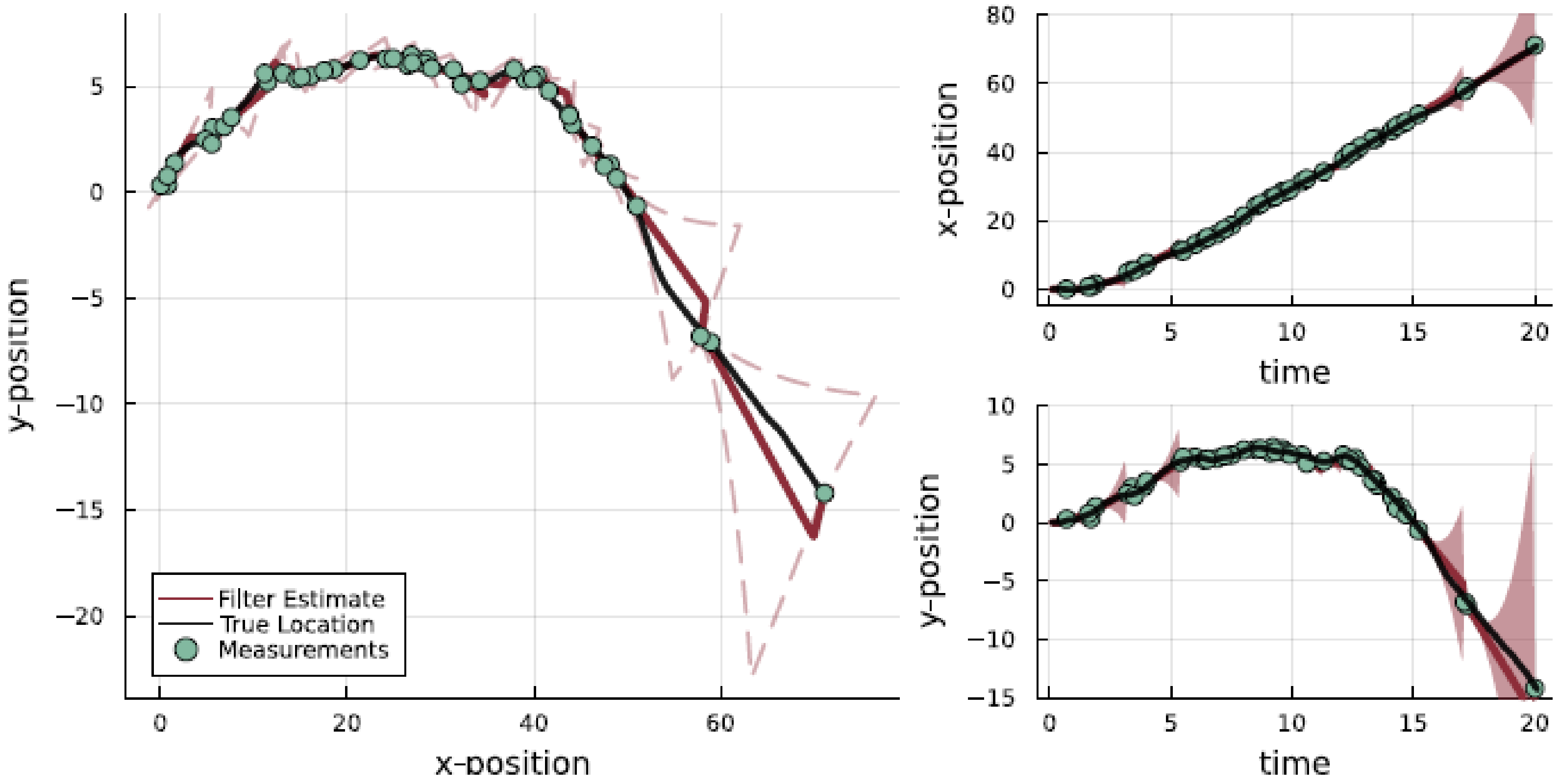
    $k \leftarrow k + 1$

  end while

  return $((\boldsymbol{\mu}_k, \Sigma_k)_{k=1}^T)$

end procedure

Example : Car tracking data



- At every step in the filtering process, only past data is known and used
  - $\rightarrow$ That's great for on-line estimation!
- **If there is one:** at the final time step T, we have seen all data

Kalman Smoother

- Kalman smoothers are used widely to estimate the state of a linear dynamical system from noisy measurements
- Goal in smoothing is to **reconstruct or approximate** the missing measurements given the **known/past** measurements.
- Since the outputs and states are jointly Gaussian, the maximum likelihood and conditional mean estimates of the missing output values are the same, and can be found as the solution of constrained least squares problem

Kalman Smoother

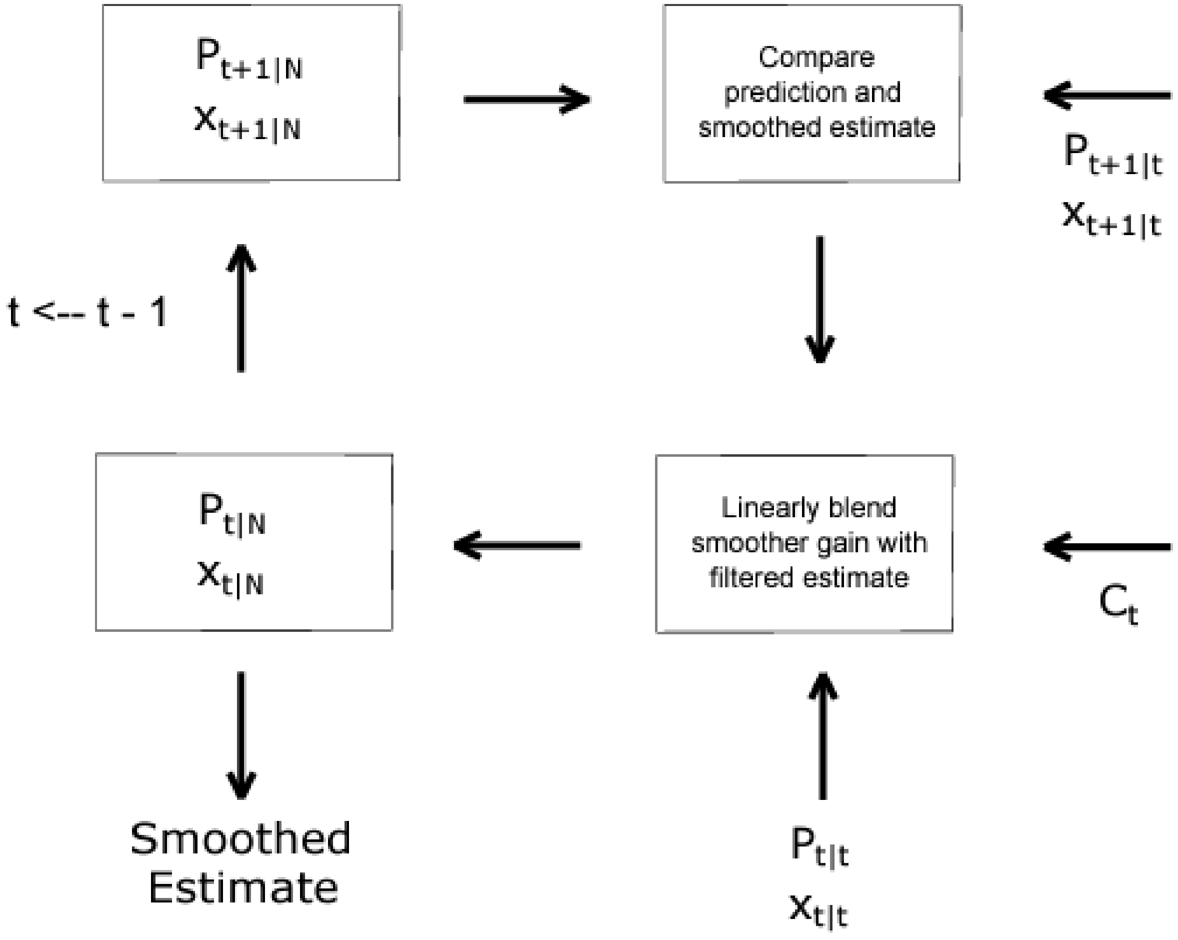


FIGURE 2.2: Kalman smoother loop

Example: Predicting, filtering and smoothing from local level model

For this example, we simulated $n = 50$ observations from the local level trend model discussed in Example 6.4. We generated a random walk

$$\mu_t = \mu_{t-1} + w_t \tag{6.51}$$

with $w_t \sim \text{iid } \mathrm{N}(0, 1)$ and $\mu_0 \sim \mathrm{N}(0, 1)$. We then supposed that we observe a univariate series y_t consisting of the trend component, μ_t, and a noise component, $v_t \sim \text{iid } \mathrm{N}(0, 1)$, where

$$y_t = \mu_t + v_t \,. \tag{6.52}$$

Example: Predicting, filtering and smoothing from local level model

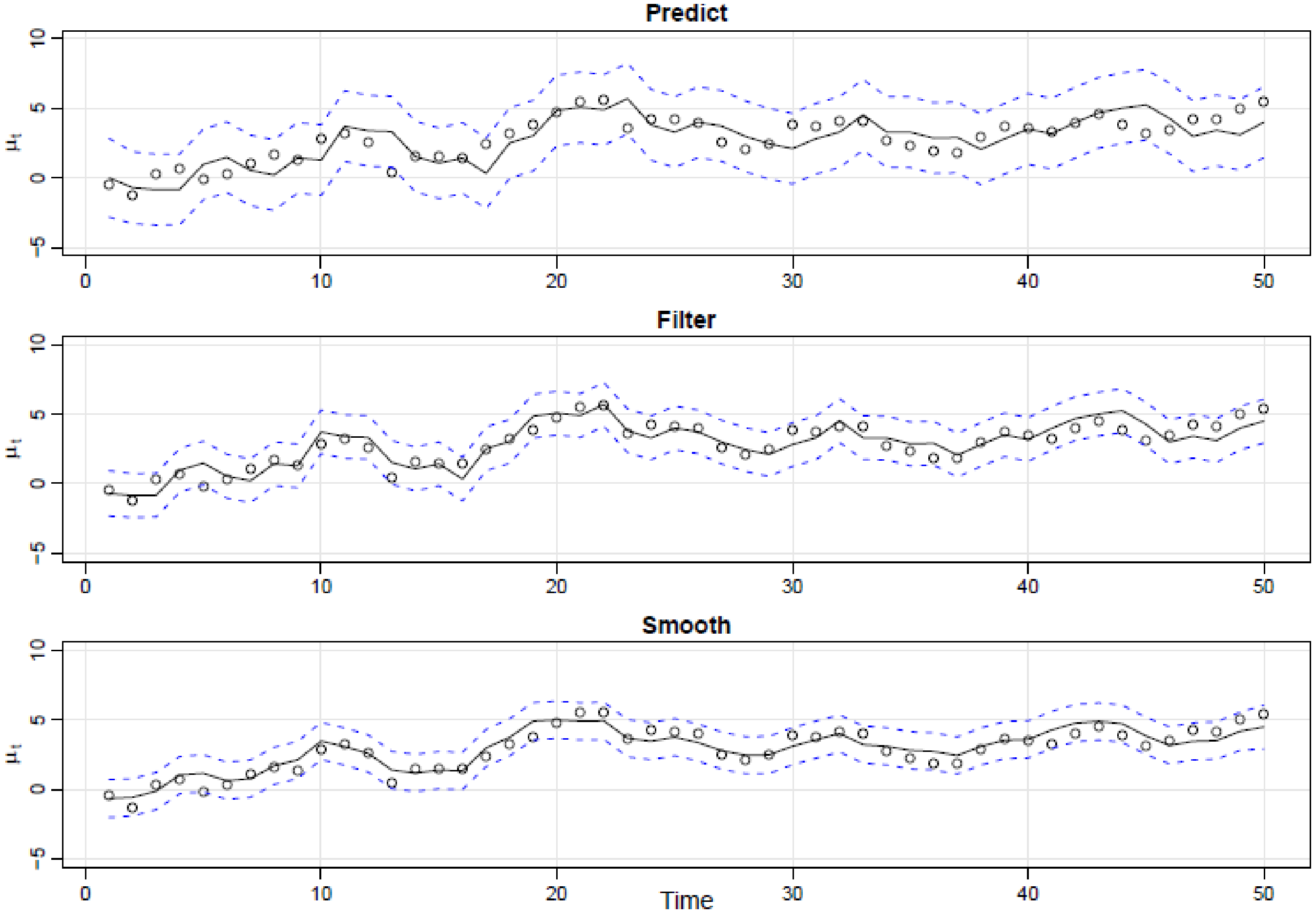


Fig. 6.4. Displays for Example 6.5. The simulated values of μ_t, for $t = 1, \ldots, 50$, given by (6.51) are shown as points. The top shows the predictions μ_t^{t-1} as a line with $\pm 2\sqrt{P_t^{t-1}}$ error bounds as dashed lines. The middle is similar, showing $\mu_t^t \pm 2\sqrt{P_t^t}$. The bottom shows $\mu_t^n \pm 2\sqrt{P_t^n}$.

Example: Predicting, filtering and smoothing from local level model

Table 6.1. First 10 Observations of Example 6.5

t	y_t	μ_t	μ_t^{t-1}	P_t^{t-1}	μ_t^t	P_t^t	μ_t^n	P_t^n
0	—	−.63	—	—	.00	1.00	−.32	.62
1	−1.05	−.44	.00	2.00	−.70	.67	−.65	.47
2	−.94	−1.28	−.70	1.67	−.85	.63	−.57	.45
3	−.81	.32	−.85	1.63	−.83	.62	−.11	.45
4	2.08	.65	−.83	1.62	.97	.62	1.04	.45
5	1.81	−.17	.97	1.62	1.49	.62	1.16	.45
6	−.05	.31	1.49	1.62	.53	.62	.63	.45
7	.01	1.05	.53	1.62	.21	.62	.78	.45
8	2.20	1.63	.21	1.62	1.44	.62	1.70	.45
9	1.19	1.32	1.44	1.62	1.28	.62	2.12	.45
10	5.24	2.83	1.28	1.62	3.73	.62	3.48	.45

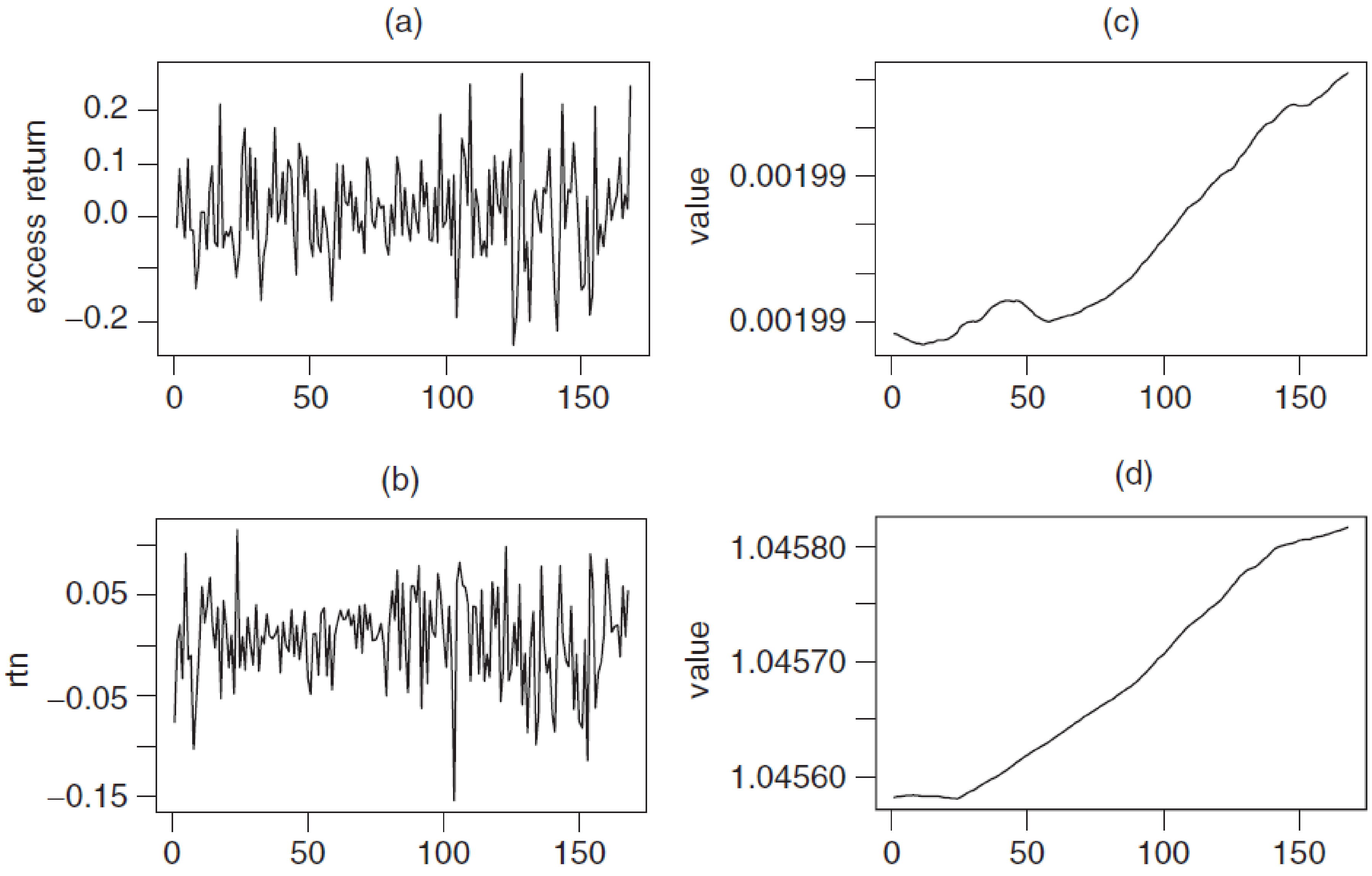


Figure 11.5. Time plots of some statistics for a time-varying CAPM applied to the monthly simple excess returns of General Motors stock. The S&P 500 composite index return is used as the market return: (a) monthly simple excess return, (b) expected returns $\boldsymbol{r}_{t|T}$, (c) α_t estimate, and (d) β_t estimate.

Review Questions

- Discuss Kalman filtering method for a dynamic system.
- Discuss State Space Model for time series data and what one can do with it.
- Discuss the applications of Kalman filter and smoothing.

Practice Questions

11.3. Consider the monthly simple excess returns of Pfizer stock and the S&P 500 composite index from January 1990 to December 2003. The excess returns are in `m-pfesp-ex9003.txt` with Pfizer stock returns in the first column.

(a) Fit a fixed-coefficient market model to the Pfizer stock return. Write down the fitted model.

(b) Fit a time-varying CAPM to the Pfizer stock return. What are the estimated standard errors of the innovations to the α_t and β_t series? Obtain time plots of the smoothed estimates of α_t and β_t.