Time series methods to forecast patent filings

Gerhard Dikta [a]

[a]Supported by the European Patent Office

- This project (TSM) is part of EPO's research project

  Improvements to methods for forecasting patent filings

- TSM focuses on short-term but mainly on long-term forecast.
- Summary of the results of the first year.
  - Exponential smoothing
  - ARIMA analysis
  - Transfer-function models

Exponential smoothing (ES): Euro-Total filings (ETT)

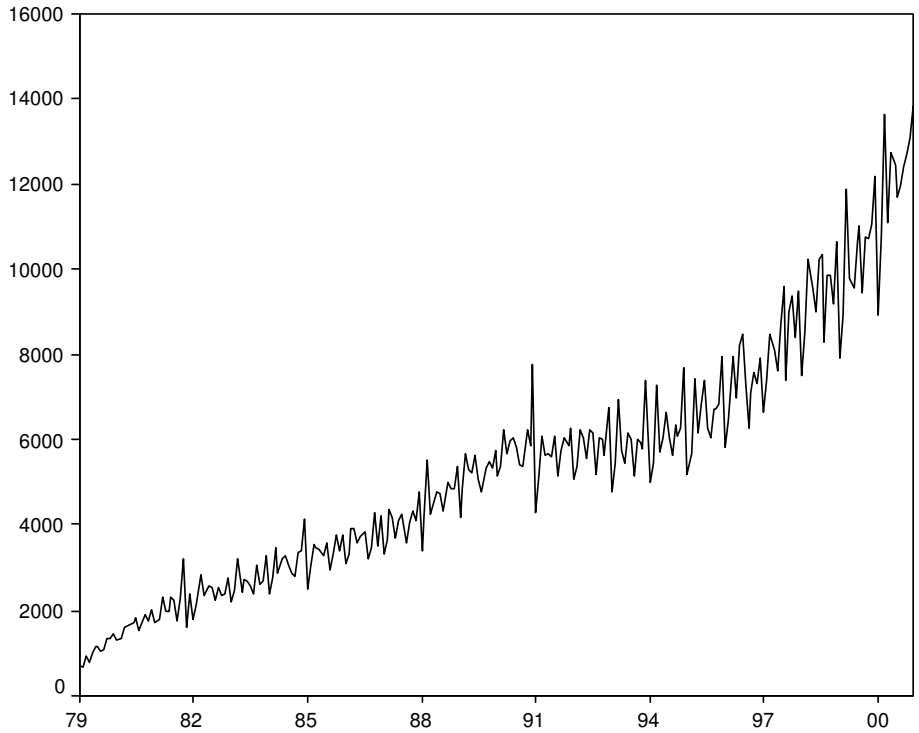

Figure 1: ETT per month, 1979 - 2000.

ES: ETT model selection

- Trend-break in 1991
  linear $\longrightarrow$ constant $\longrightarrow$ linear
  $\Rightarrow$ linear trend model
- Increasing fluctuation
  $\Rightarrow$ multiplicative seasonal component
- Model

$$(*) \quad X_i = (a_i + b_i \cdot i) \cdot s_i + \varepsilon_i$$

- Winter's method of three parameter ES.

ES: ETT forecast

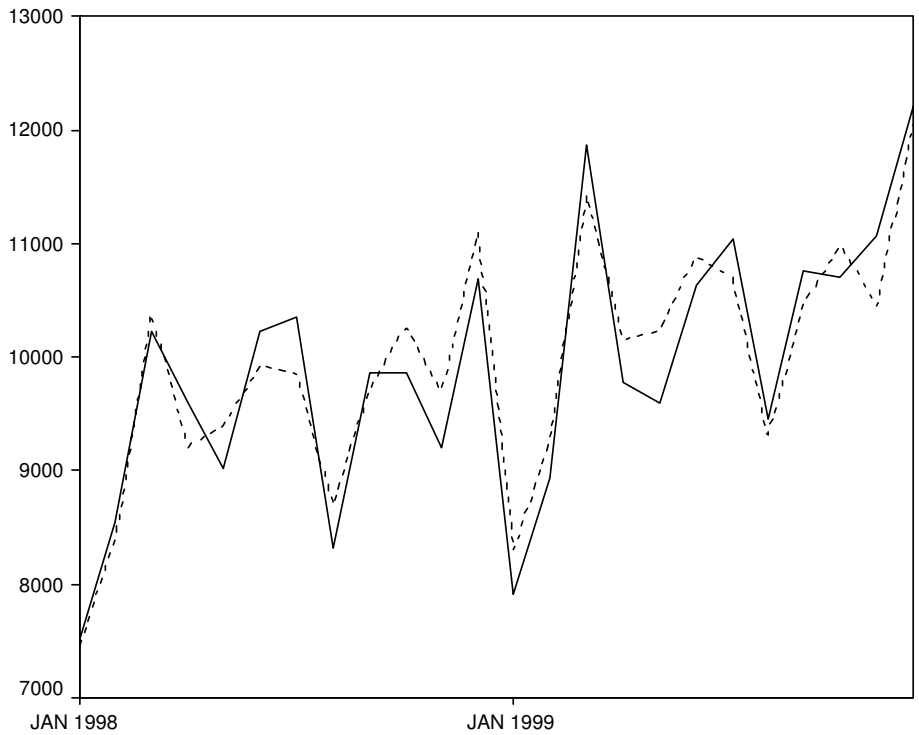

Figure 2: ETT per month, based on period 1979-1998, forecast 1999. (Actual-solid line)

ES: ETT results

- Seasonal components can be assumed to be stable.
- The latest filing figures have an influence on the linear term but also those from previous years!

ARIMA: General assumptions

- Stationarity:
  - Time series (TS) stays at a constant level.
  - Dependence structure is time-stable
- ARMA(p, q) process $(X_t)_t$: Let $(\varepsilon_s)_s$ be white noise.

$$(*) \quad X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

- ARIMA(p, d, q) $(X_t)_t$:
  - $(\nabla^d(X_t))_t$ is ARMA(p, q)
  - $\nabla X_t = X_t - X_{t-1}$, $\nabla^d X_t = \nabla(\nabla^{d-1} X_t)$

ARIMA: Application

- Stationarity:
  - Use transformation to get a homoscedastic TS.
  - Find degree d of differencing to remove the trend.
- Identification: Find the order p and q of the transformed TS.
- Parameter Estimation: Fit ARIMA(p, d, q).
- Diagnostics: Analyze the residuals.
- Forecast: Based on the fitted model and the estimated residuals a forecast together with standard errors can be obtained.

ARIMA: ETTyUSA forecast

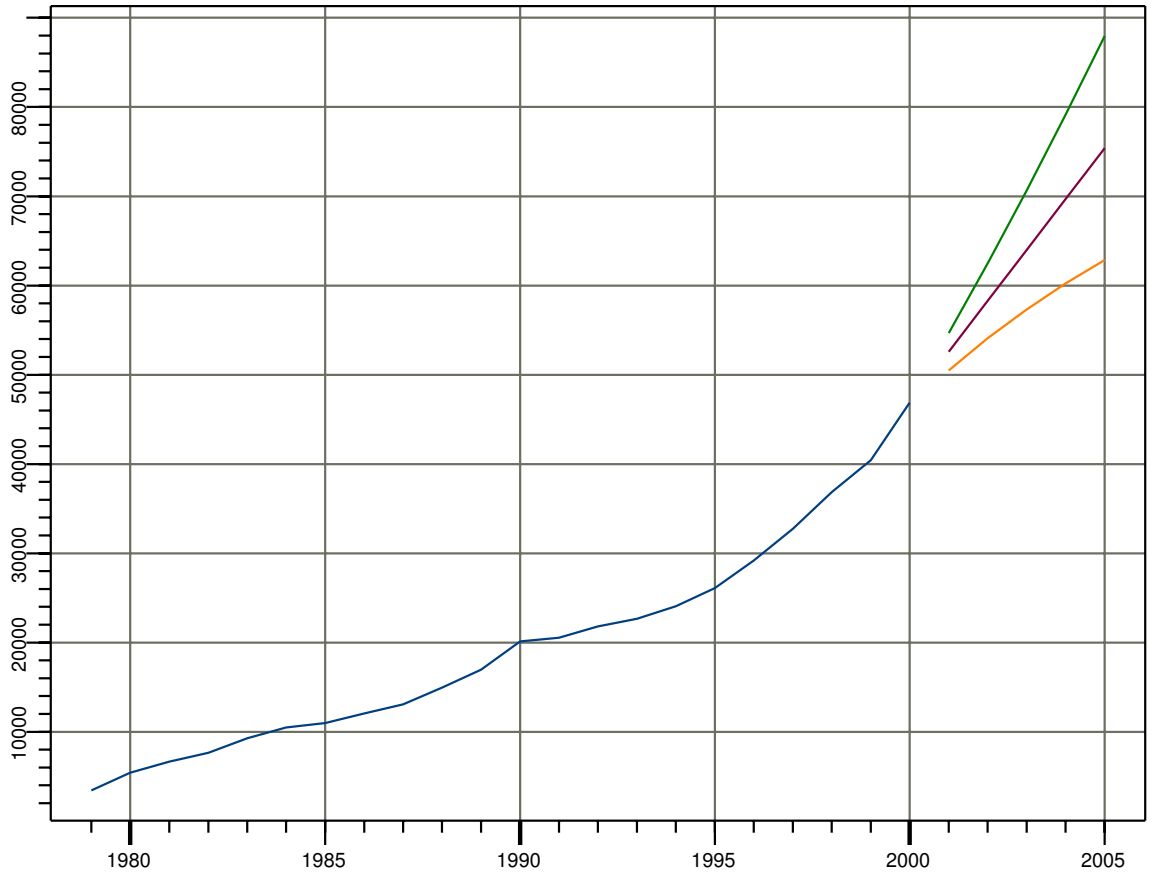

Figure 3: Forecast based on ARIMA$(0, 2, 1)$ for ETTyUSA.

ARIMA: Results

- Analysis:
  - Filings: Euro-direct, Euro-PCT international phase, and Euro-Total.
  - Blocs: Total, EPC, USA, Japan, Other.
  - Period: 1979-2000.
- Results:
  - Forecasted 2001 figures fit well to the actual ones. Exception: Bloc Other.
  - No substantial difference between aggregated blocwise forecast and forecast based on total filings.

ARIMA: Problems

- Sample size: 22 years of filings (1979-2000).
  Differencing 2-times $\Rightarrow$ sample size = 20.
  - Small sample size.
  - Not possible to model long-memory dependence.
- Precision: Standard deviation of forecast error is large, i.e 1047 in the case of ARIMA$(0, 2, 1)$ and ETTyUSA for one-year-ahead forecast.

Transfer function model (TFM): General assumptions

- ARMA processes:
  - Input $(X_t)_t$.
  - Output $(Y_t)_t$.
  - Error $(W_t)_t$.
  - $(X_t)_t$ independent of $(W_t)_t$.
- TFM(r, s, d, p, q): $(W_t)_t$ is ARMA(p, q) and

$$(*) \quad Y_t = \sum_{k=1}^{r} \omega_k Y_{t-k} + \sum_{k=0}^{s} \delta_k X_{t-d-k} + W_t$$

TFM: Application

- Stationarity: Transformation of input and output.
- Identification of r, s, d: Use Prewhitening to find the order r, s, d of the transformed TS.
- Parameter Estimation: Fit the model.
- Identification of p, q: Find the order p and q of the residuals.
- Diagnostics: Test of the assumptions.
- Forecast: Based on the fitted model and the estimated residuals a forecast together with standard errors are obtained.

TFM: R&D-expenditures in USA (RDUSA) and ETTyUSA

- RDUSA: Business expenditures in R&D in USA (source OECD) deflated by GDP (source WTO).
- Stationarity:
  - Input: $X = \nabla^3 \log(\text{RDUSA})$.
  - Output: $Y = \nabla^2$ ETTyUSA.
- Identification of r, s, d:
  - Input: $(X_t)_t$ is ARMA$(3, 0)$.
  - Prewhitening: $r = 0$, $s = 1$, $d = 4$
  - Identified model: $p = 0$, $q = 0$

$$(*) \quad Y_t = \delta_0 X_{t-4} + \delta_1 X_{t-5} + c + W_t$$

TFM: Results and Problems

- Results: RDUSA-ETTyUSA
  - Observed time-lag between RDUSA and ETTyUSA 4-8 years.
  - TFM-forecast above ARIMA.
  - Standard deviation of the one-year-ahead TFM-forecast is 905 (ARIMA approach 1047).
- Problems:
  - Small sample size of filing data.
  - Deflator?