Forecasting for Flexible Design: Adaptive Trend Fitting with Stochastic Forecast Errors

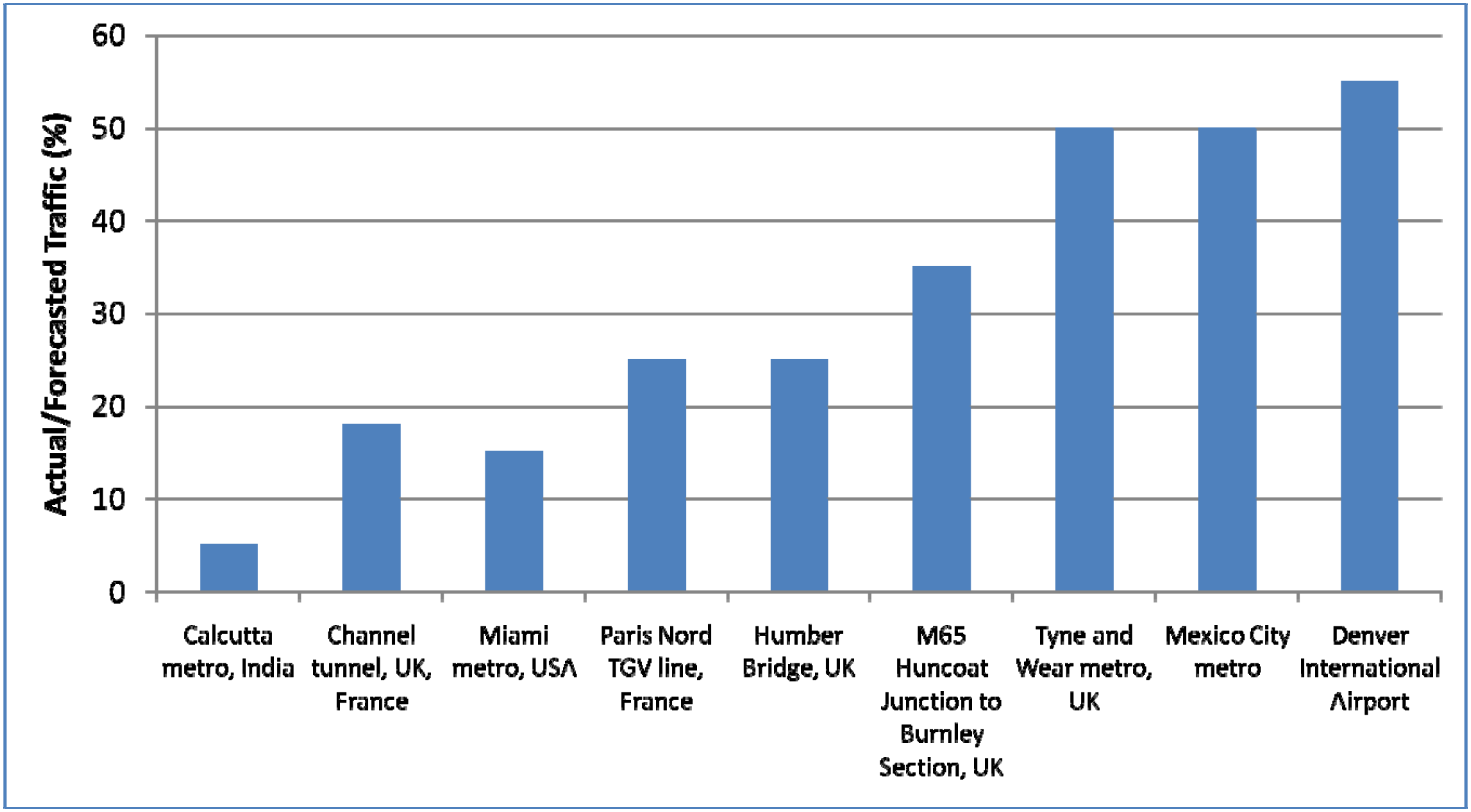
June 2009

Yun S. Lee

Judge Business School, University of Cambridge

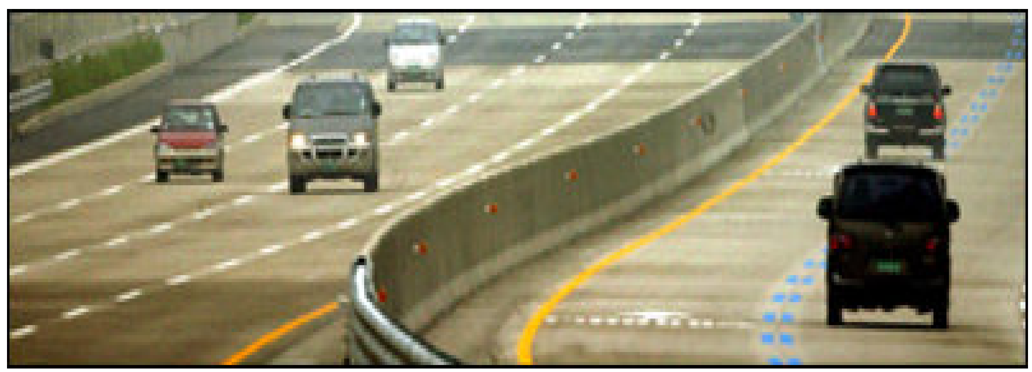
ysL27@cam.ac.uk

Reliance on a deterministic forecast can lead to a white elephant



Forecast errors in infrastructure projects in the transport sector (Flyvbjerg, 2003)

Airport express, Korea (6% of the forecast)

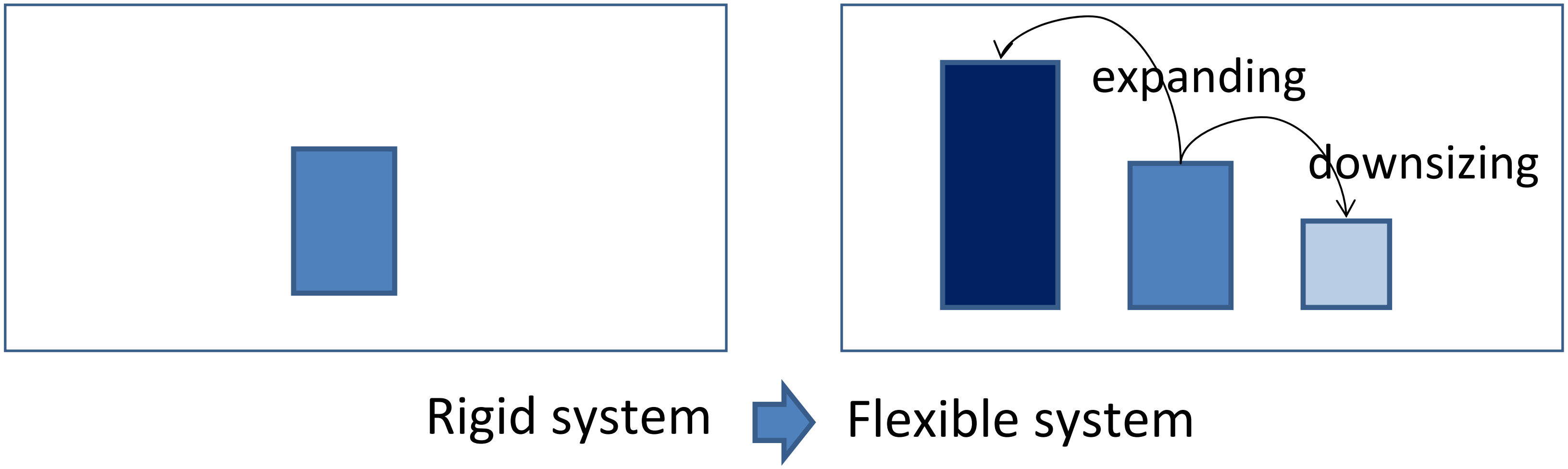


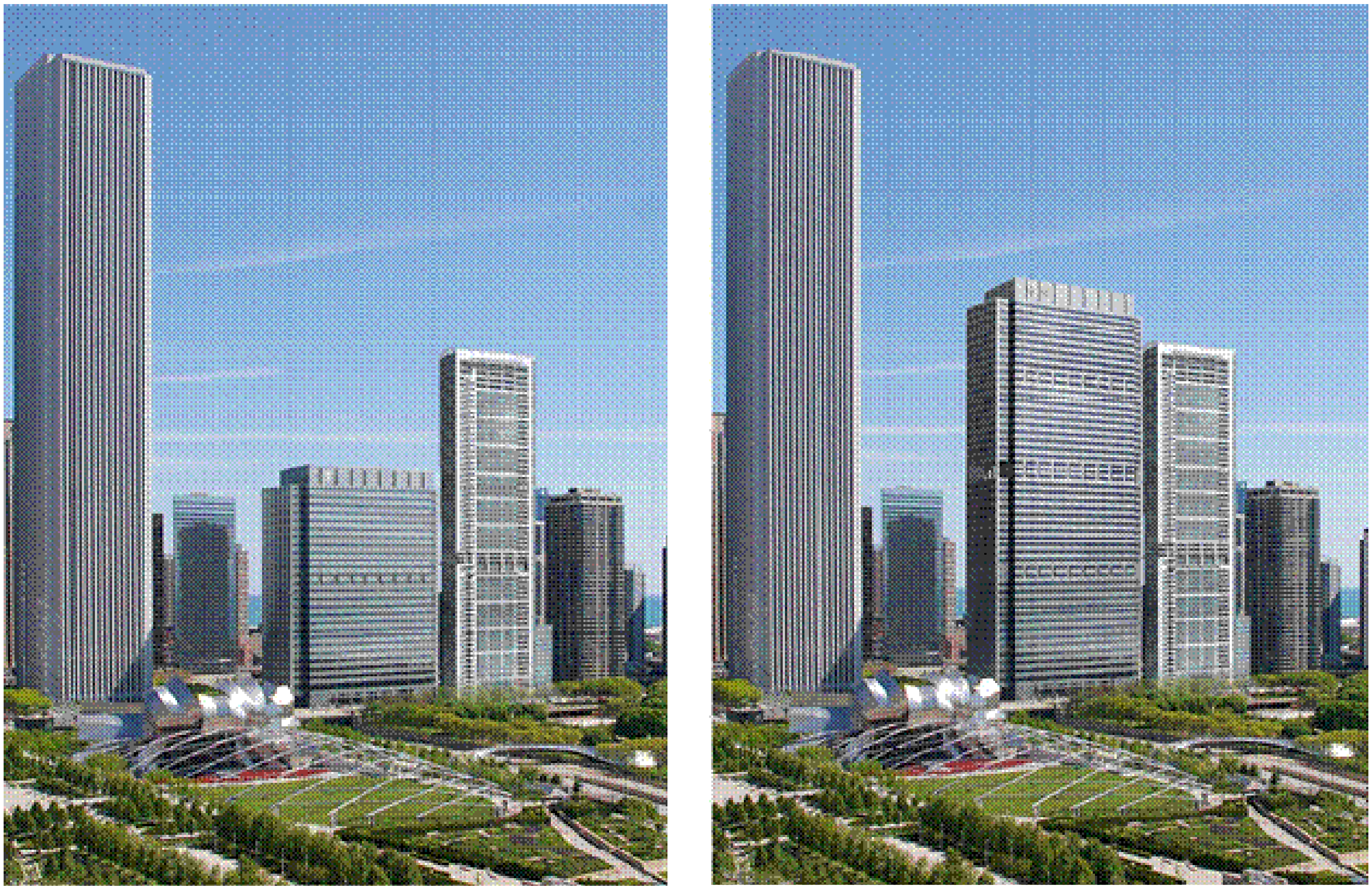
Incheon international airport expressway, Korea (40% of the forecast)

Managing uncertainty: Design Flexibility

- Flexibility is a proactive tool to manage uncertainty.

- We can introduce flexibility in
  - Strategic level e.g. altering the size or usage of a building
  - Tactical level e.g. movable partitions , shell spaces
  - Operational level e.g. flexible furniture systems





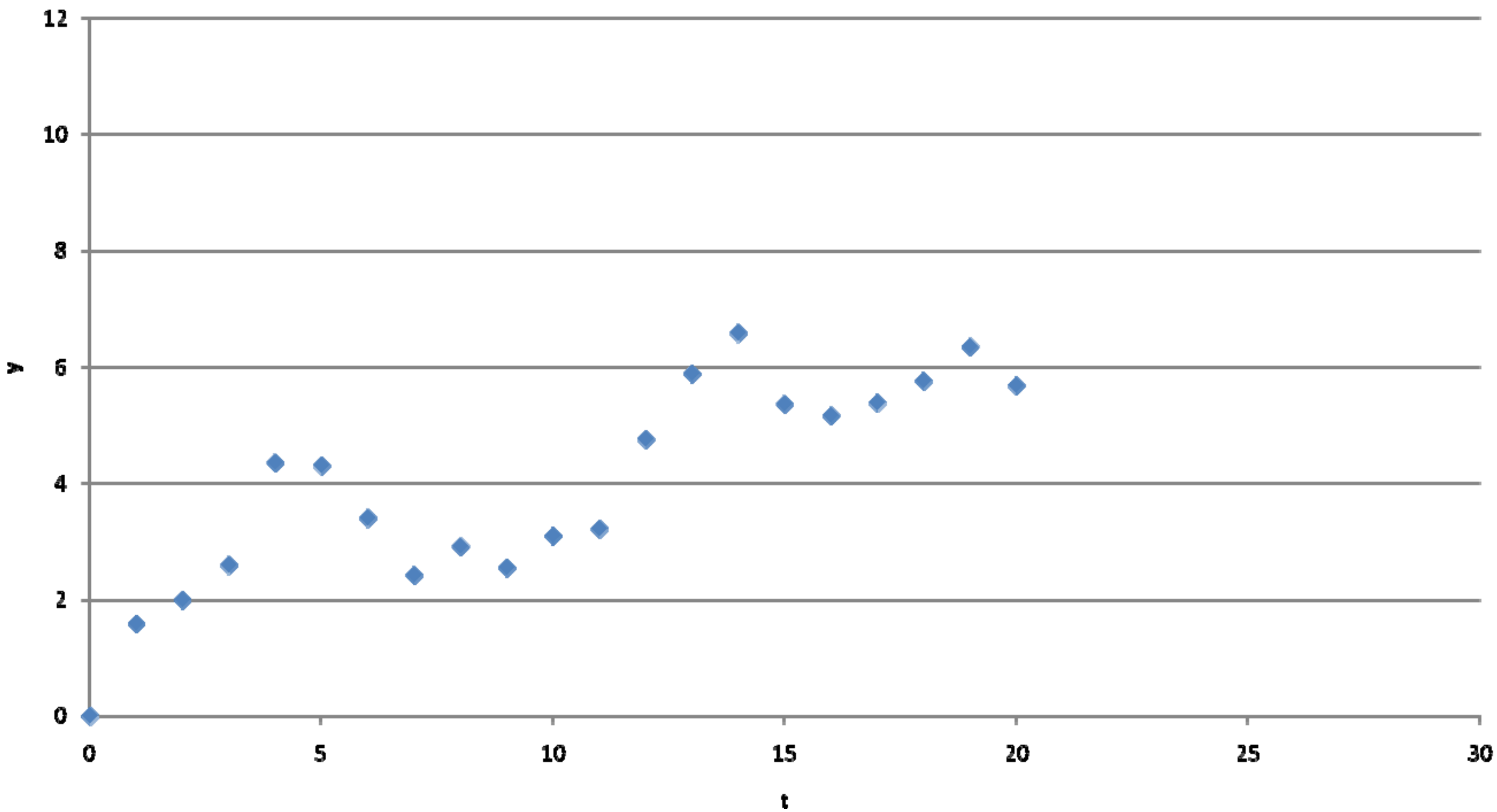
Health Care Service Corporation Building in Chicago: Phase 1 (left, 30 stories) was competed in 1997 and Phase 2 (right, 54 stories) is scheduled for 2010. Source: Goettsch Partners, 2008 and Pearson and Wittels, 2008.

Research Question:

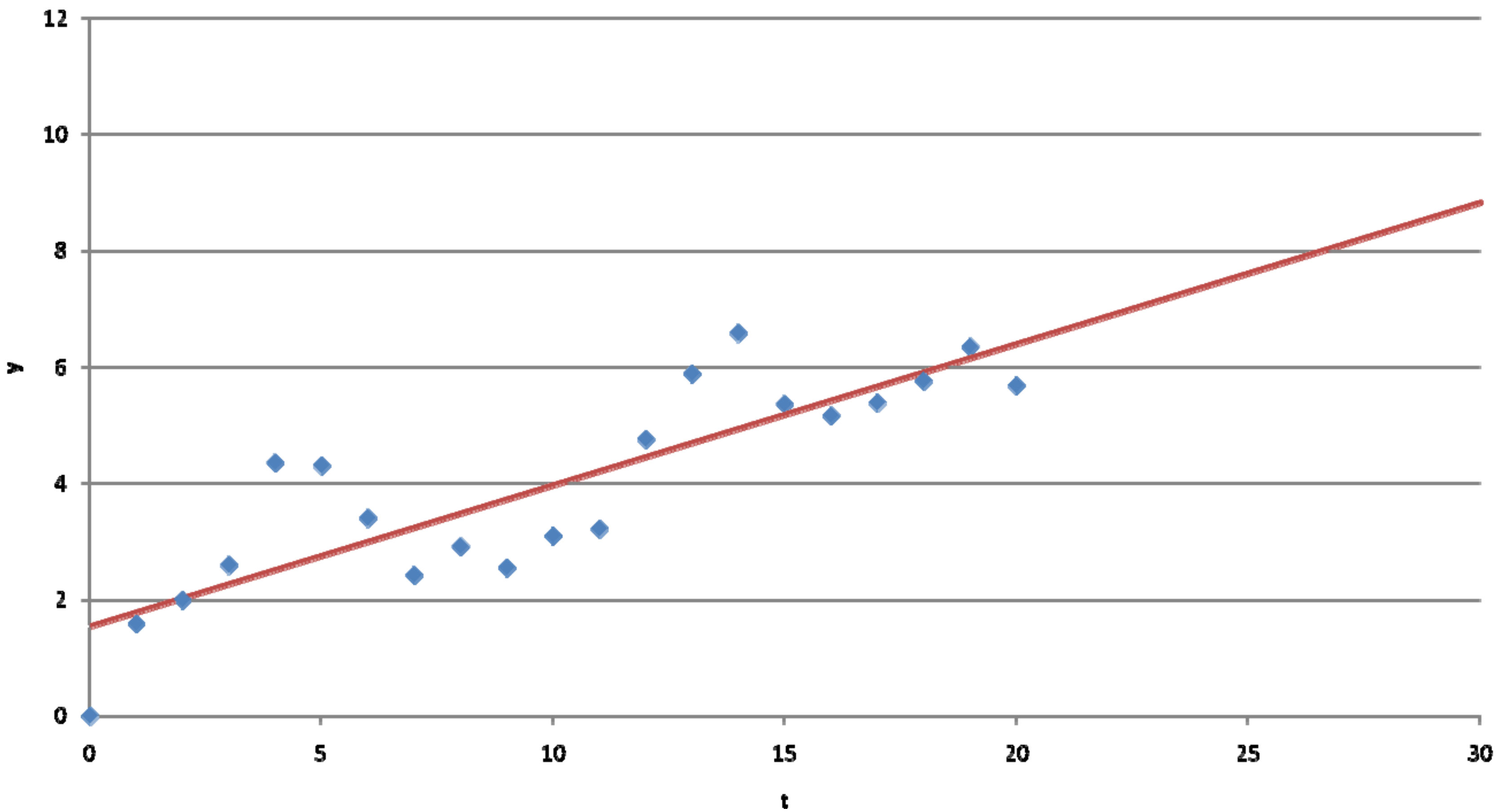
How can we forecast demand for flexible design?

- ***Forecasts need to exhibit uncertainty.***
- ***Forecasts need to address the dynamics of the uncertainty over projection horizon.***

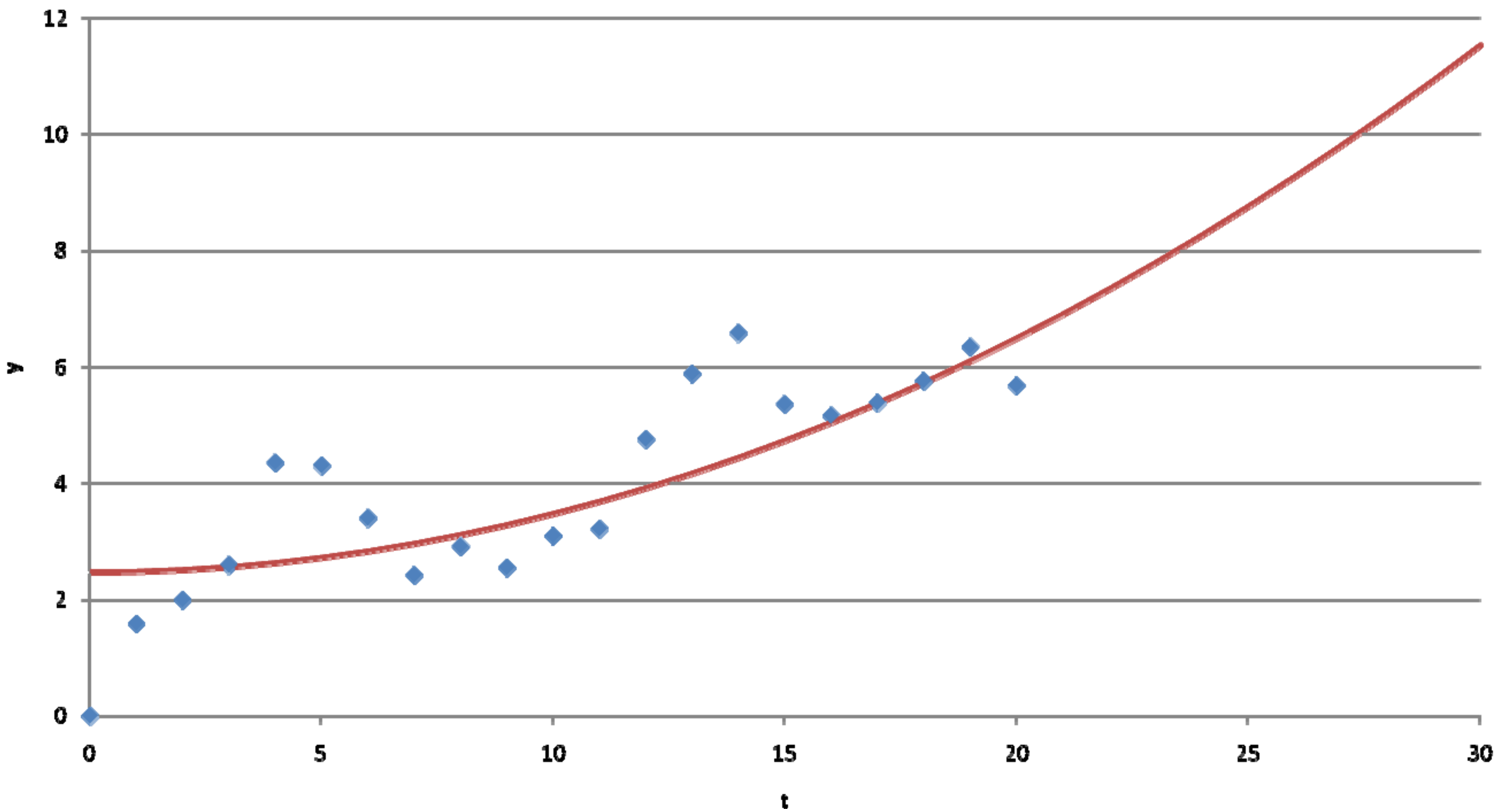
Forecasting:
Base period=[0,20] Forecasting horizon=[20,30]



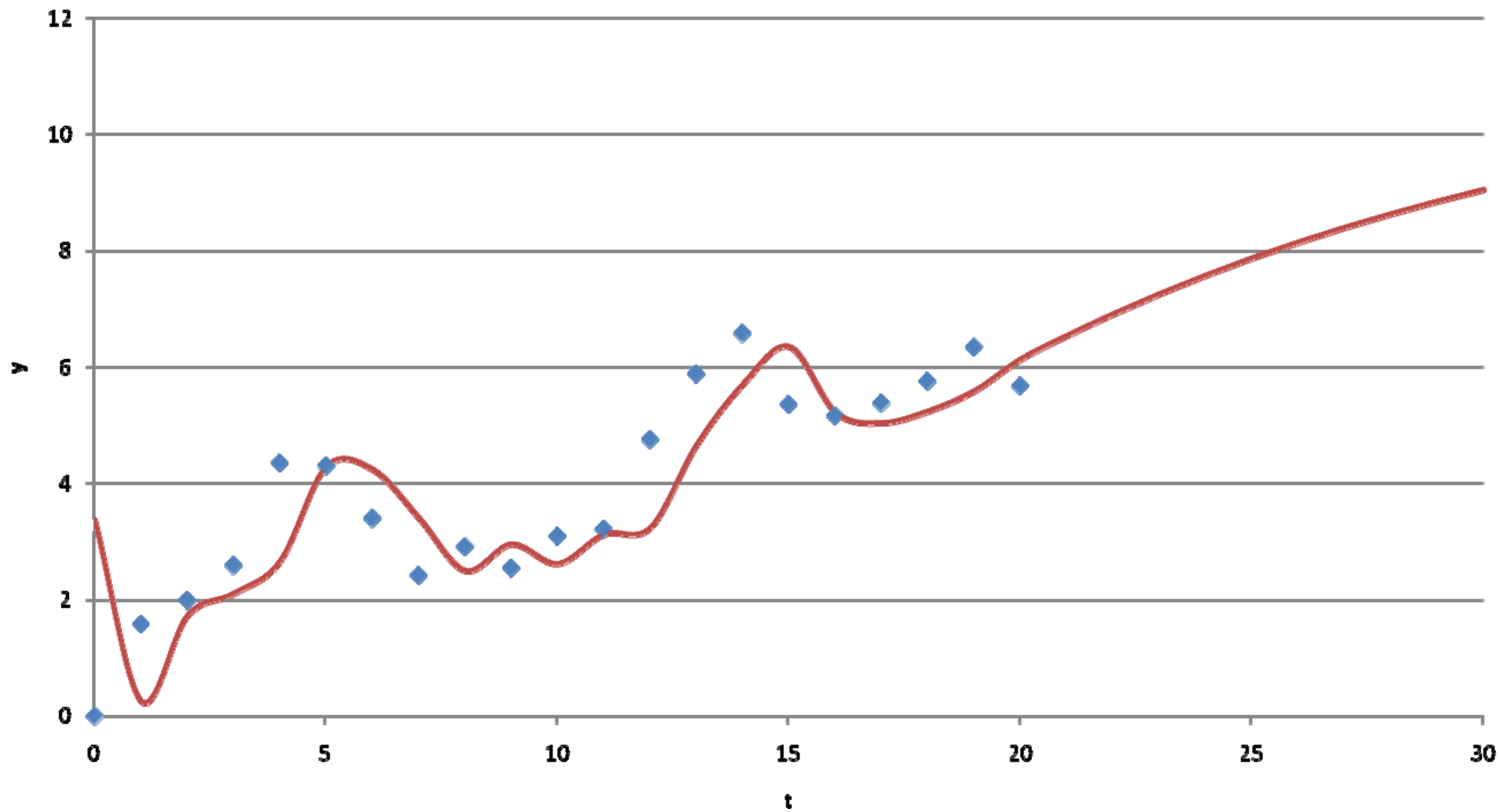
Deterministic Forecasting: linear trend fit



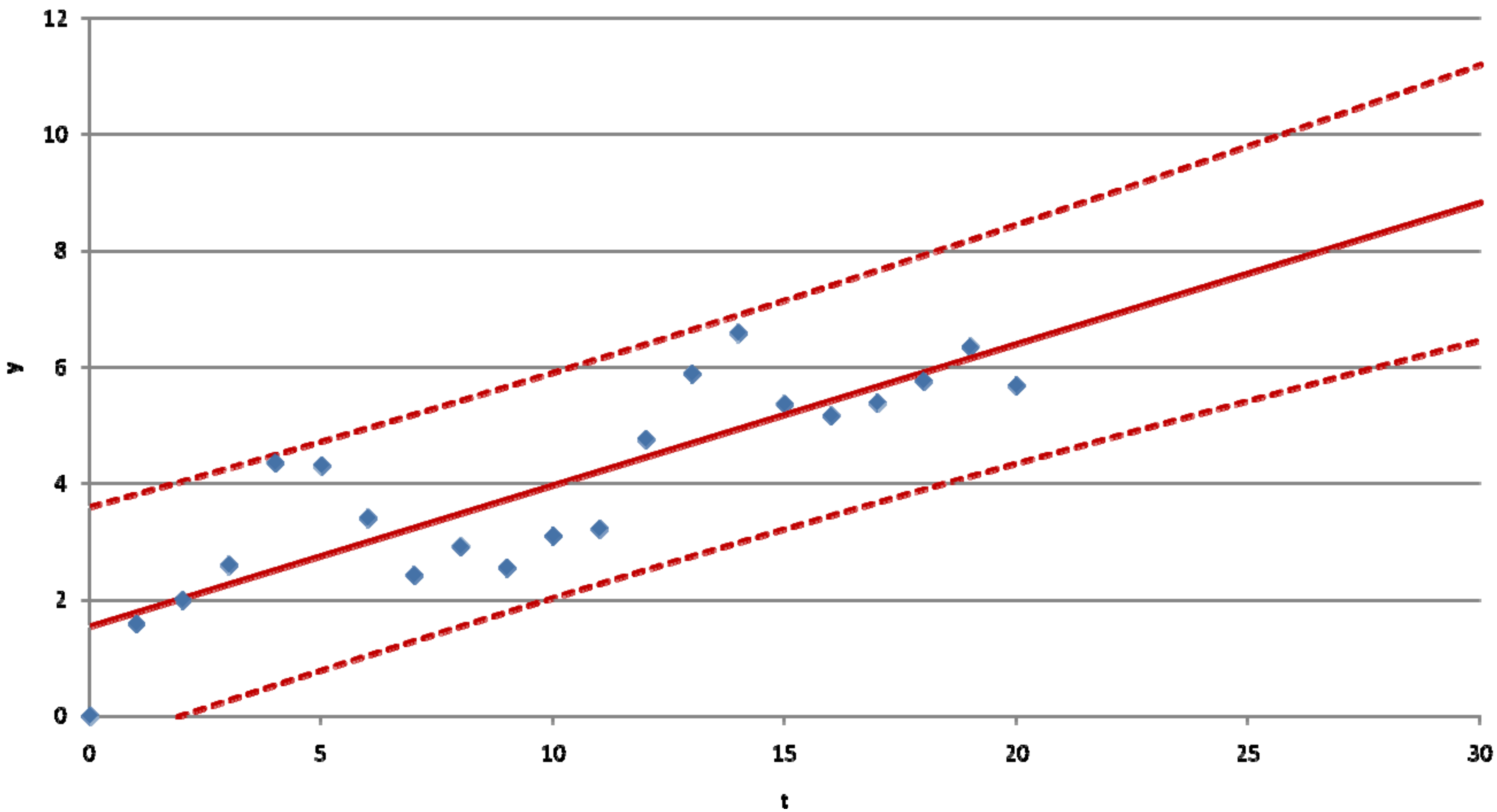
Deterministic Forecasting: quadratic trend fit



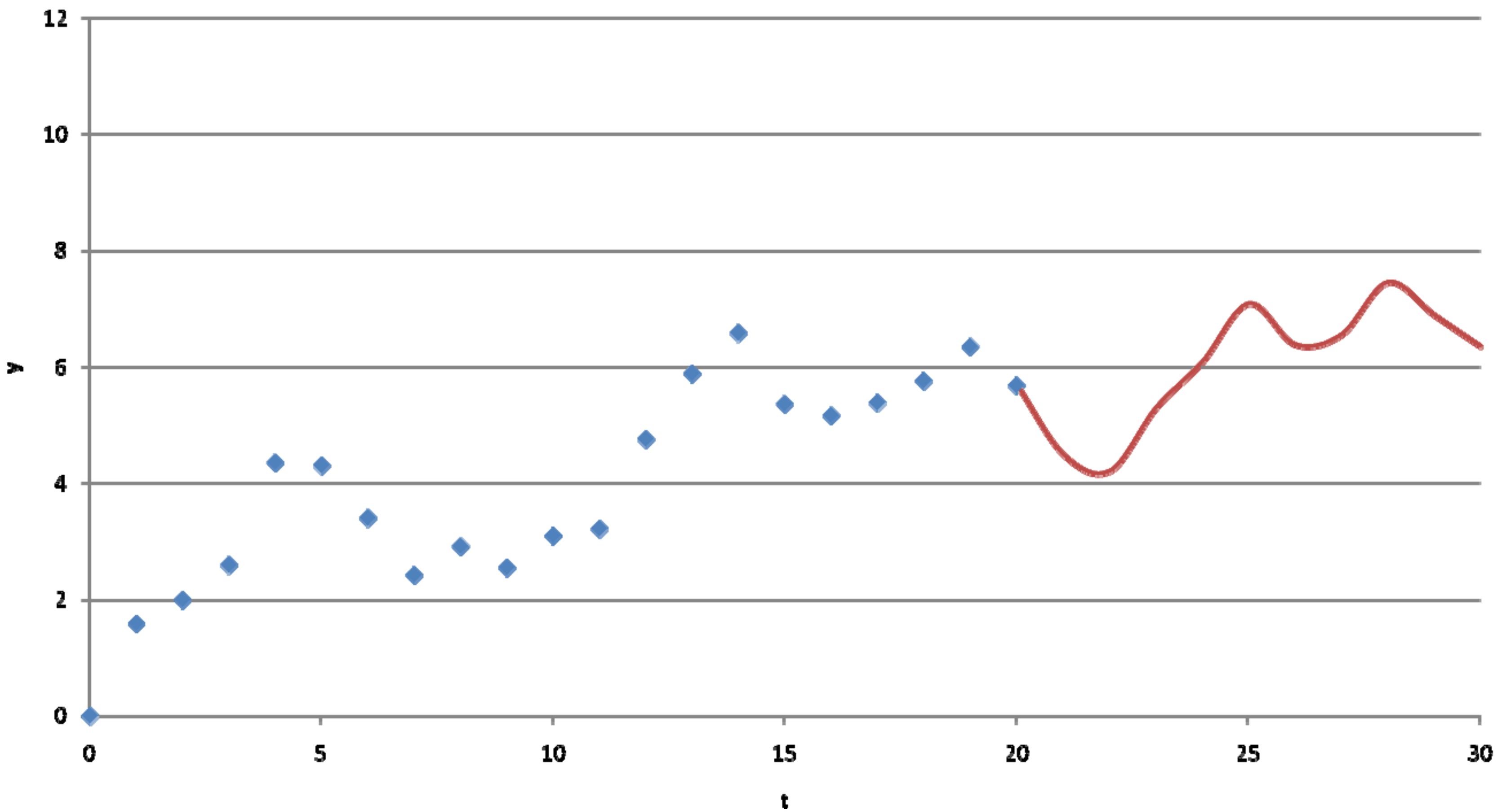
Deterministic Forecasting: AR(1)



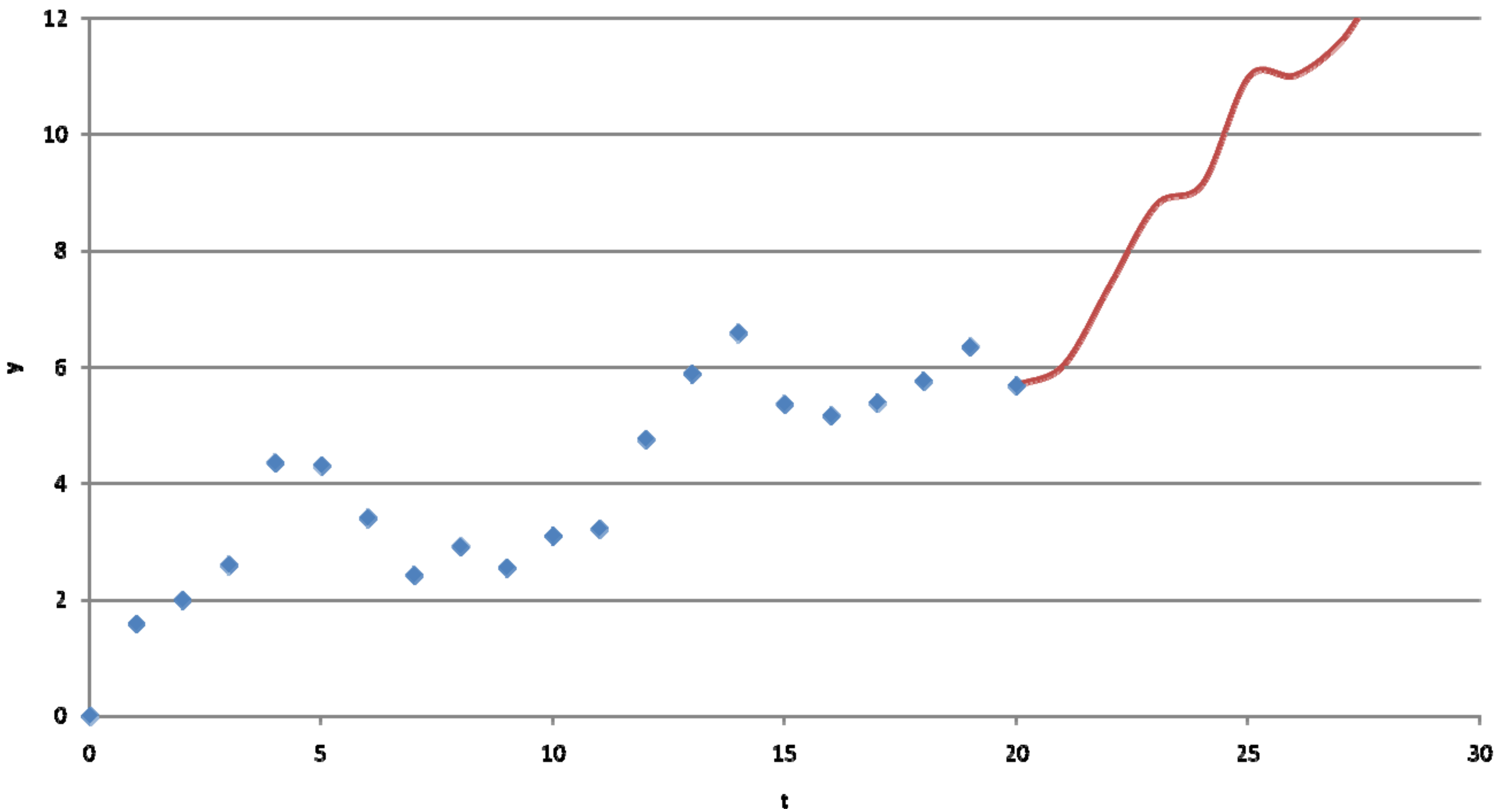
Forecasting with 95% confidence intervals
linear trend fit



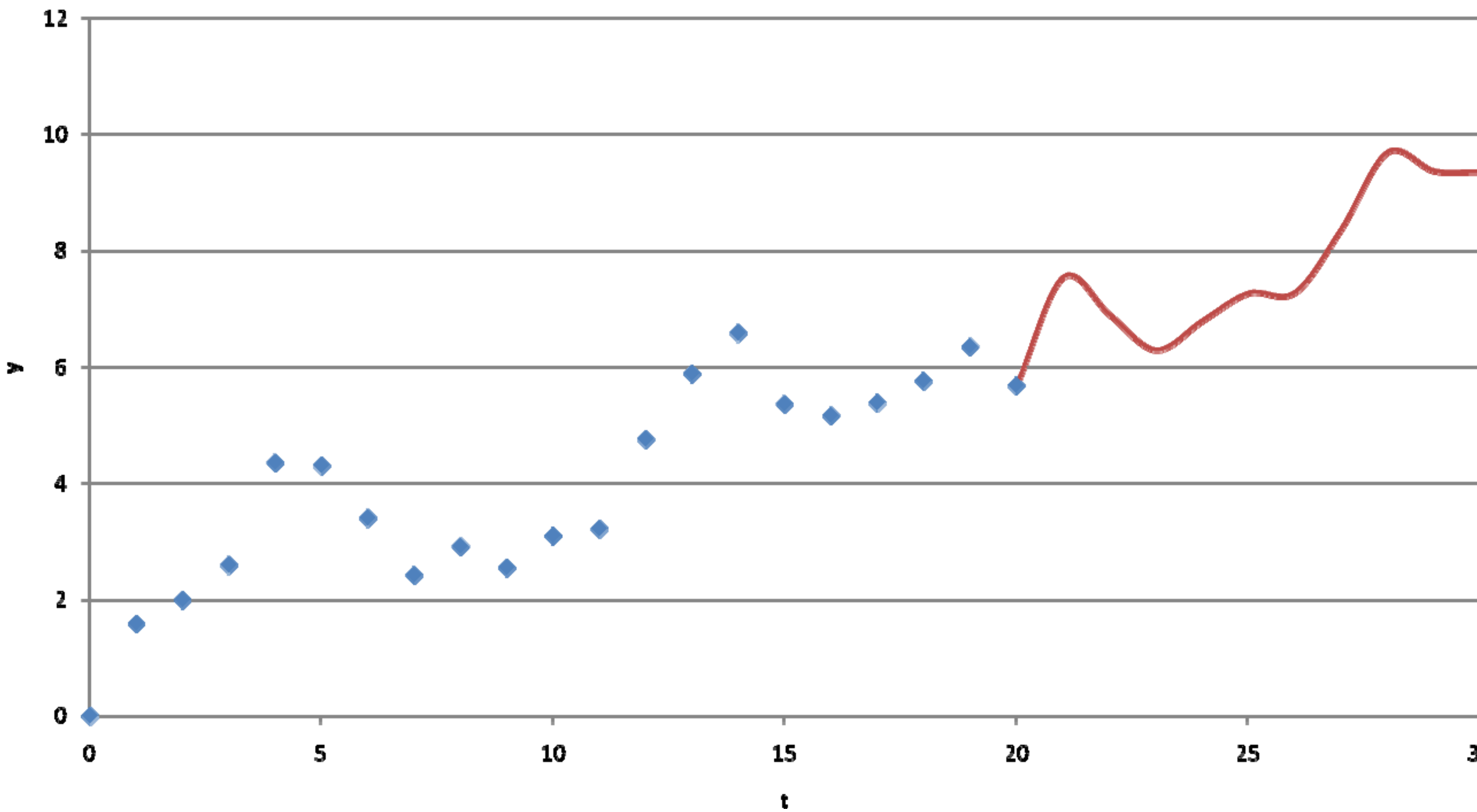
Dynamic Stochastic Forecasting



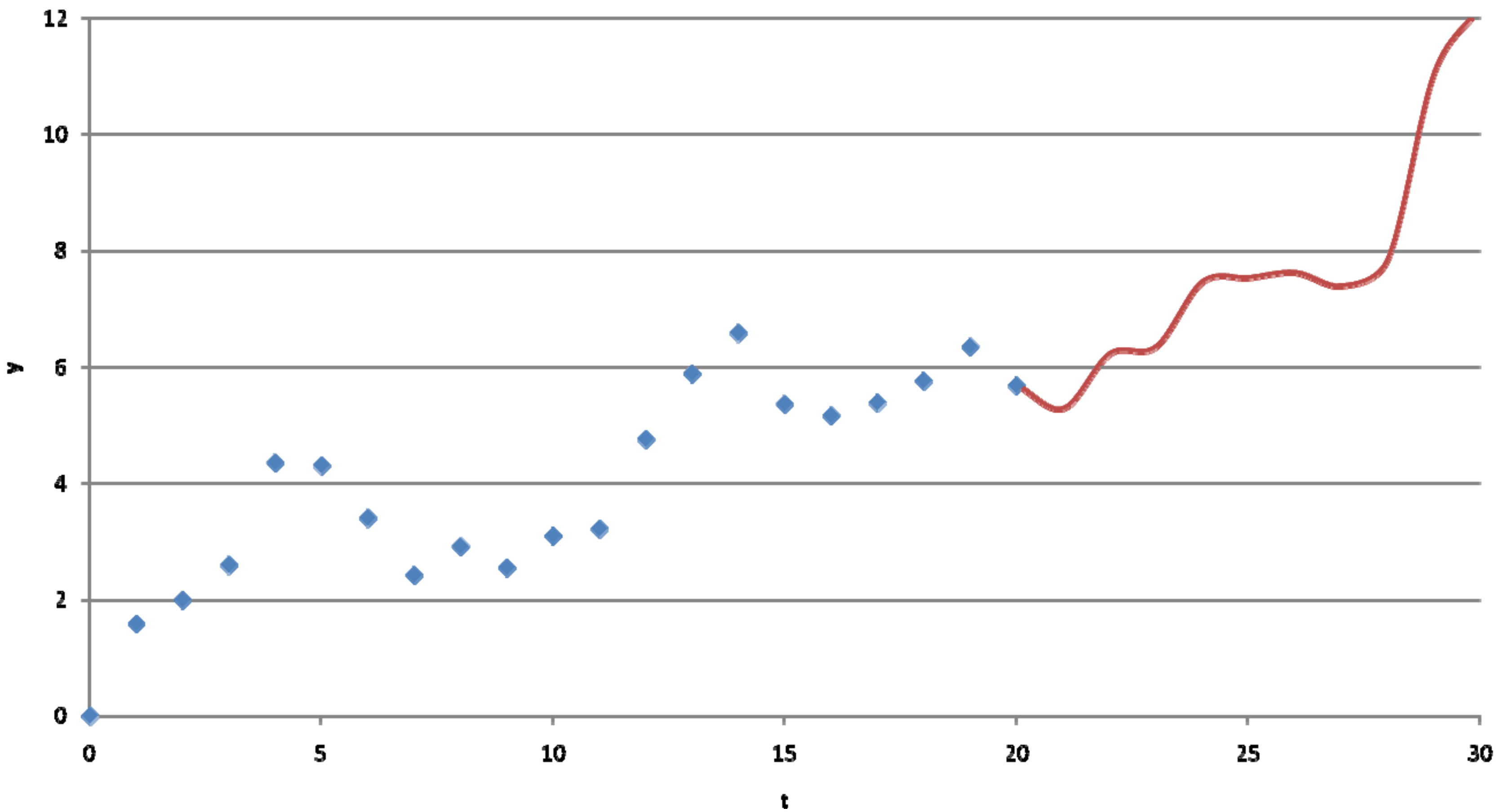
Dynamic Stochastic Forecasting



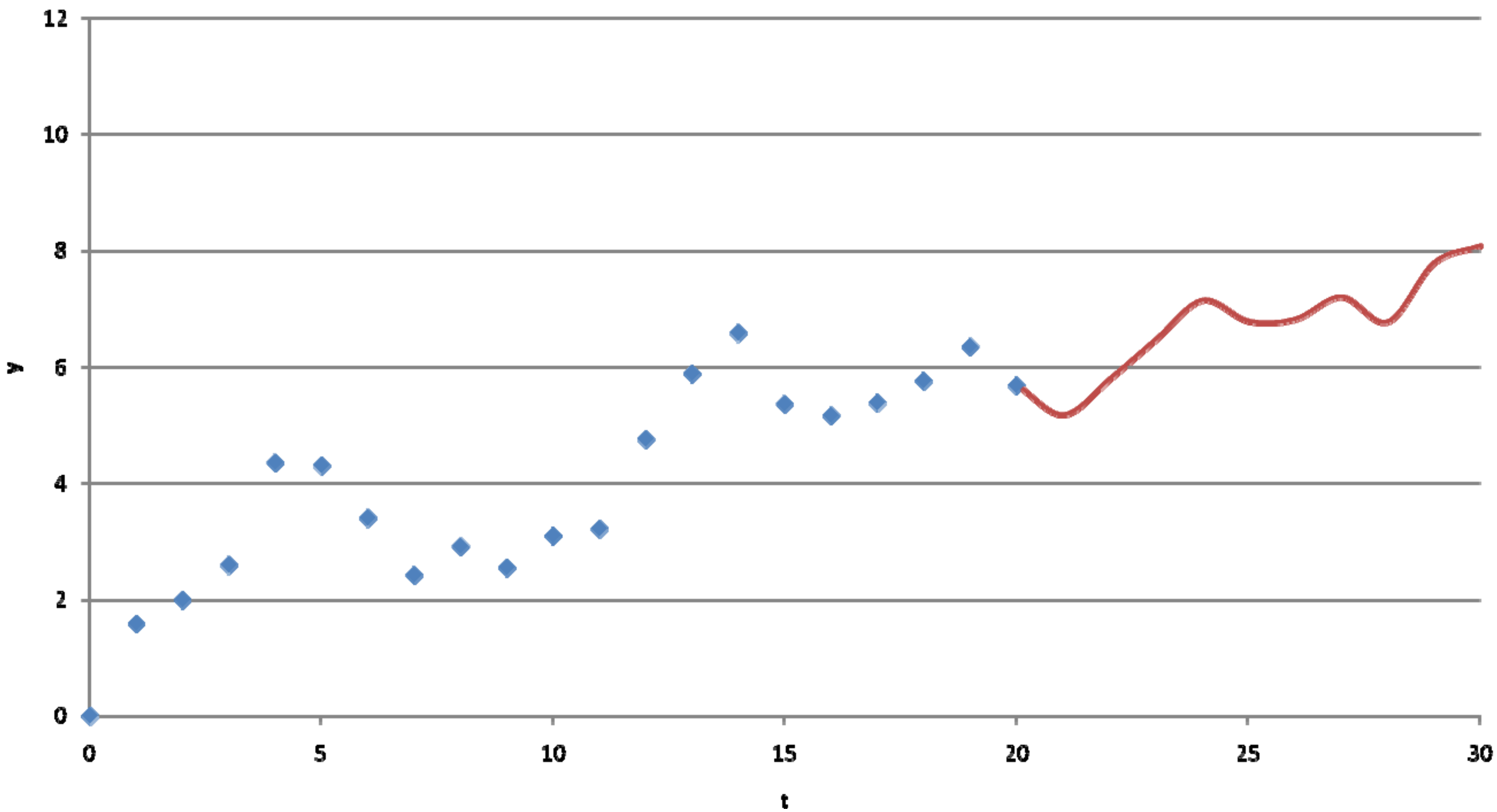
Dynamic Stochastic Forecasting



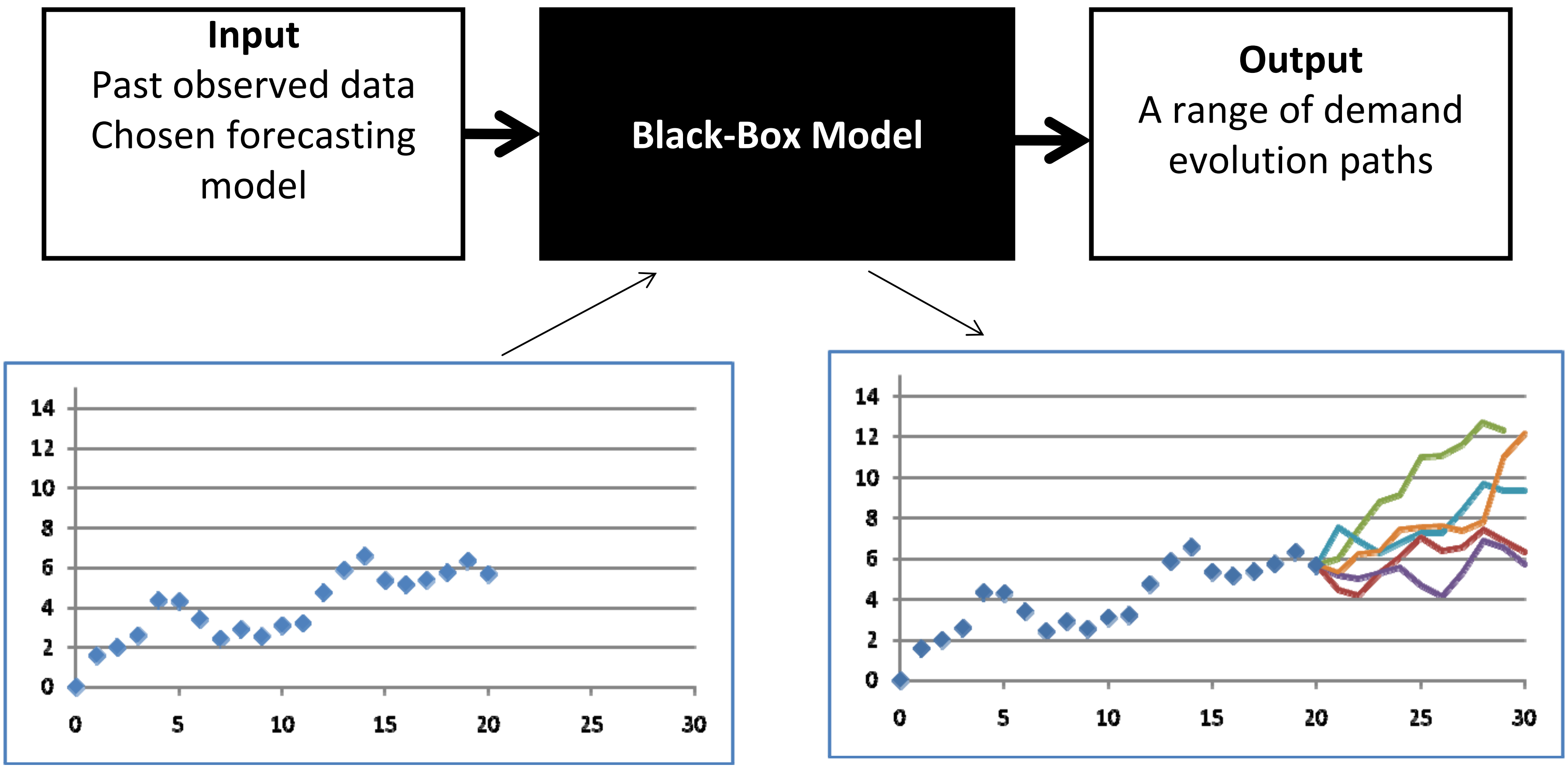
Dynamic Stochastic Forecasting



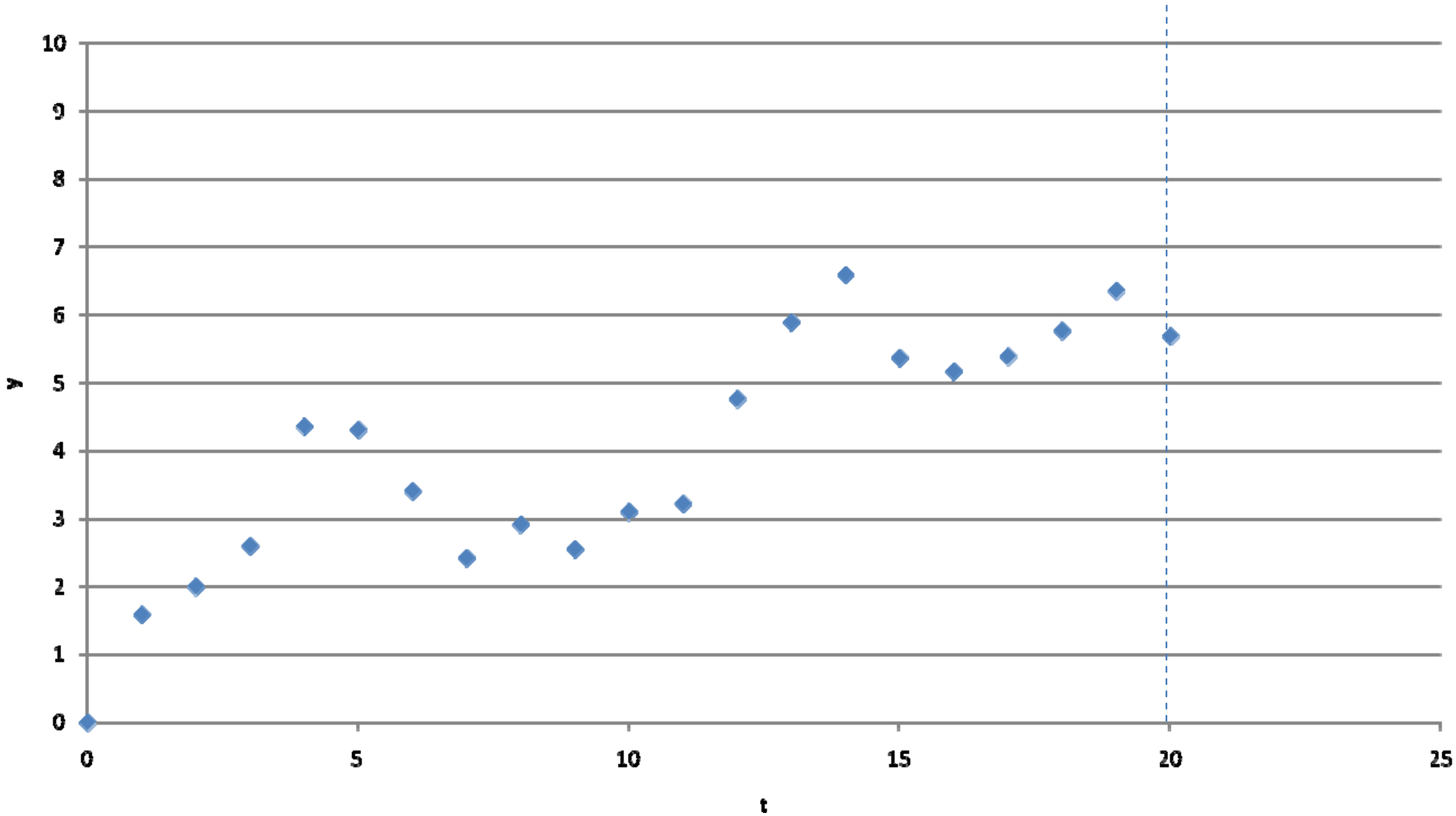
Dynamic Stochastic Forecasting



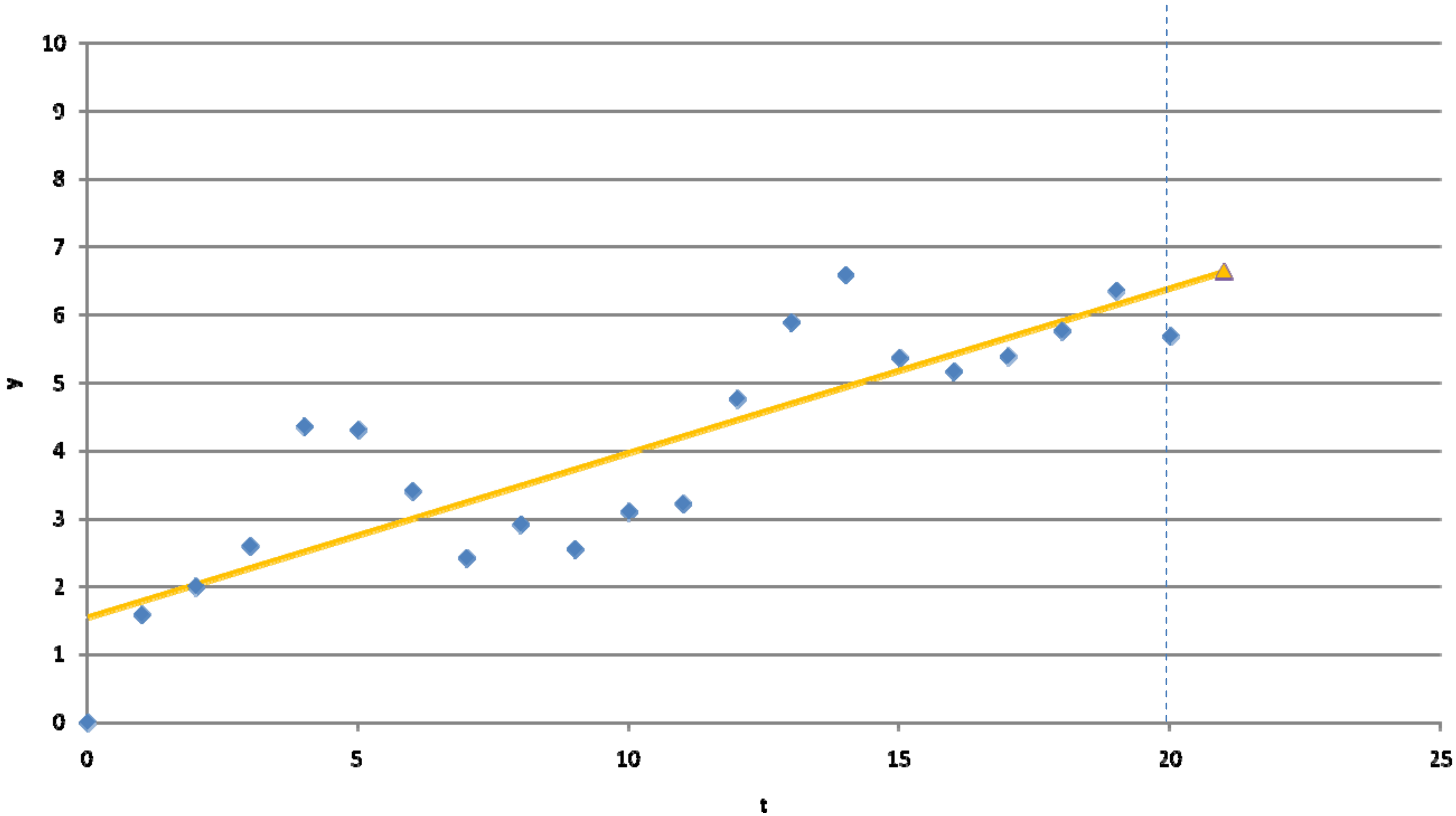
Forecasting for flexible systems



Proposed forecasting framework with linear fitting model



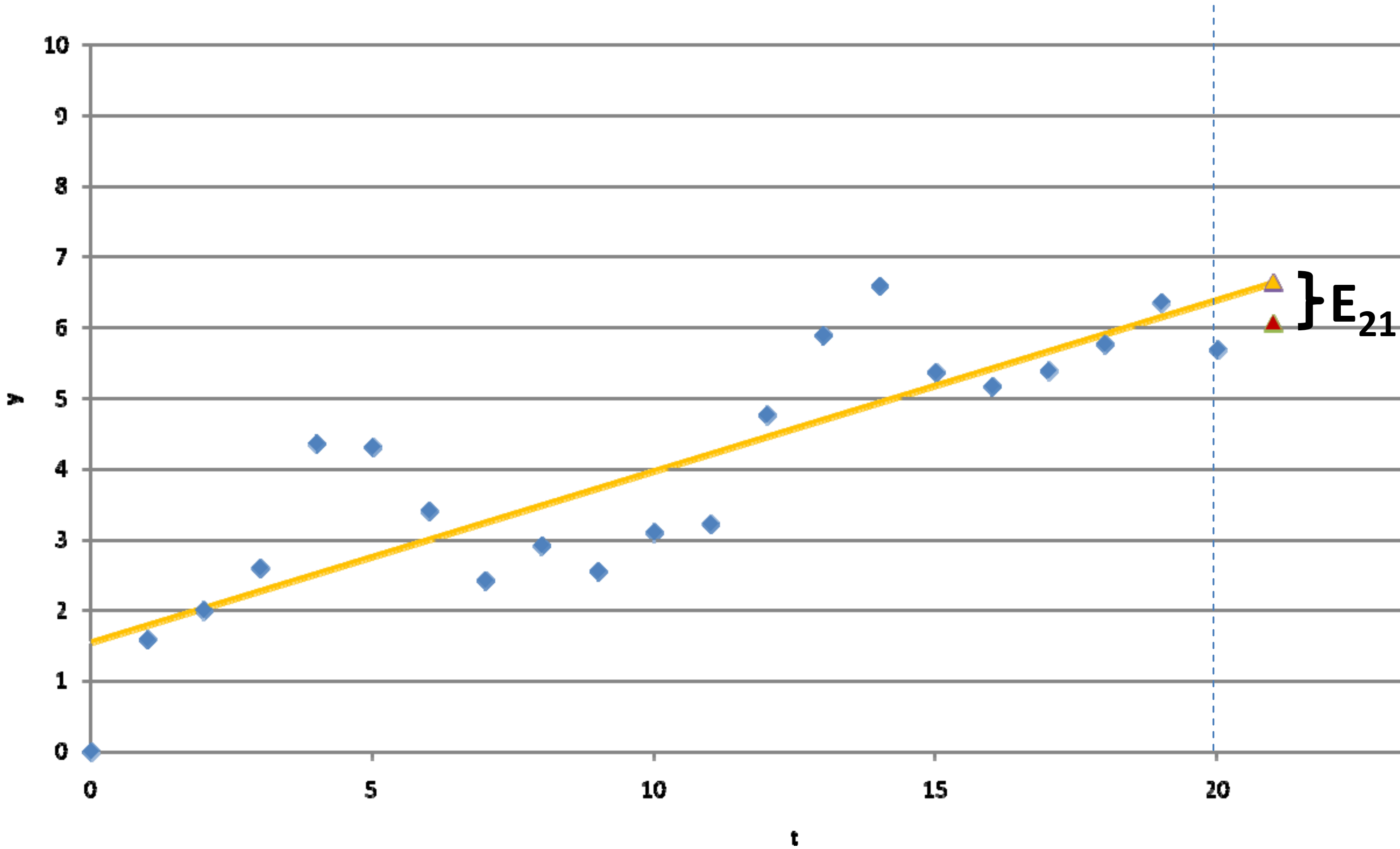
Stochastic forecast
= Deterministic forecast + stochastic forecast error



$$y_{t+1} = \hat{y}_{t+1} + E_{t+1} = \hat{\alpha} + \hat{\beta}(t+1) + E_{t+1} \quad where\ E_{t+1} \sim N(0, \sigma^2)$$

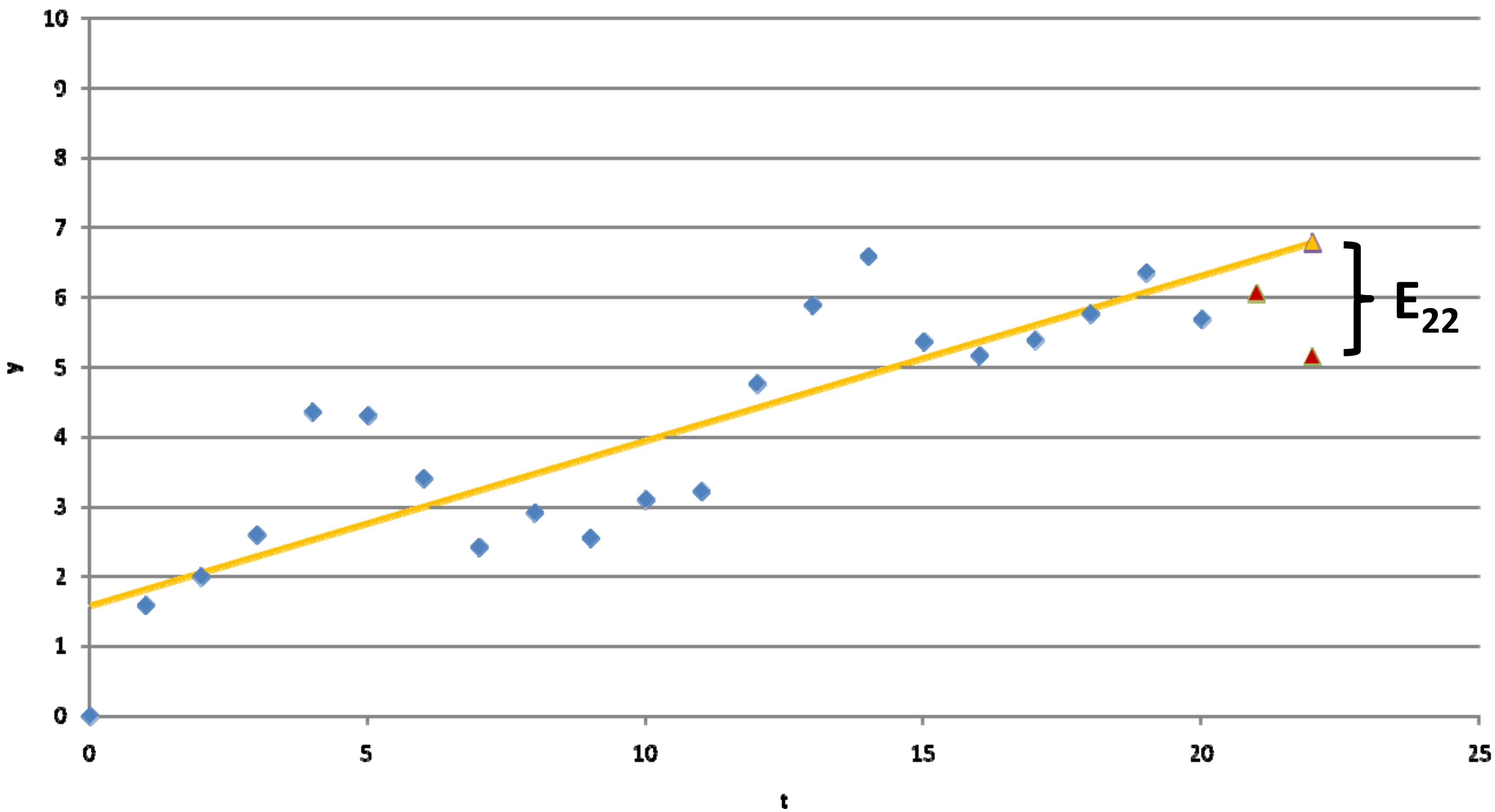
Stochastic forecast

= Deterministic forecast + stochastic forecast error



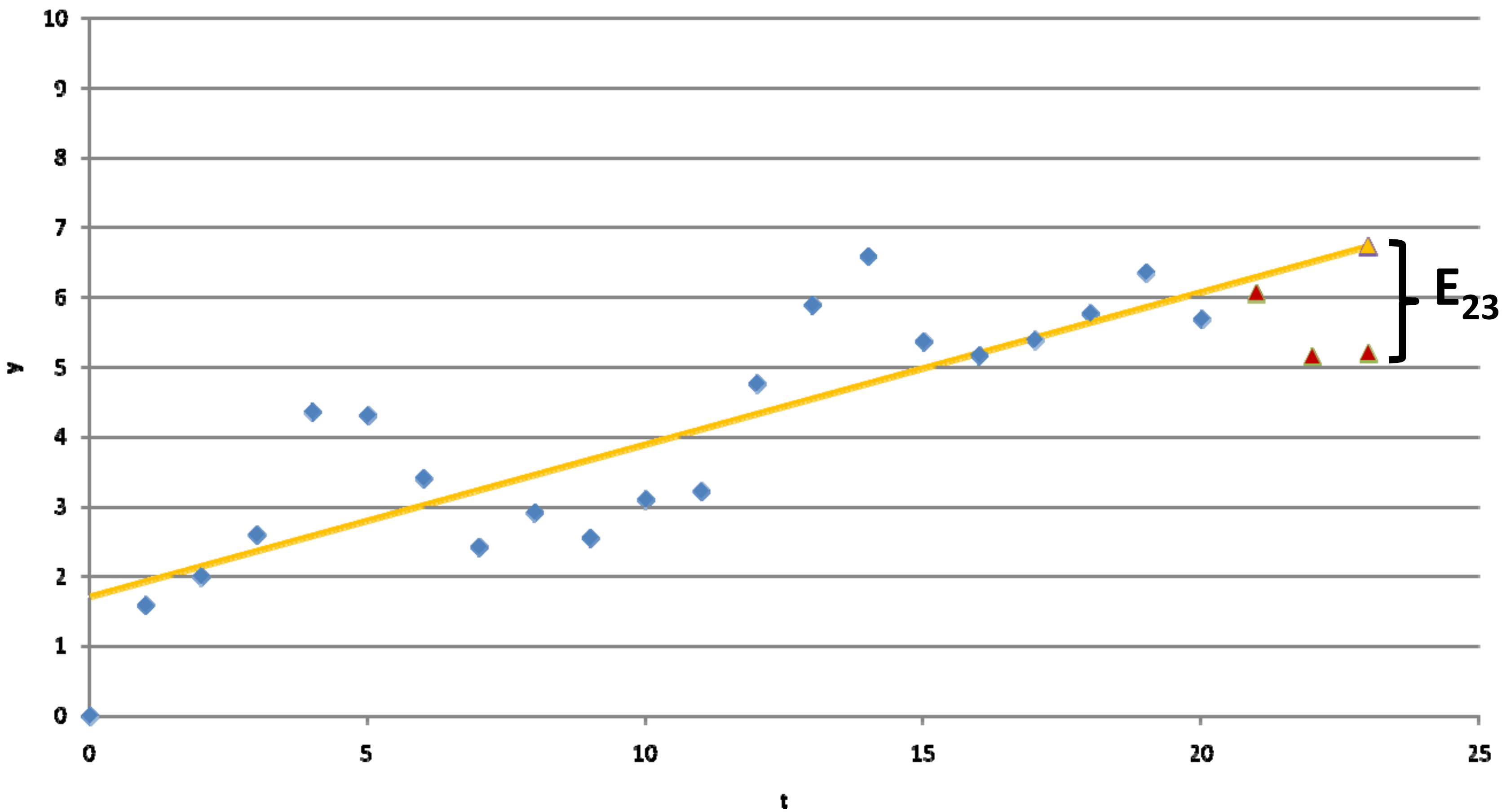
$$y_{t+1} = \hat{y}_{t+1} + E_{t+1} = \hat{\alpha} + \hat{\beta}(t+1) + E_{t+1} \quad where\ E_{t+1} \sim N(0, \sigma^2)$$

Stochastic forecast
= Deterministic forecast + stochastic forecast error



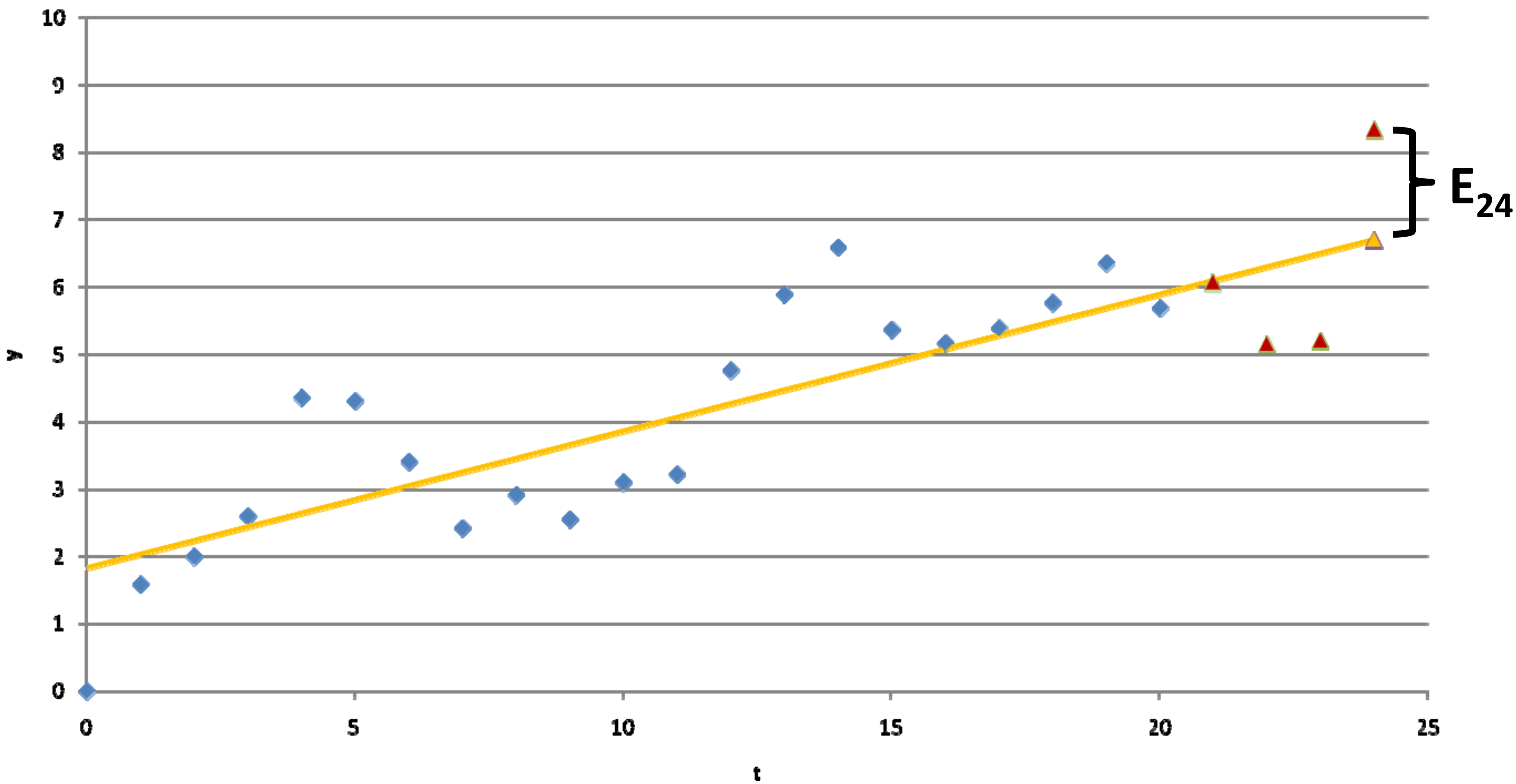
$$y_{t+1} = \hat{y}_{t+1} + E_{t+1} = \hat{\alpha} + \hat{\beta}(t+1) + E_{t+1} \quad where\ E_{t+1} \sim N(0, \sigma^2)$$

Stochastic forecast
= Deterministic forecast + stochastic forecast error



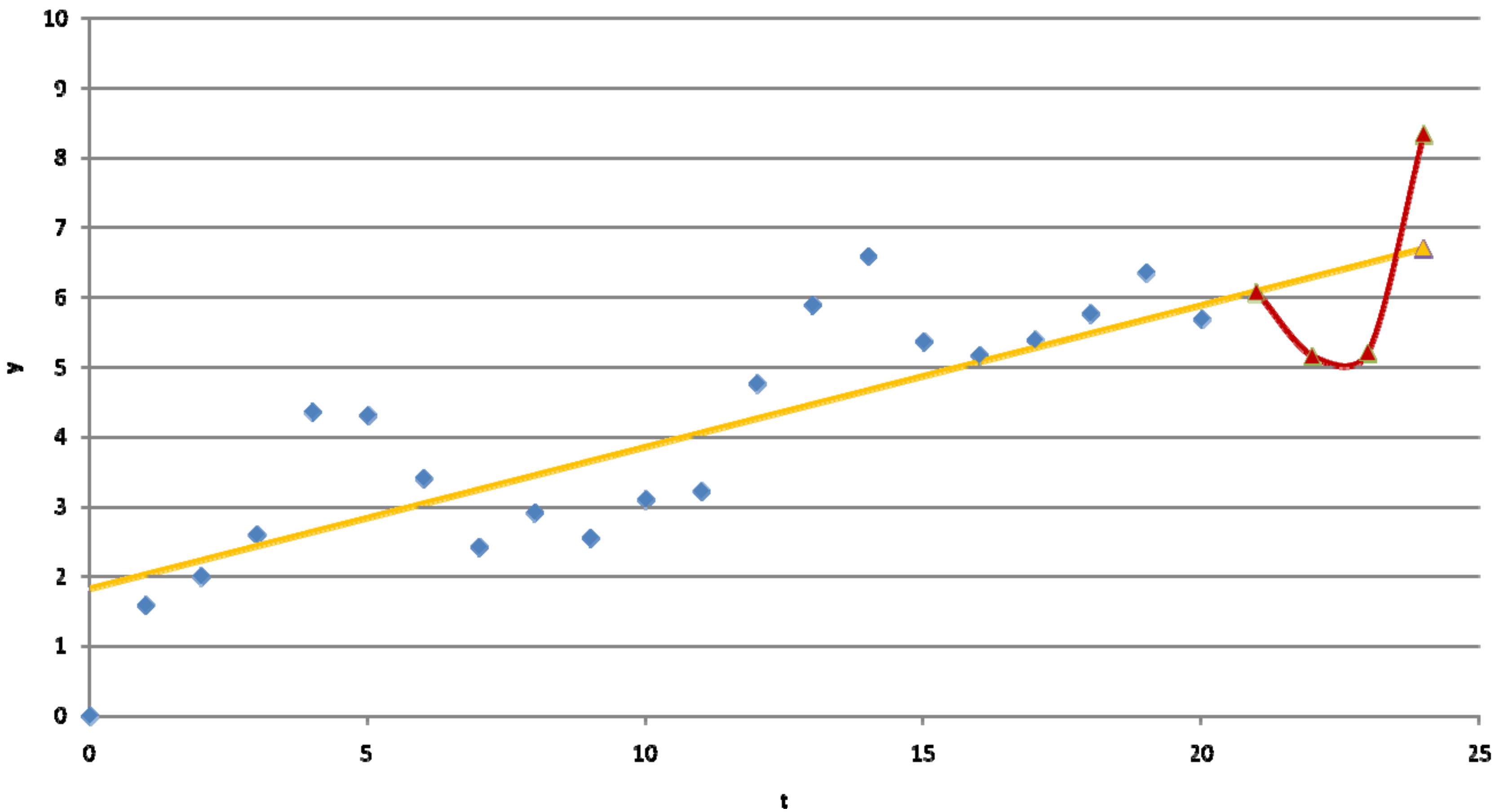
$$y_{t+1} = \hat{y}_{t+1} + E_{t+1} = \hat{\alpha} + \hat{\beta}(t+1) + E_{t+1} \quad where\ E_{t+1} \sim N(0, \sigma^2)$$

Stochastic forecast
= Deterministic forecast + stochastic forecast error



$$y_{t+1} = \hat{y}_{t+1} + E_{t+1} = \hat{\alpha} + \hat{\beta}(t+1) + E_{t+1} \quad where\ E_{t+1} \sim N(0, \sigma^2)$$

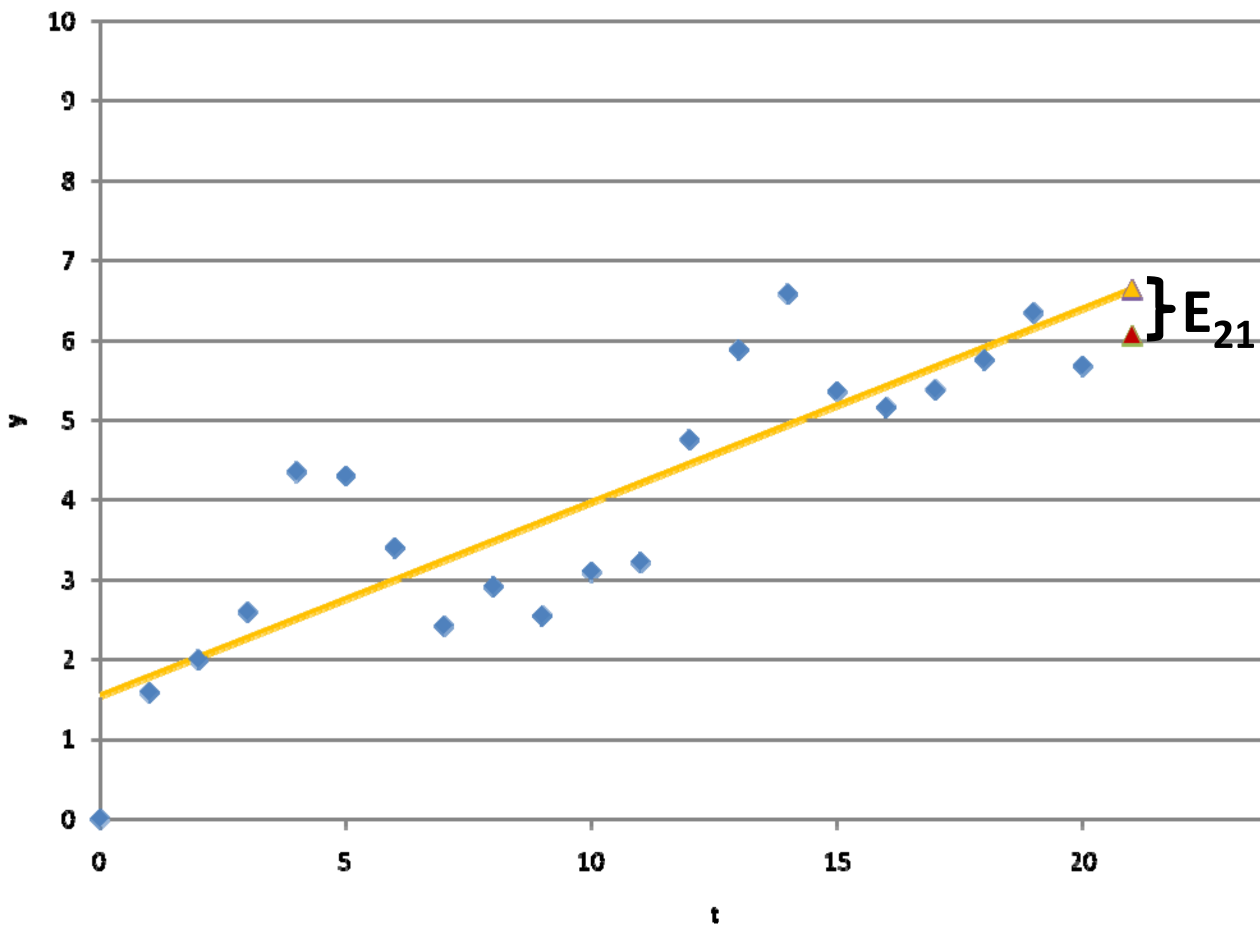
Stochastic forecast
= Deterministic forecast + stochastic forecast error



Adaptive Trend Fitting with Stochastic Forecast Errors

How can we assume a distribution of future forecast error, E_{t+1}?

$$E_{t+1} = y_{t+1} - \hat{y}_{t+1}$$



How can we assume a distribution of E_{t+1}?

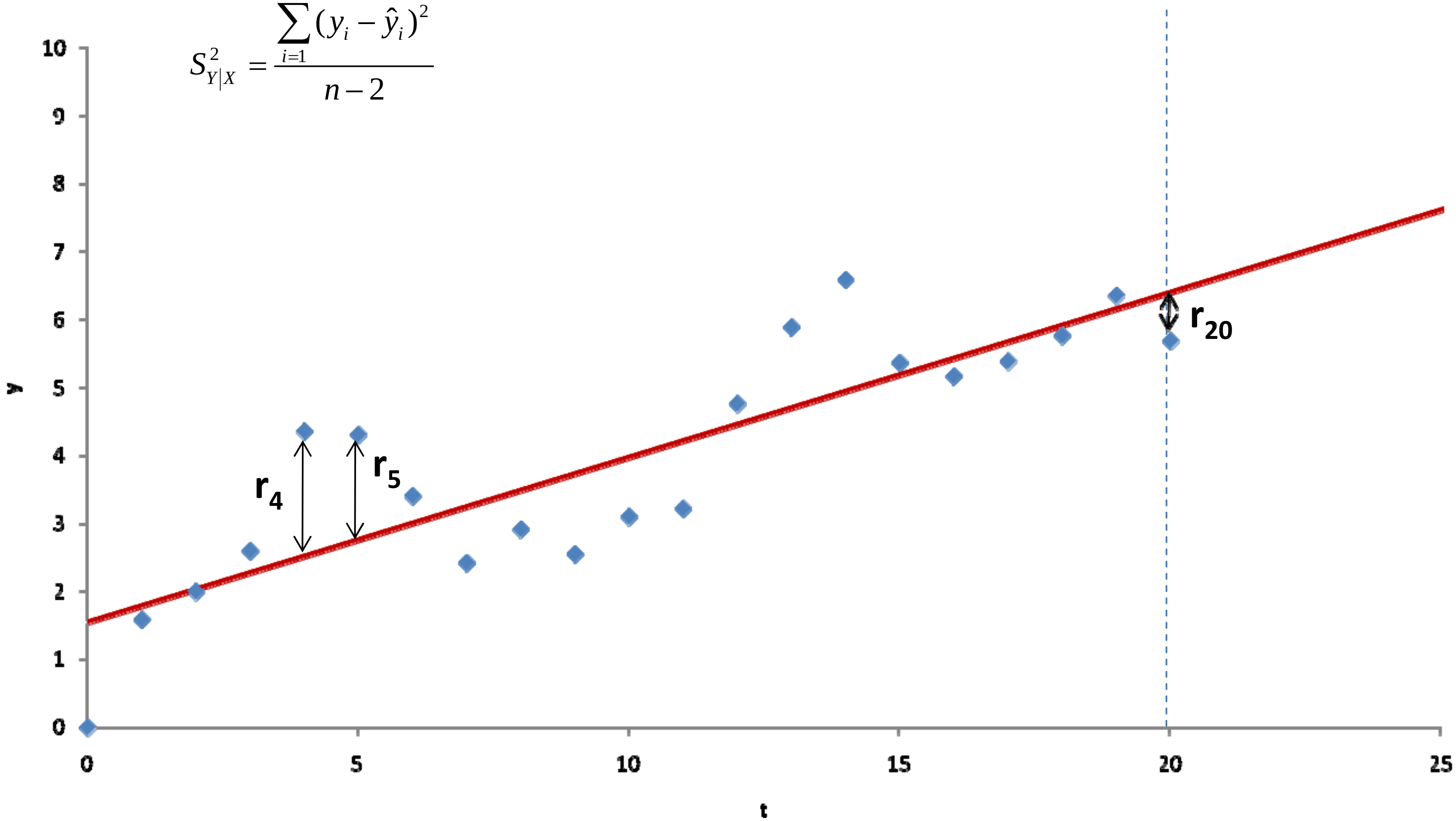
$$E_{t+1} = y_{t+1} - \hat{y}_{t+1}$$

- based on the in-sample estimation of the fitting method (residuals)

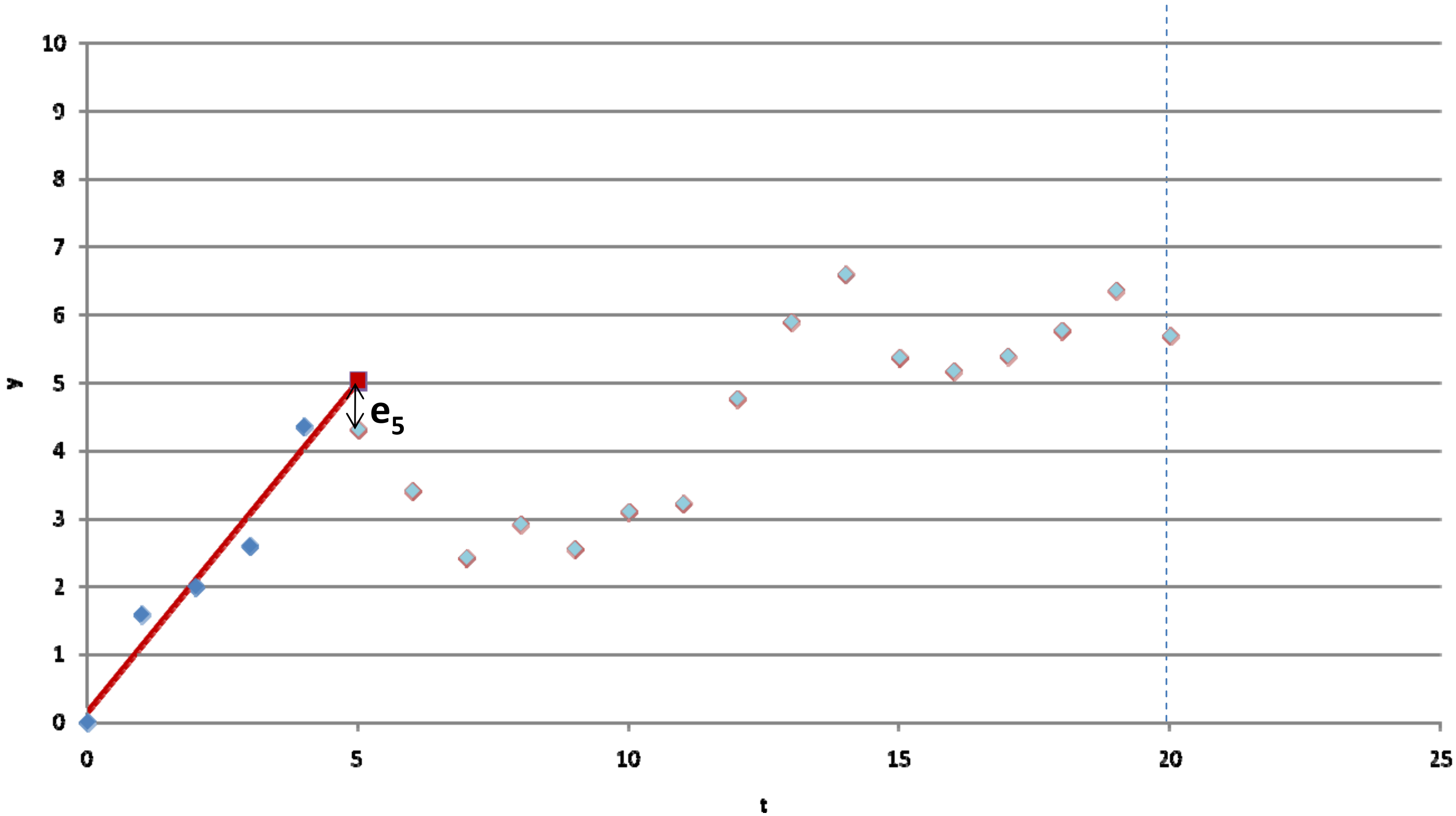
- based on the out-of-sample forecast evaluation of the fitting method (historical forecast errors)

Based on in-sample estimation

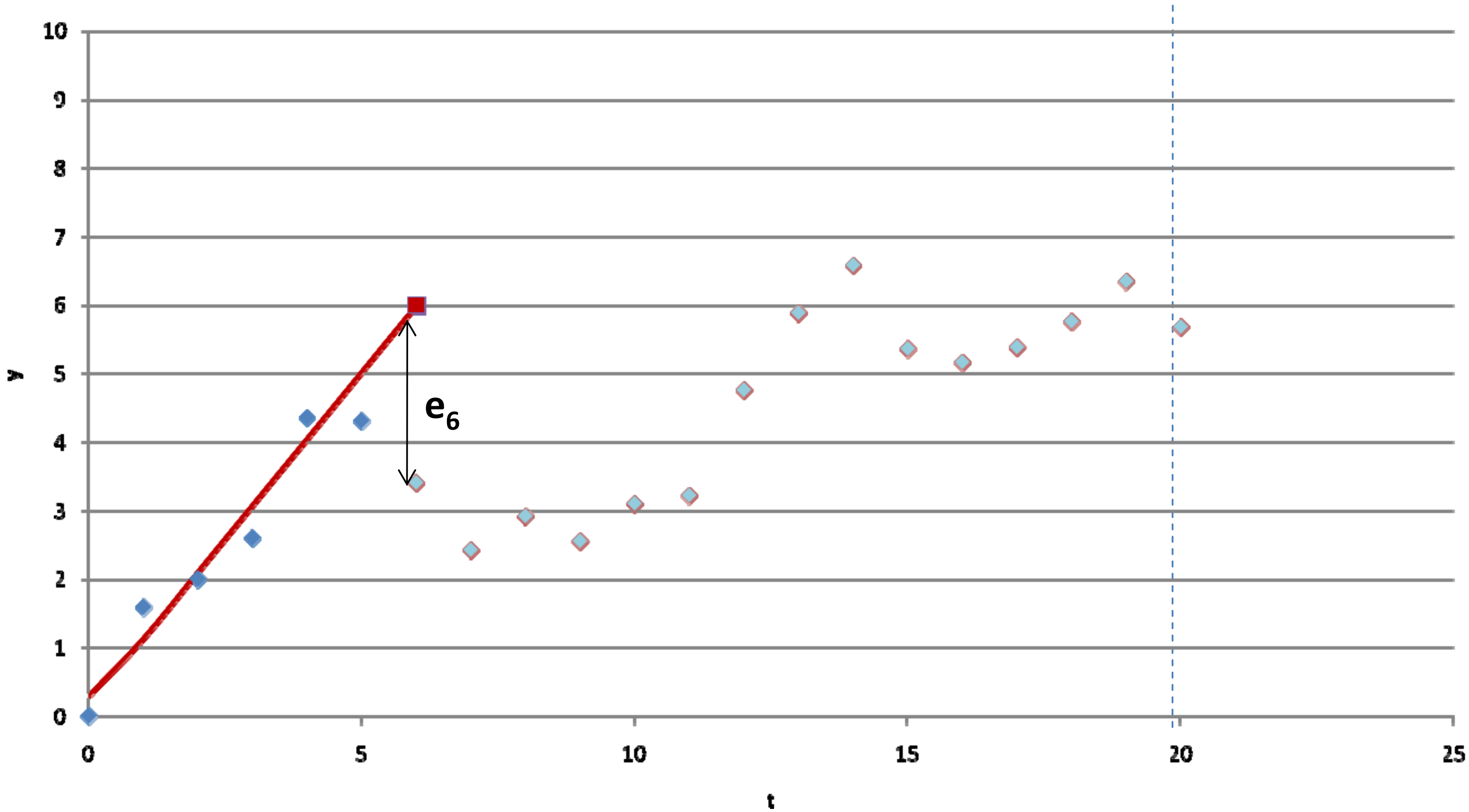
$$S_{Y|X}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$$



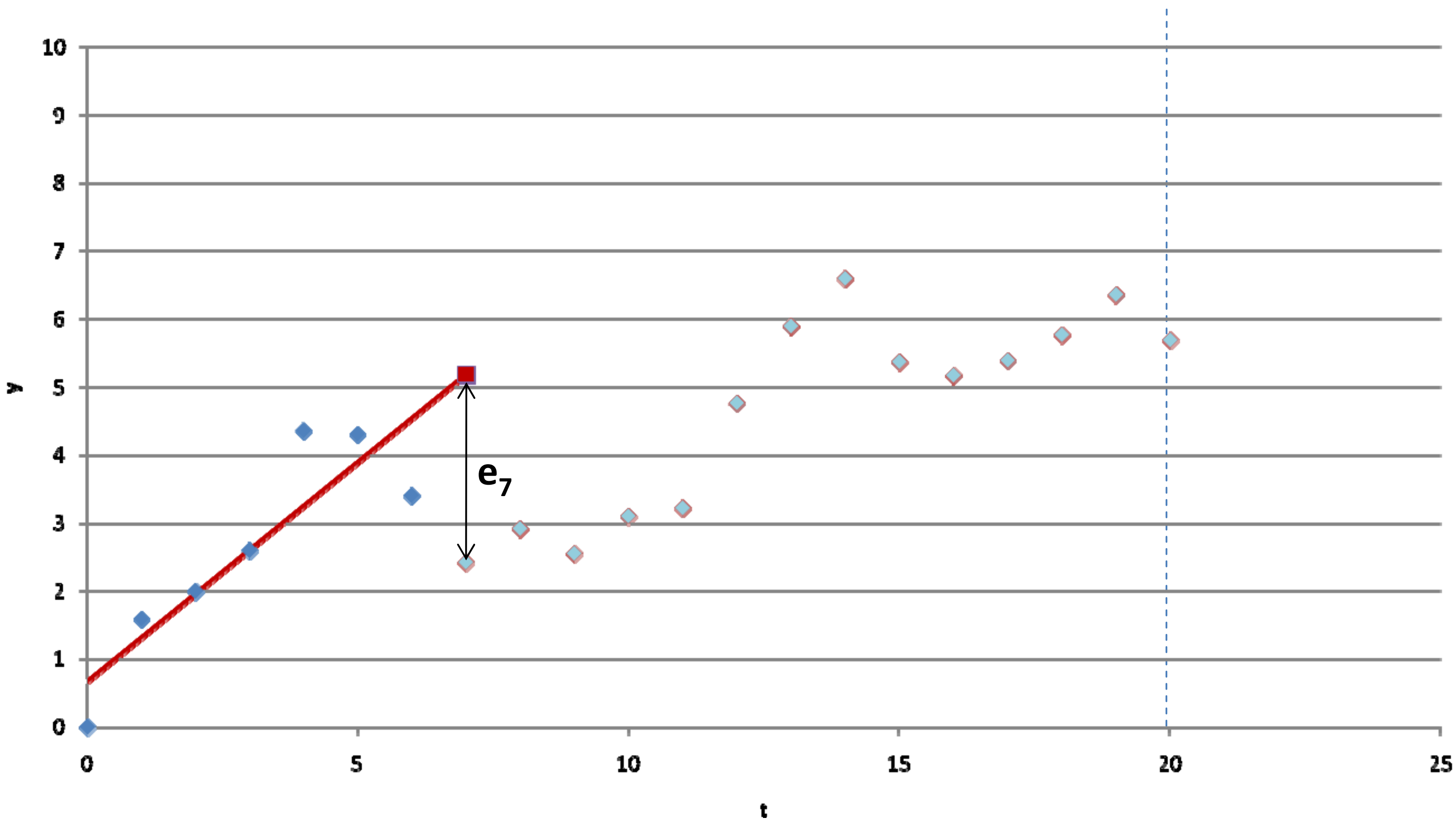
Based on out-of-sample forecast evaluation



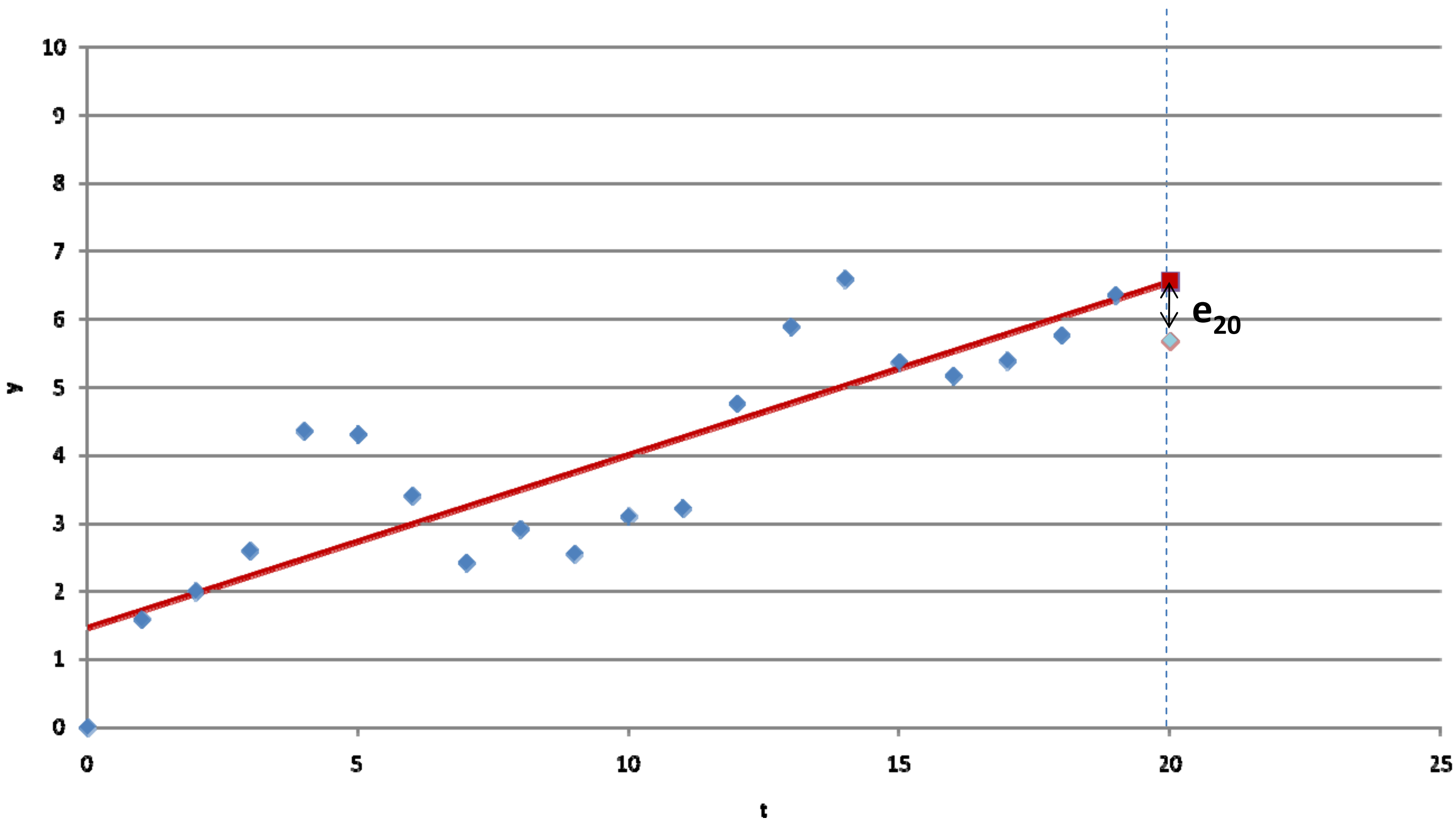
Based on out-of-sample forecast evaluation



Based on out-of-sample forecast evaluation



Based on out-of-sample forecast evaluation



How can we assume a distribution of future forecast error?

– The distribution of future forecast error is assumed by the analysis of historical forecast errors.

How can we assume a distribution of future forecast error?

– The distribution of future forecast error is assumed by the analysis of historical forecast errors.

– We cannot simply sample from the historical forecast errors since the recurrent forecasts rely on the same general method and share historical data. Forecast errors are dependent and correlated.

– We need to deal with the non-stationary characteristics of forecast errors.

Statistical properties of the proposed framework

- To illustrate the statistical properties of the proposed framework, we investigate its probabilistic behaviour for the case that the preferred projection method is linear trend fitting and the data generation process is a random walk.

Statistical properties of the proposed framework

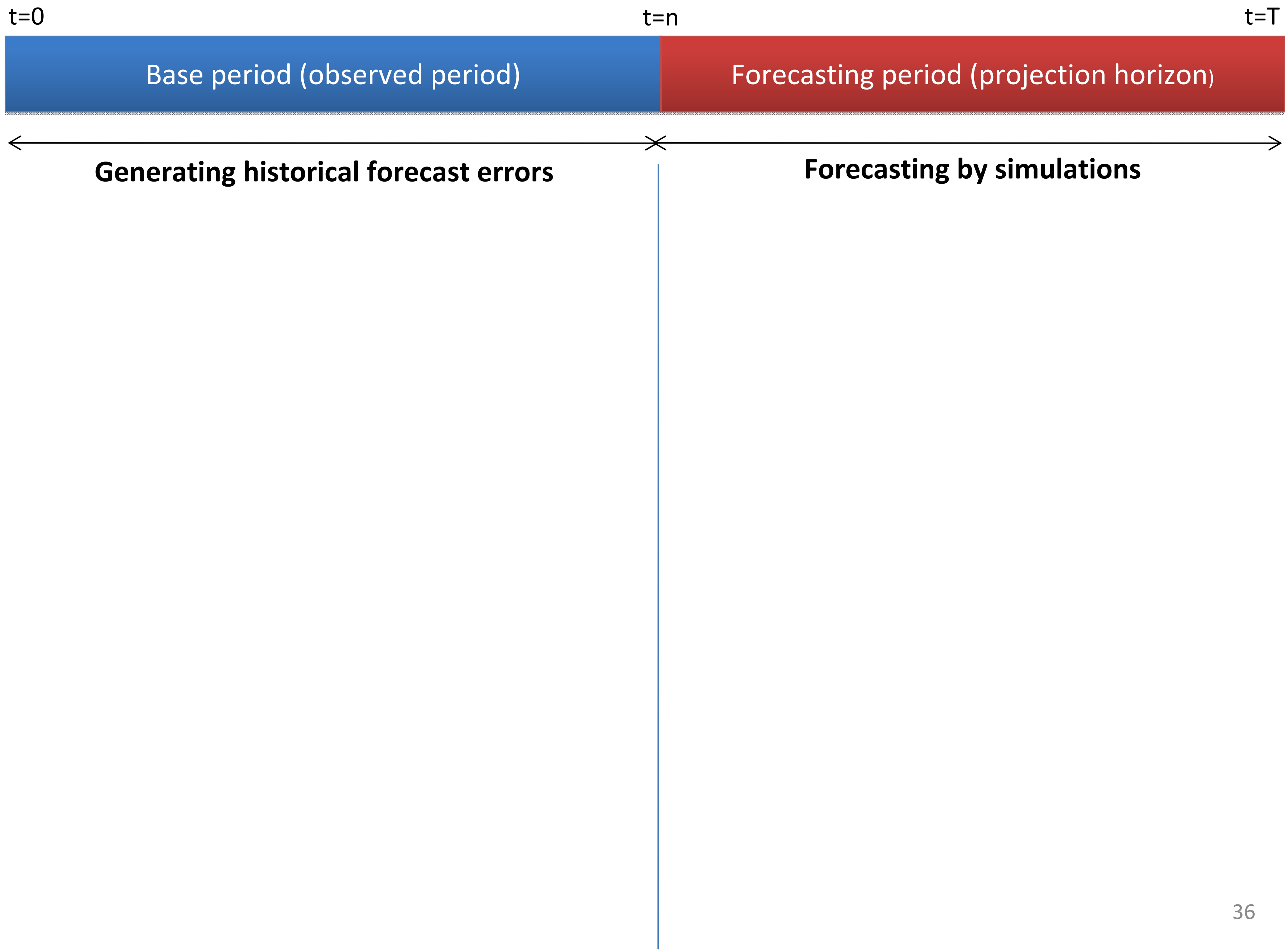
- To illustrate the statistical properties of the proposed framework, we investigate its probabilistic behaviour for the case that the preferred projection method is linear trend fitting and the data generation process is a random walk

- Actual Process is assumed by an additive random walk process.

$$y_0 = 0$$

$$y_{t+1} = y_t + \varepsilon_t \sqrt{\Delta t} = \sum_{i=0}^{t} \varepsilon_i$$

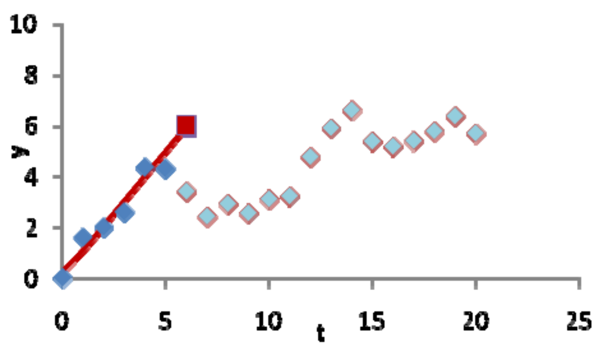
where ε_i is i.i.d and ~$N(0, \sigma_\varepsilon{}^2)$


t=0
t=n
t=T
Base period (observed period)
Forecasting period (projection horizon)
Generating historical forecast errors
Forecasting by simulations

t=0 | t=n | t=T

Base period (observed period) | **Forecasting period (projection horizon)**

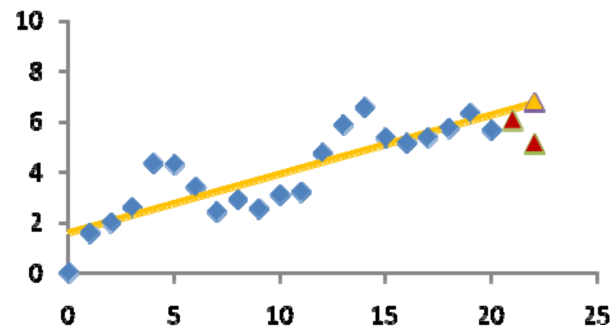
Generating historical forecast errors



$$\hat{y}_{t+1} = f(t+1; y_0, y_1, \ldots, y_t) = a_t + b_t(t+1)$$

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1}$$

Forecasting by simulations



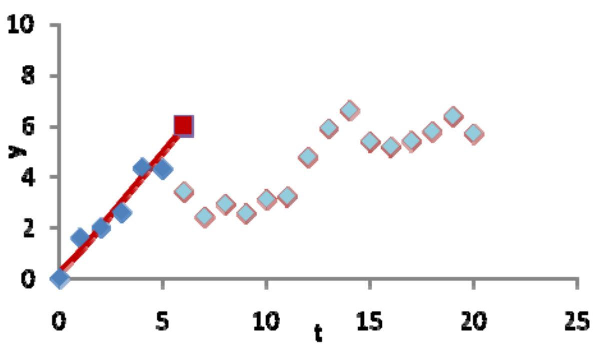
$$\tilde{y}_{t+1} = f(t+1; y_0, y_1, \ldots, y_n, \tilde{y}_{n+1}, \tilde{y}_{n+2}, \ldots, \tilde{y}_t) + \tilde{E}_{t+1}$$

$$= \tilde{a}_t + \tilde{b}_t(t+1) + \tilde{E}_{t+1}$$

t=0 t=n t=T

Base period (observed period) | **Forecasting period (projection horizon)**

Generating historical forecast errors



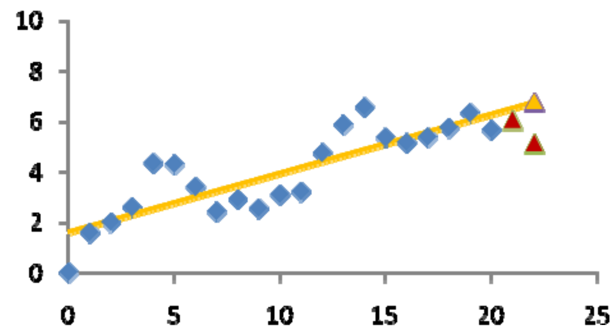
$$\hat{y}_{t+1} = f(t+1; y_0, y_1, \ldots, y_t) = a_t + b_t(t+1)$$

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1}$$

$$a_t = \bar{y} - \bar{t}\, b_t = \frac{\sum_{i=0}^{t} y_i}{t+1} - \frac{t}{2} b_t$$

$$b_t = \frac{\sum_{i=0}^{t}(t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=0}^{t}(t_i - \bar{t})^2} = \frac{\sum_{i=0}^{t}(t_i - \frac{t}{2}) y_i}{\sum_{i=0}^{t}(t_i - \frac{t}{2})^2}$$

Forecasting by simulations



$$\tilde{y}_{t+1} = f(t+1; y_0, y_1, \ldots, y_n, \tilde{y}_{n+1}, \tilde{y}_{n+2}, \ldots, \tilde{y}_t) + \tilde{E}_{t+1}$$
$$= \tilde{a}_t + \tilde{b}_t(t+1) + \tilde{E}_{t+1}$$

$$\tilde{a}_t = \frac{\sum_{i=0}^{n} y_i + \sum_{i=n+1}^{t} \tilde{y}_i}{t+1} - \frac{t}{2} b_t$$

$$\tilde{b}_t = \frac{\sum_{i=0}^{n}(t_i - \frac{t}{2}) y_i + \sum_{i=n+1}^{t}(t_i - \frac{t}{2}) \tilde{y}_i}{\sum_{i=0}^{t}(t_i - \frac{t}{2})^2}$$

Analysis of historical forecast errors ($0 \le t \le n$)

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1}$$

Analysis of historical forecast errors (0≤t≤n)

Actual demand at t+1:

$$y_{t+1} = \sum_{i=0}^{t} \varepsilon_i$$

$$where\ \varepsilon \sim N(0, \sigma_{\varepsilon}^{2})$$

Analysis of historical forecast errors (0≤t≤n)

Actual demand at t+1:

$$y_{t+1} = \sum_{i=0}^{t} \varepsilon_i$$

$$where\ \varepsilon \sim N(0, \sigma_{\varepsilon}^{\ 2})$$

Predicted demand at t+1:

$$\begin{aligned}\hat{y}_{t+1} &= a_t + b_t(t+1) \\ &= \overline{y} + b_t(t+1-\overline{t}) \\ &= \frac{\sum_{i=0}^{t} y_i}{t+1} + \frac{\sum_{i=0}^{t}(t_i - \frac{t}{2})y_i}{\sum_{i=0}^{t}(t_i - \frac{t}{2})^2}(t+1-\frac{t}{2}) \\ &\ldots\end{aligned}$$

Analysis of historical forecast errors (0≤t≤n)

Actual demand at t+1:

$$y_{t+1} = \sum_{i=0}^{t} \varepsilon_i$$

$$where\ \varepsilon \sim N(0, \sigma_\varepsilon^{\ 2})$$

Predicted demand at t+1:

$$\hat{y}_{t+1} = a_t + b_t(t+1)$$
$$= \overline{y} + b_t(t+1-\overline{t})$$
$$= \frac{\sum_{i=0}^{t} y_i}{t+1} + \frac{\sum_{i=0}^{t}(t_i - \frac{t}{2})y_i}{\sum_{i=0}^{t}(t_i - \frac{t}{2})^2}(t+1-\frac{t}{2})$$
$$\ldots$$

$$\hat{y}_{t+1} = \sum_{i=0}^{t-1} \varepsilon_i B_i^t$$

$$where\ B_i^t = \sum_{j=i+1}^{t}\left[\frac{1}{t+1} + \frac{(t+1-\frac{t}{2})}{\sum_{k=0}^{t}(t_k - \frac{t}{2})^2}(t_j - \frac{t}{2})\right]$$

Analysis of historical forecast errors (0≤t≤n)

Actual demand at t+1:

$$y_{t+1} = \sum_{i=0}^{t} \varepsilon_i$$

$$where\ \varepsilon \sim N(0, \sigma_\varepsilon{}^2)$$

Predicted demand at t+1:

$$\hat{y}_{t+1} = a_t + b_t(t+1)$$

$$= \bar{y} + b_t(t+1-\bar{t})$$

$$= \frac{\sum_{i=0}^{t} y_i}{t+1} + \frac{\sum_{i=0}^{t}(t_i - \frac{t}{2})y_i}{\sum_{i=0}^{t}(t_i - \frac{t}{2})^2}(t+1-\frac{t}{2})$$

$$\ldots$$

$$\hat{y}_{t+1} = \sum_{i=0}^{t-1} \varepsilon_i B_i^t$$

$$where\ B_i^t = \sum_{j=i+1}^{t}\left[\frac{1}{t+1} + \frac{(t+1-\frac{t}{2})}{\sum_{k=0}^{t}(t_k - \frac{t}{2})^2}(t_j - \frac{t}{2})\right]$$

Analysis of historical forecast errors (0≤t≤n)

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1} = \sum_{i=0}^{t} \varepsilon_i - \sum_{i=0}^{t-1} \varepsilon_i B_i^t = \sum_{i=0}^{t-1} \varepsilon_i (1 - B_i^t) + \varepsilon_t$$

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1} = \sum_{i=0}^{t} \varepsilon_i - \sum_{i=0}^{t-1} \varepsilon_i B_i^t = \sum_{i=0}^{t-1} \varepsilon_i (1 - B_i^t) + \varepsilon_t$$

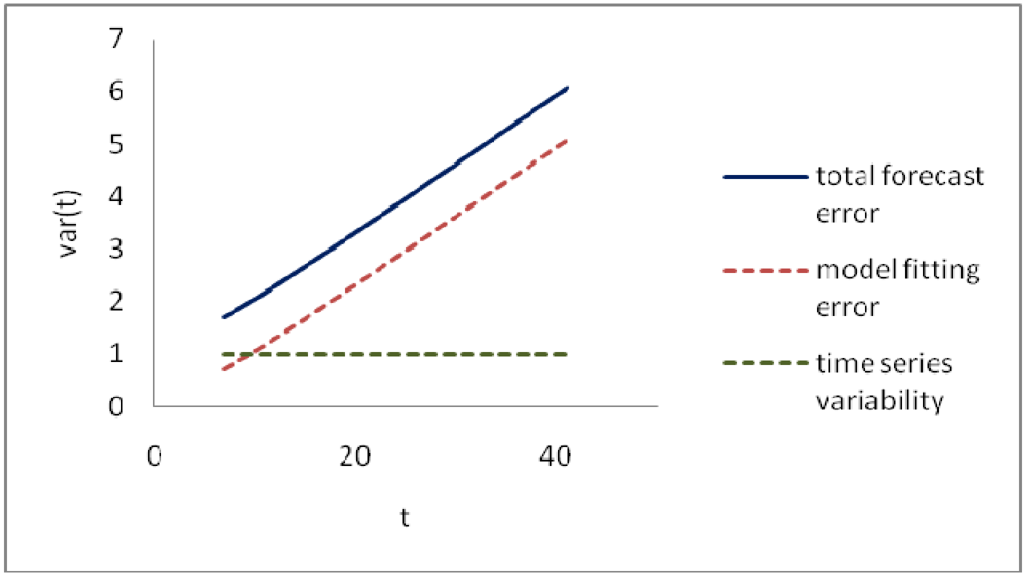
- Characteristic 1: e_t is normally distributed.
- Characteristic 2: The expected value of e_t is zero.

$$E\left[e_{t+1}\right] = E\left[\sum_{i=0}^{t-1} \varepsilon_i (1 - B_i^t)\right] + E\left[\varepsilon_t\right] = 0$$

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1} = \sum_{i=0}^{t} \varepsilon_i - \sum_{i=0}^{t-1} \varepsilon_i B_i^t = \sum_{i=0}^{t-1} \varepsilon_i (1 - B_i^t) + \varepsilon_t$$

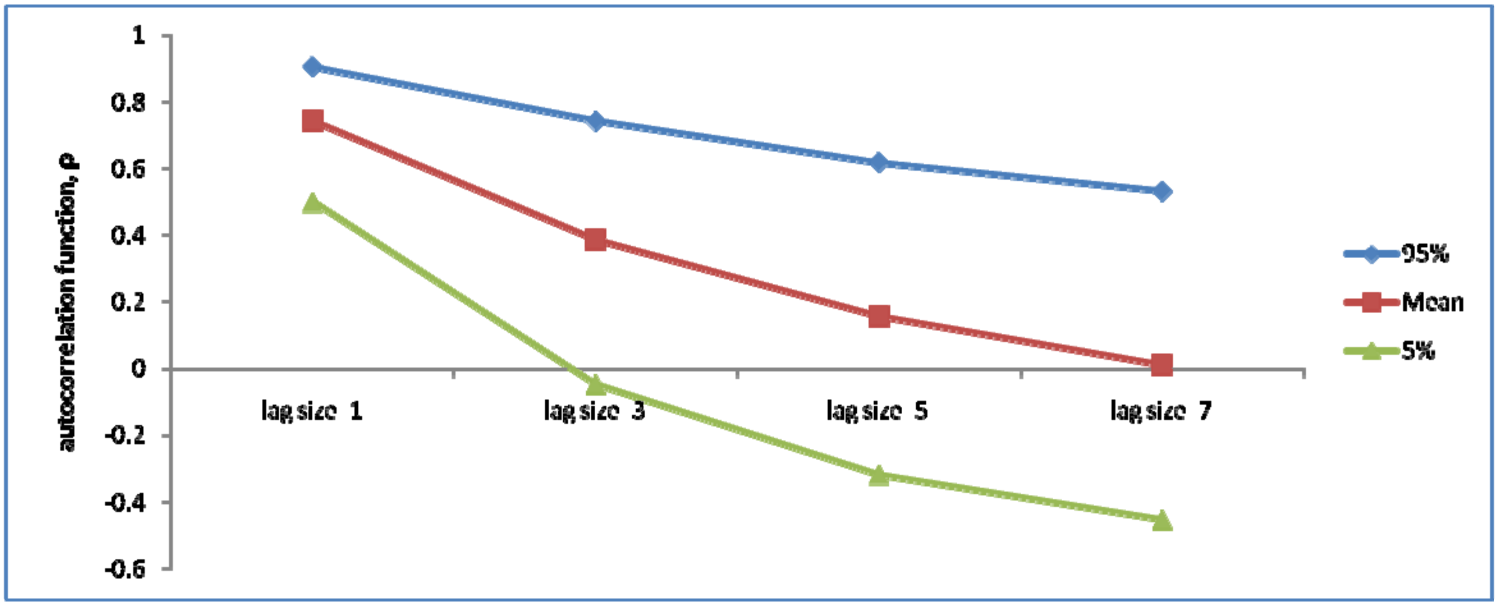
- Characteristic 3: The variance of e_t increases as time increases.

$$\sigma_{e_{t+1}}^2 = \sum_{i=0}^{t-1} (1 - B_i^t)^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2$$



$$e_{t+1} = y_{t+1} - \hat{y}_{t+1} = \sum_{i=0}^{t} \varepsilon_i - \sum_{i=0}^{t-1} \varepsilon_i B_i^t = \sum_{i=0}^{t-1} \varepsilon_i (1 - B_i^t) + \varepsilon_t$$

- Characteristic 4: e_t terms are autocorrelated.

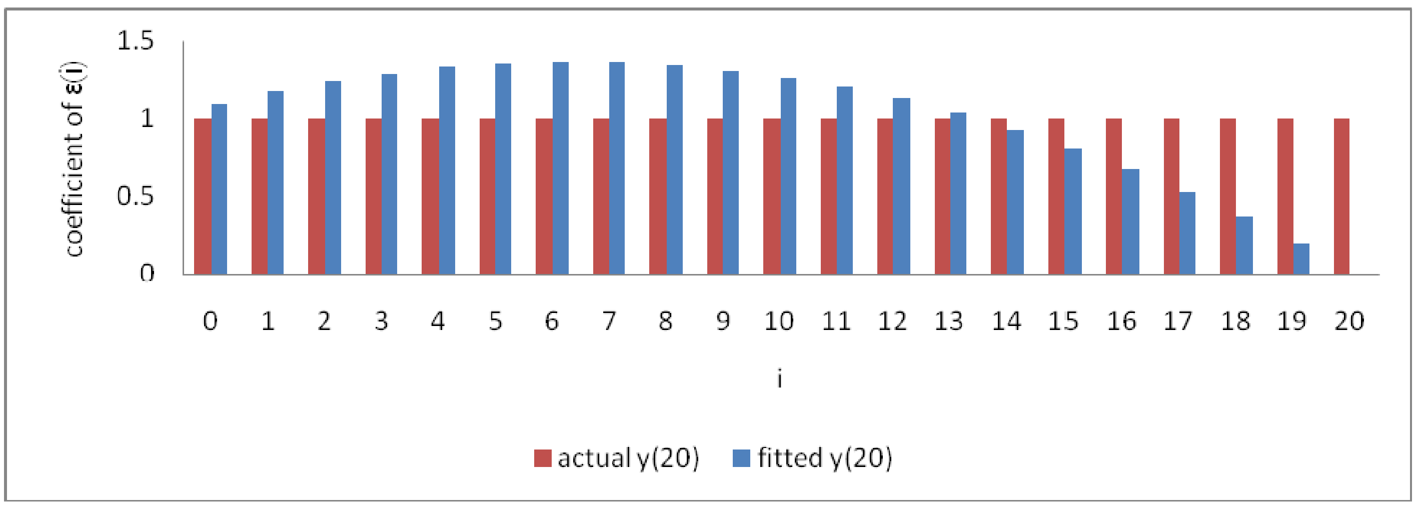


time series variability

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1} = \sum_{i=0}^{t} \varepsilon_i - \sum_{i=0}^{t-1} \varepsilon_i B_i^t = \sum_{i=0}^{t-1} \varepsilon_i (1 - B_i^t) + \varepsilon_t$$

Model fitting error

- Characteristic 5: e_t expresses model fitting error and time series variability.

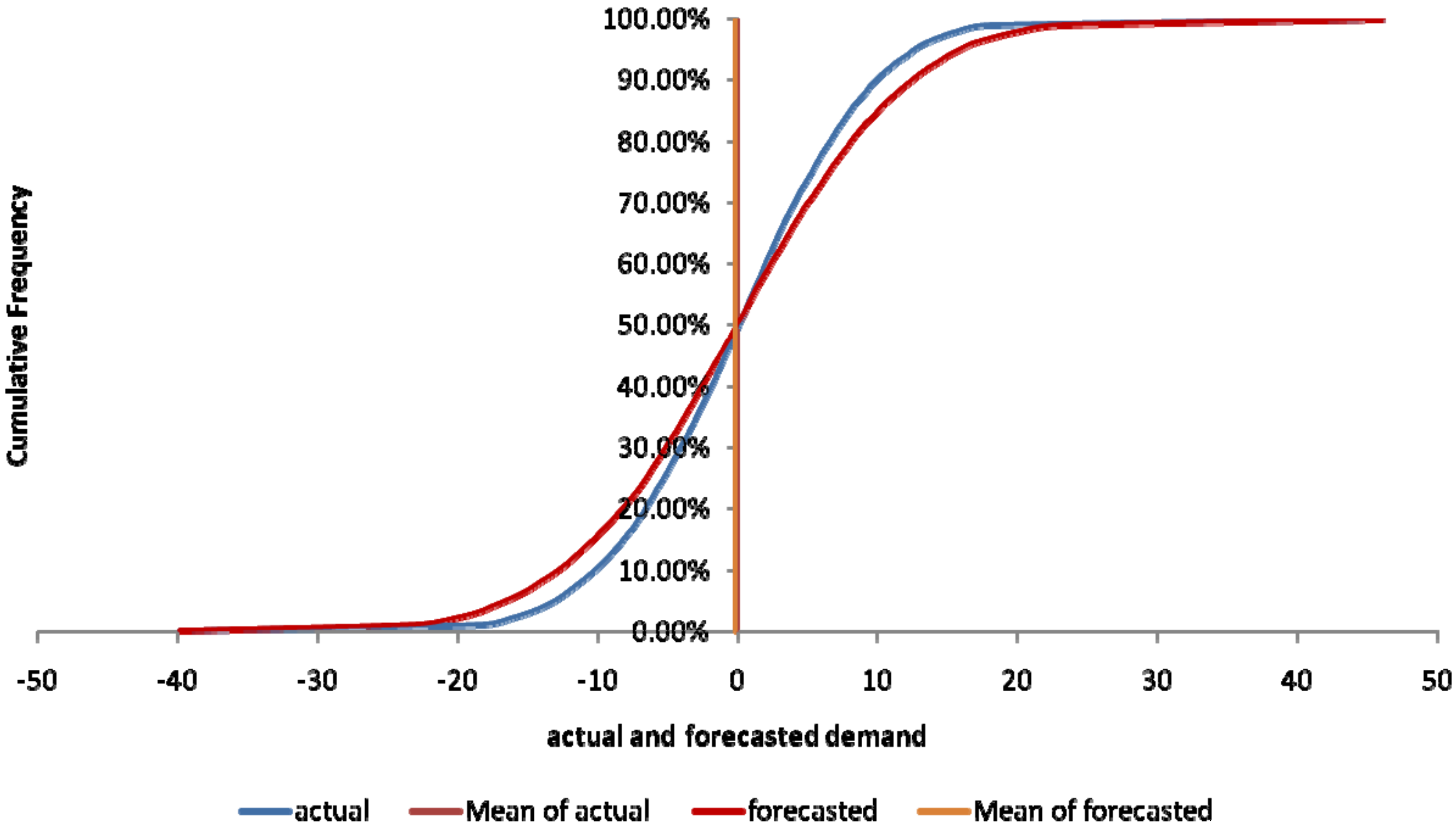


Coefficients of where $0 \le i \le 20$ for the actual and fitted demand at t=20

Assumptions in the distribution of future forecast errors (n≤t≤T)

- Future forecast errors are normally distributed.
- The distribution has expected value zero.
- The variance of future forecast error will increase at the same rate as in the past.
- A first order autoregressive (AR(1)) model describes the autocorrelation structure of forecast errors: $E_t = c + \phi E_{t-1} + u_t$

Simulation result



Cumulative distributions of the actual and forecasted demand at t=60 using the base period [0,40]**.** The normal random variable of a random walk process is assumed to have the variance of one. Simulation result is based on 10,000 Monte Carlo trials.

Conclusion

- Key assumption: The future performance of the fitting method (future forecast error) is dependent upon the performance of the same method in the past (historical forecast error).

- We illustrated that forecast errors are neither independent nor identically distributed.

- We explained how to deal with the non-stationary characteristics of forecast errors in the case of a random walk underlying time series.

- Simulation result suggests that the forecasted demand is an unbiased estimate of the actual demand and it has a higher variance compared to actual process.