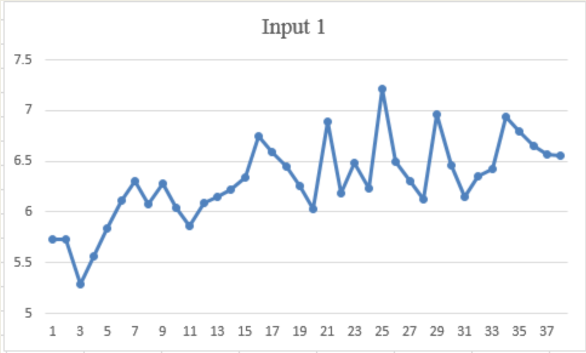
TOPIC 6. TIME SERIES ANALYSIS

Introduction to Time Series

- Time series analysis accounts for the fact that data points taken over time may have an internal structure that should be accounted for.
- Time Series Definition:
  - *An ordered sequence of values of a variable at equally spaced time intervals*
- Time Series Data Example

Expressed in Excel

1	Time	Input 1
2	10/1/08 12:03 AM	5.727265
3	10/1/08 12:03 AM	5.73588
4	10/1/08 12:03 AM	5.292308
5	10/1/08 12:03 AM	5.564445
6	10/1/08 12:03 AM	5.836582
7	10/1/08 12:04 AM	6.108719
8	10/1/08 12:04 AM	6.307732
9	10/1/08 12:04 AM	6.074213
10	10/1/08 12:04 AM	6.279366
11	10/1/08 12:04 AM	6.036408
12	10/1/08 12:04 AM	5.857479
13	10/1/08 12:05 AM	6.08507
14	10/1/08 12:05 AM	6.15545
15	10/1/08 12:05 AM	6.225831
16	10/1/08 12:05 AM	6.342936
17	10/1/08 12:05 AM	6.747737
18	10/1/08 12:05 AM	6.5956
19	10/1/08 12:06 AM	6.443463
20	10/1/08 12:06 AM	6.260338
21	10/1/08 12:06 AM	6.027213
22	10/1/08 12:06 AM	6.890164
23	10/1/08 12:06 AM	6.187493
24	10/1/08 12:06 AM	6.488
25	10/1/08 12:07 AM	6.227664
26	10/1/08 12:07 AM	7.211753
27	10/1/08 12:07 AM	6.502004
28	10/1/08 12:07 AM	6.300909
29	10/1/08 12:07 AM	6.130266
30	10/1/08 12:07 AM	6.965285
31	10/1/08 12:08 AM	6.464217
32	10/1/08 12:08 AM	6.143574
33	10/1/08 12:08 AM	6.352184
34	10/1/08 12:08 AM	6.426536
35	10/1/08 12:08 AM	6.940796
36	10/1/08 12:08 AM	6.797923
37	10/1/08 12:09 AM	6.65505
38	10/1/08 12:09 AM	6.566047
39	10/1/08 12:09 AM	6.557975

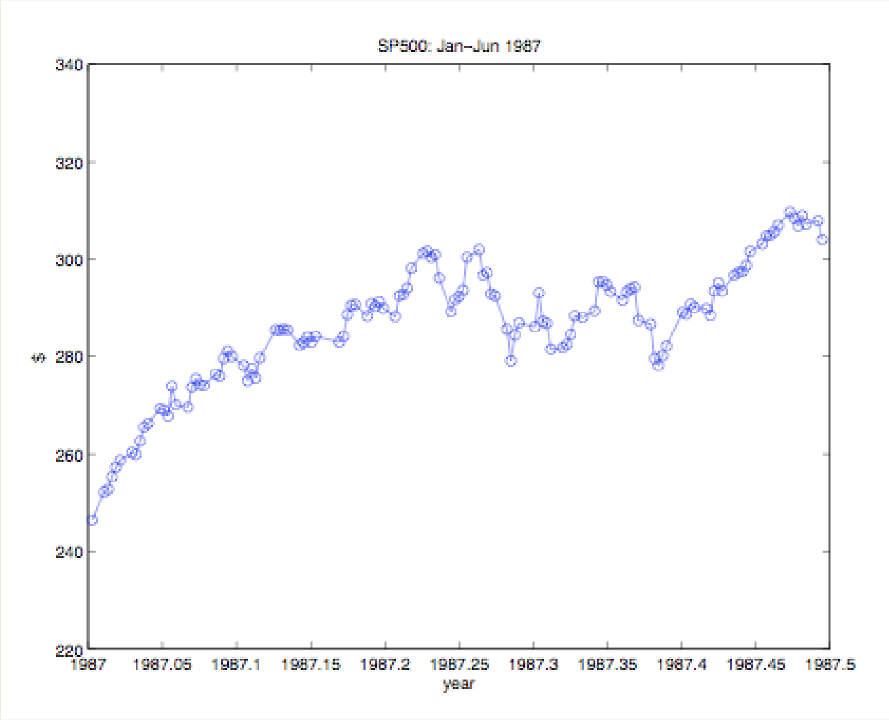


Introduction to Time Series

- Application Goals of Time Series Analyses
  - *Obtain an understanding of the underlying forces and structure that produced the observed data*
  - *Fit a model and proceed to forecasting, monitoring or even feedback and feedforward control*
- Popular Applications
  - *Economic and Sales Forecasting*
  - *Budgetary Analysis*
  - *Stock Market Analysis*
  - *Process and Quality Control*
  - *Inventory Studies*

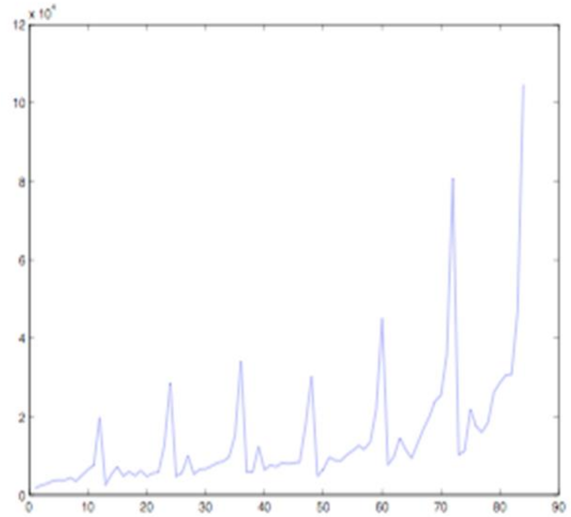
Introduction to Time Series

- Real Time Series Examples



Monthly sales for a souvenir shop at a beach resort town in Queensland.

(Makridakis, Wheelwright and Hyndman, 1998)



Time Series Model

- A time series model specifies the joint distribution of the sequence $\{X_t\}$ of random variables.

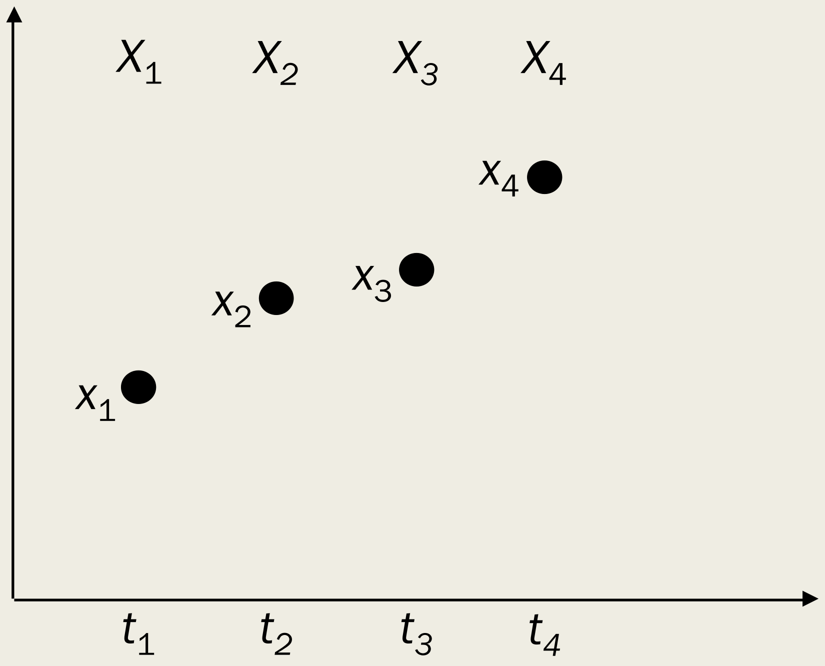
For example:

$P[X_1 \leq x_1, \ldots, X_t \leq x_t]$ for all t and $x_1, \ldots, x_t$.

Notation:

$X_1, X_2, \ldots$ is a stochastic process.

$x_1, x_2, \ldots$ is a single realization.



Simple Time Series Modeling

- Step 1: Plot the time series, look for trends, seasonal components, step changes, outliers.
  - *Trend T_t is typically a linear model depending on time t,* $T_t = \beta_0 + \beta_1 t$
  - *Seasonal component S_t describes the repeated cycles depending on t example is* $S_t = \sum_i (\beta_i \cos(\lambda_i t) + \beta_i' \sin(\lambda_i t))$
- Step 2: Transform data so that residuals are stationary.
  - *Estimate and subtract T_t, S_t*
  - *Differencing*
  - *Nonlinear transformations (log, square-root)*
- Step 3: Fit model to residuals

Time Series Modeling

- Differencing and Trend

Define the lag-1 difference operator,

$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$

where B is the backshift operator, $BX_t = X_{t-1}$.

If $X_t = \beta_0 + \beta_1 t + Y_t$, then $X_{t-1} = \beta_0 + \beta_1(t-1) + Y_{t-1}$ and

$\nabla X_t = \beta_1 + \nabla Y_t$

If $X_t = \sum_{i=0}^{k} \beta_i t^i + Y_t$, then

$\nabla^k X_t = k!\beta_k + \nabla^k Y_t$

where $\nabla^k X_t = \nabla(\nabla^{k-1} X_t)$ and $\nabla^1 X_t = \nabla X_t$

Time Series Modeling

Differencing and seasonal variation

Define the lag-s **difference operator**,

$$\nabla_s X_t = X_t - X_{t-s} = (1 - B^s) X_t,$$

where B^s is the backshift operator applied s times, $B^s X_t = B(B^{s-1} X_t)$ and $B^1 X_t = B X_t$.

If $X_t = T_t + S_t + Y_t$, and S_t has period s (that is, $S_t = S_{t-s}$ for all t), then

$$\nabla_s X_t = T_t - T_{t-s} + \nabla_s Y_t.$$

Time Series Modeling

- Stationarity

$\{X_t\}$ is **strictly stationary** if

for all $k, t_1, \ldots, t_k, x_1, \ldots, x_k$, and h,

$$P(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k) = P(x_{t_1+h} \leq x_1, \ldots, X_{t_k+h} \leq x_k).$$

i.e., shifting the time axis does not affect the distribution.

We shall consider **second-order properties** only.

Time Series Modeling

- Weak Stationarity

Suppose that $\{X_t\}$ is a time series with $E[X_t^2] < \infty$.
Its **mean function** is

$$\mu_t = E[X_t].$$

Its **autocovariance function** is

$$\begin{aligned}\gamma_X(s,t) &= \mathrm{Cov}(X_s, X_t) \\ &= E[(X_s - \mu_s)(X_t - \mu_t)].\end{aligned}$$

We say that $\{X_t\}$ is **(weakly) stationary** if

1. μ_t is independent of t, and
2. For each h, $\gamma_X(t+h, t)$ is independent of t.

In that case, we write

$$\gamma_X(h) = \gamma_X(h, 0).$$

Time Series Modeling

- Example 1 of checking stationarity

Example: i.i.d. noise, $\mathrm{E}[X_t] = 0, \mathrm{E}[X_t^2] = \sigma^2$. We have

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2 & \text{if } h = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

1. $\mu_t = 0$ is independent of t.
2. $\gamma_X(t+h, t) = \gamma_X(h, 0)$ for all t.

So $\{X_t\}$ is stationary.

Similarly for any white noise (uncorrelated, zero mean), $X_t \sim WN(0, \sigma^2)$.

Time Series Modeling

- Example 2 of checking stationarity

Example: Random walk, $S_t = \sum_{i=1}^{t} X_i$ for i.i.d., mean zero $\{X_t\}$.
We have $\mathrm{E}[S_t] = 0, \mathrm{E}[S_t^2] = t\sigma^2$, and

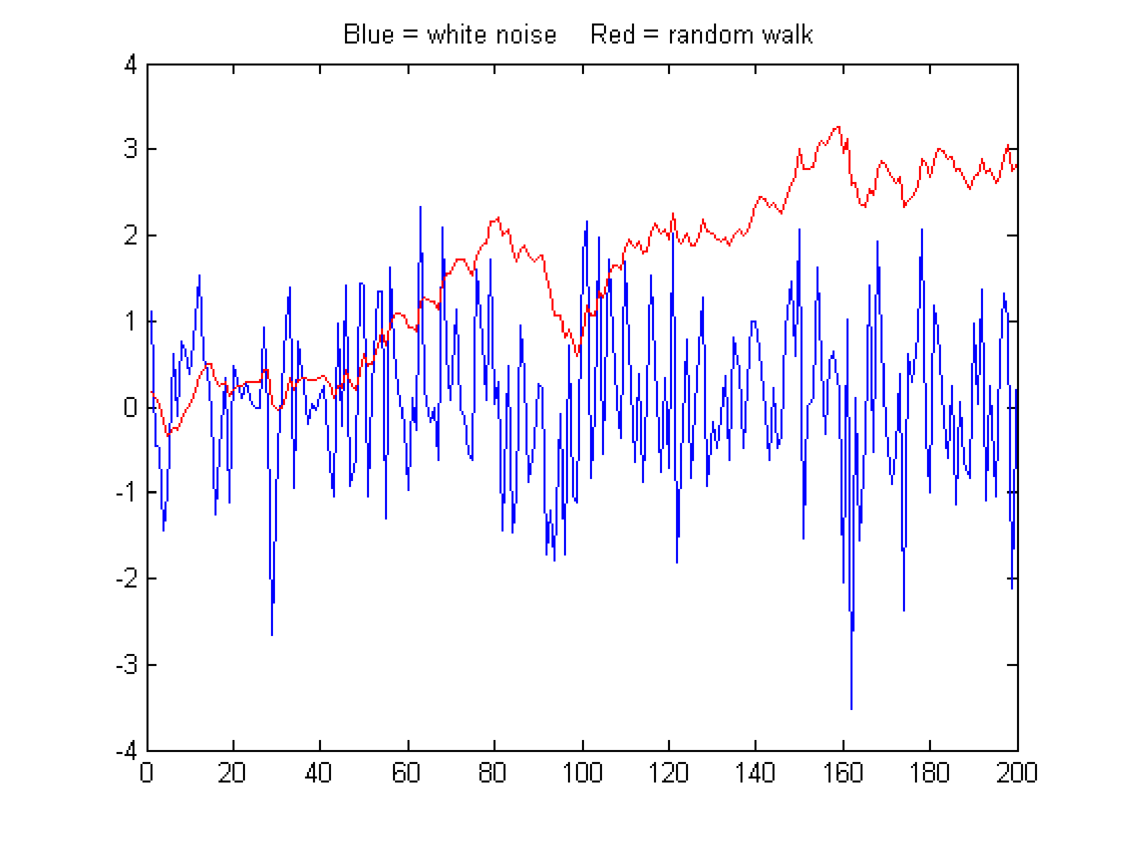
$$\begin{aligned}
\gamma_S(t+h,t) &= \mathrm{Cov}(S_{t+h}, S_t) \\
&= \mathrm{Cov}\left(S_t + \sum_{s=1}^{h} X_{t+s}, S_t\right) \\
&= \mathrm{Cov}(S_t, S_t) = t\sigma^2.
\end{aligned}$$

1. $\mu_t = 0$ is independent of t, but
2. $\gamma_S(t+h,t)$ is not.

So $\{S_t\}$ is not stationary.

Time Series Modeling

- Graphs of the two examples



https://terpconnect.umd.edu/~toh/spectrum/CaseStudies.html

Time-series Modeling Methods

- Famous Classical Techniques:
  - *Box-Jenkins Autoregressive and Moving Average (ARMA) Method for univariate case*
  - *Box-Jenkins Multivariate Models*
  - *Holt-Winters Exponential Smoothing*
- **In this class, we will learn the Autoregressive Model which is the foundation of the ARMA model. More advanced time-series methods will be taught in advanced courses.**

Autoregressive Model

1	Time	Input 1
2	10/1/08 12:03 AM	$\boxed{5.727265}$
3	10/1/08 12:03 AM	$\boxed{5.73588}$
4	10/1/08 12:03 AM	$\boxed{5.292308}$
5	10/1/08 12:03 AM	5.564445
6	10/1/08 12:03 AM	5.836582
7	10/1/08 12:04 AM	6.108719
8	10/1/08 12:04 AM	6.307732
9	10/1/08 12:04 AM	6.074213
10	10/1/08 12:04 AM	6.279366
11	10/1/08 12:04 AM	6.036408
12	10/1/08 12:04 AM	5.857479
13	10/1/08 12:05 AM	6.08507
14	10/1/08 12:05 AM	6.15545
15	10/1/08 12:05 AM	6.225831
16	10/1/08 12:05 AM	6.342936
17	10/1/08 12:05 AM	6.747737
18	10/1/08 12:05 AM	6.5956
19	10/1/08 12:06 AM	6.443463
20	10/1/08 12:06 AM	6.260338
21	10/1/08 12:06 AM	6.027213
22	10/1/08 12:06 AM	6.890164
23	10/1/08 12:06 AM	6.187493
24	10/1/08 12:06 AM	6.488
25	10/1/08 12:07 AM	6.227664
26	10/1/08 12:07 AM	7.211753
27	10/1/08 12:07 AM	6.502004
28	10/1/08 12:07 AM	6.300909
29	10/1/08 12:07 AM	6.130266
30	10/1/08 12:07 AM	6.965285
31	10/1/08 12:08 AM	6.464217
32	10/1/08 12:08 AM	6.143574
33	10/1/08 12:08 AM	6.352184
34	10/1/08 12:08 AM	6.426536
35	10/1/08 12:08 AM	6.940796
36	10/1/08 12:08 AM	6.797923
37	10/1/08 12:09 AM	6.65505
38	10/1/08 12:09 AM	6.566047
39	10/1/08 12:09 AM	6.557975

X_t

X_{t-1}

X_{t-2}

Autoregressive (AR) Model Formulation:

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t$$

where X_t is the time series, W_t is white noise, and

$$\delta = (1 - \sum_{i=1}^{p} \phi_i)\mu$$

where μ is the process mean. p is called the order of the AR model, or AR(p).

Autoregressive Model

- Examples:

If $p = 1$, we will write the AR model with order 1, AR(1), as follows:

$X_t = \Phi_1 X_{t-1} + W_t$

How can we check the stationarity of AR(1)?

Autoregressive Model

Example: AR(1) process (**AutoRegressive**):

$$X_t = \phi X_{t-1} + W_t, \qquad \{W_t\} \sim WN(0, \sigma^2).$$

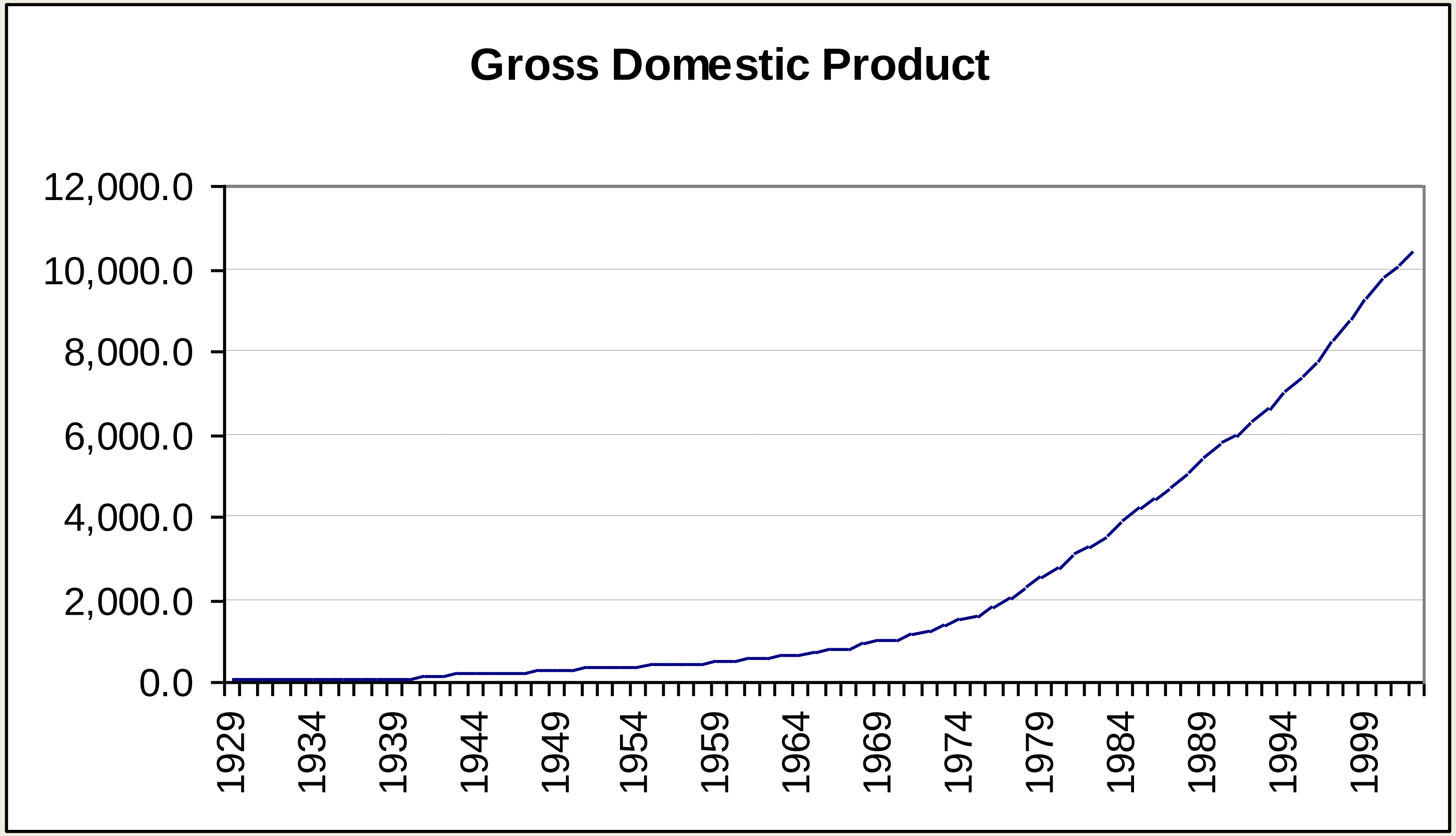
Assume that X_t is stationary and $|\phi| < 1$. Then we have

$$\begin{aligned}
\mathrm{E}[X_t] &= \phi \mathrm{E} X_{t-1} \\
&= 0 \qquad \text{(from stationarity)} \\
\mathrm{E}[X_t^2] &= \phi^2 \mathrm{E}[X_{t-1}^2] + \sigma^2 \\
&= \frac{\sigma^2}{1 - \phi^2} \qquad \text{(from stationarity)},
\end{aligned}$$

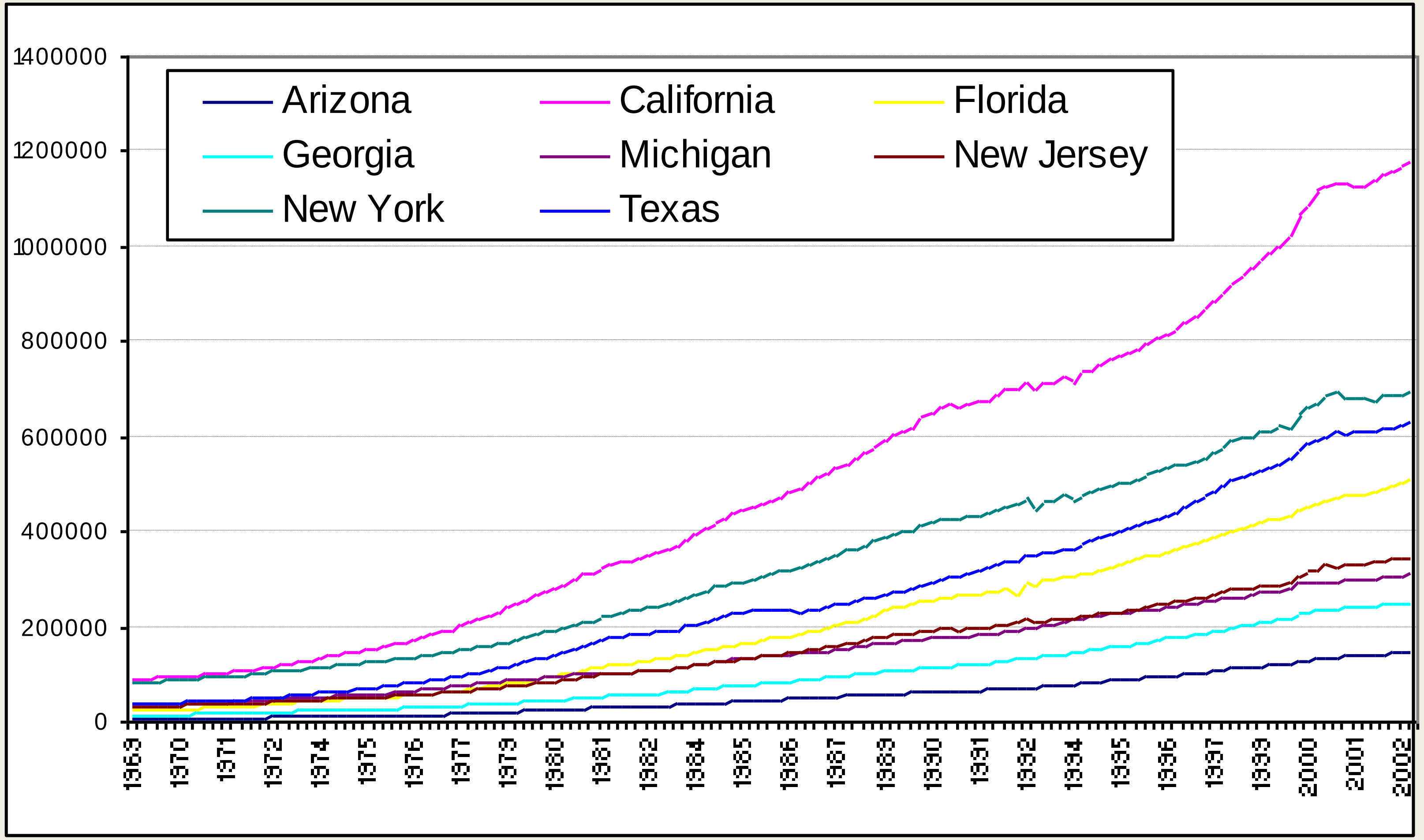
Forecasting Examples

- Forecasting 1 time series
  - *GDP, New construction expenditures, personal savings of British workers*
- Forecasting 10 - 100 time series
  - *Unemployment by states, T-bond demand by locations, % of student returns by departments*
- Forecasting 1000 - millions time series
  - *Daily stock prices/volume of all stocks trading, sales demand by products by locations of a supermarket, customer level spendings*

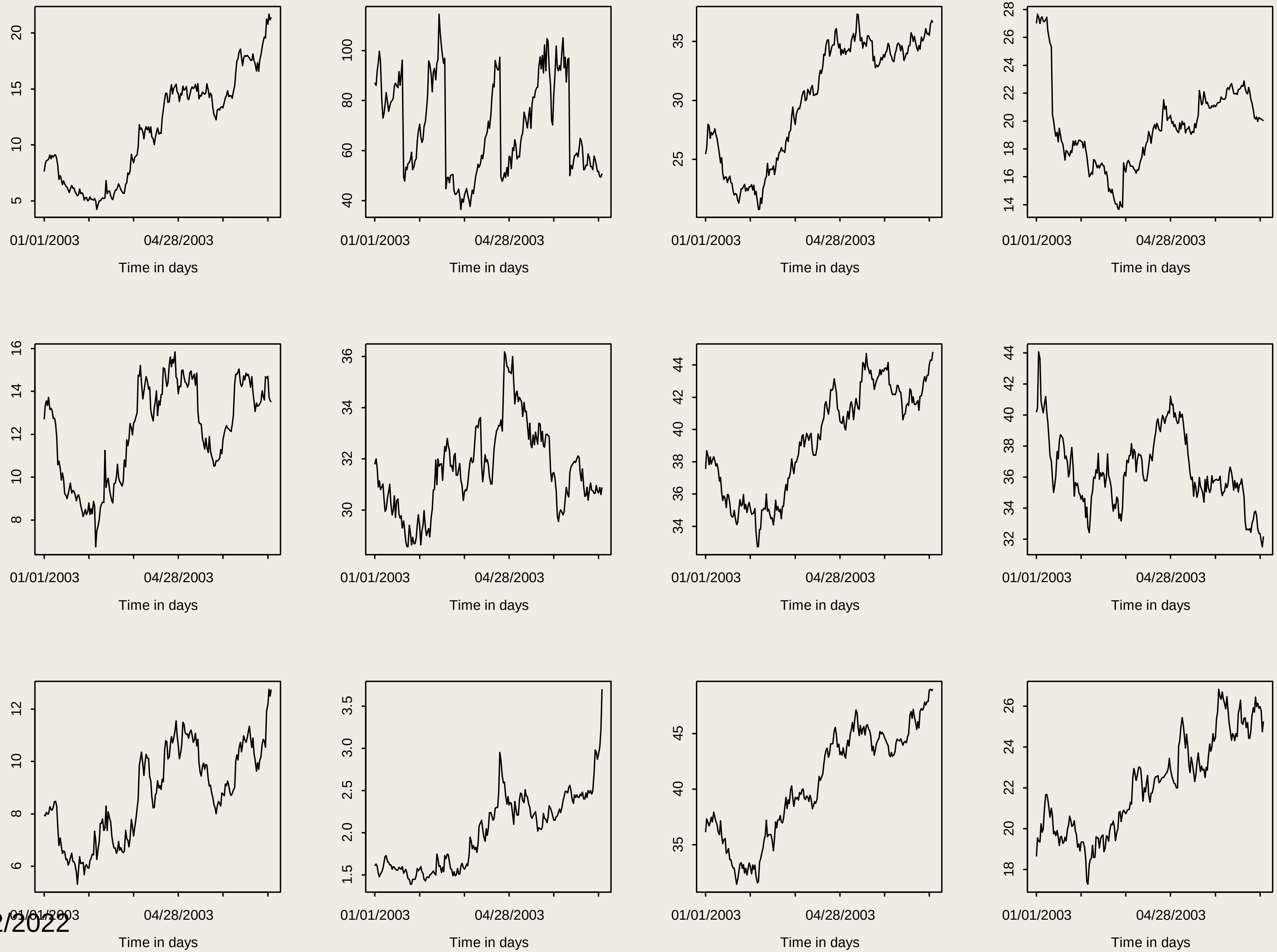
Examples : GDP



Examples : Ave. Personal Incomes



Airline Biotech Bank Telecom



Forecasting Examples

- Customer Churn Modeling: hotel chain, telecom, online grocery
- A retail business has to determine how much it can sell by products and locations in the coming season so that the company can order and distribute
- A financial institution is eager to predict short-term and long-term trends of stock/index price to set buy/hold/sell signal
- A health care management firm wants to know patient admission by sickness by hospitals to allocate its resources
- A telecommunication company wants to predict its demand by technologies and locations to allocate capital spending
- An university wants to forecast students by departments and campus to decide how many professors to hire