LESZEK KLUKOWSKI
INSTYTUT BADAŃ SYSTEMOWYCH PAN

REVIEW OF METHODS OF FORECASTING OF TIME SERIES

LESZEK KLUKOWSKI

IBS PAN, WARSZAWA, 21.03.2014

Several last decades are the period of intensive development of forecasting methods. The following factors stimulate this fact:

- **development of theory of forecasts – mainly in the area of stochastic methods and artificial neuronal networks,**
- **dynamic increase of capacity of computing systems,**
- **complexity of economic phenomena especially reflected at financial markets and availability of huge data sets in computer systems.**

Existing methods allow modeling and forecasting of phenomena with significant complexity, variability and variety. However, there exists the need for further works – in the area of theory and application – in this field.

Monographies:

C.W.J. Granger;

J.D. Hamilton;

D.C. Montgomery, Ch.L. Jennings, M. Kulahci

G.E.P. Box, G.M. Jenkins;

D.C. Montgomery, L.A. Johnson;

H. Tong;

P.H. Frances, D. van Dijk

Journals: forecasting, time series, econometrics

Noble prizes: Jan Tinbergen, C.W.J. Granger

PLAN OF PRESENTATION

1. Introduction to the theory of forecasting.
2. Main directions of development of forecasting methods.

- univariate linear time series models,
- combination of forecasts,
- multiple linear time series models,
- univariate nonlinear time series models.

3. Some comparisons of empirical forecasts.
4. Summary and conclusions.

1. Introduction to the theory of forecasting.

2. Main directions of development of forecasting methods

- univariate linear time series models:

  trend (regression), exponential smoothing, ARIMA;

- combination of forecasts:

  linear combination and others (nonparametric approach, artificial neural networks);

- multiple linear time series models:

  bivariate ARMA;

- univariate nonlinear time series models (having multivariate extensions):

  Kalman's filter, ARCH and GARCH, regime switching, artificial neural networks.

3. Some comparisons of empirical forecasts.

4. Summary and conclusions.

5. Basic literature.

1. Introduction to the theory of forecasting

The reasons of application of time series forecasting

- speed and low costs,
- formalization of forecasting process and known properties of forecasts,
- limited requirements about data,
- broad spectrum of methods
- reference point to other methods,
- component for combining forecasts,
- often, the only formalized method of forecasting.

Background for forecasting

• Forecasting – prediction of future events or processes based on rational bases.

• Assumptions – there exists a "mechanism" of predicted process (usually stochastic process), which can be identified, with a use of the statistical methods and extrapolated in the future. The stochastic process (SP) is the sequence of dependent in some way random variables. Identification – on the basis of time series (TS), i.e. some set of observations - finite realization of the process.

• The TS is used for determination of: a type of SP (model), estimation of its parameters, validation of the model (ex'ante, ex'post), extrapolation of a model (forecasting) and determining of precision of forecasts.

• The forecast can be in point or interval form, which are analogy to point and interval estimators; for point forecasts is possible to determine measurements of accuracy, usually variance (or SD) of prediction error.

• Optimal forecast minimizes (usually) the variance of error.

- Area of application of forecasting methods:
  - Financial market,
  - Economy, monetary policy, public finance,
  - Operational management,
  - Industrial processes,
  - Demography.

- Main groups of linear methods:
  - Regression models,
  - Exponential smoothing methods,
  - Autoregressive moving averages models (ARIMA),
  - Transfer function and intervention models (ARMAX not discussed),
  - Combaining forecasts.

- Main groups of non-linear methods:
  - Autoregressive conditional heteroscedastic models (ARCH),
  - Generalized autoregressive conditional heteroscedastic models (GARCH),
  - State space models,
  - Regime-switching models: threshold autoregressive, Markov-switching (TAR, SETAR, STAR, MSW),
  - Artificial networks models.

- The forecasting process (in general):
  - Problem definition,
  - Data collection,
  - Data analysis,
  - Model selection and fitting (estimation),
  - Model validation,
  - Forecasting model deployment,
  - Monitoring forecasting model performance.

• Basing statistical tools

Time series: $y_t, t = 1, ..., T;$

forecast: $\hat{y}_{T+h}, h \geq 1,$

forecast error: $e_{t+h} = y_{t+h} - \hat{y}_{t+h},$

stationary TS:

- **strictly - if joint probability distribution of** $y_t, y_{t+1}, ..., y_{t+n}$ **is the same as** $y_{t+k}, y_{t+k+1}, ..., y_{t+k+n}$**;**
- **weakly - if the expected value** $E(y_t) = \mu_y$ **and variance** $\sigma_y^2 = E(y - \mu_y)^2$ **are constants and covariance function** $\gamma_k = Cov(y_t, y_{t+k}),\ k = 0, 1, 2, ..., K$ **is a function of lag** k**.**

Sample autocovariance c_k **and autocorrelation** r_k **functions:**

$$c_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y}),\quad k = 0, 1, 2, ..., K;\quad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0}.$$

Forecasting errors (for comparisons of models):

Mean error (one-step-ahead): $ME = \frac{1}{n}\sum_{t=1}^{n} e_t$,

Mean absolute deviation: $MAD = \frac{1}{n}\sum_{t=1}^{n} \left| e_t \right|$,

Mean square error: $MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$,

Mean absolute percent: $MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{e_t}{y_t} 100 \right|$.

Criteria for model adequacy:

Mean square error of residuals: $s^2 = \frac{1}{T-p}\sum_{t=1}^{T} e_t^2$

(*p* **– number of estimated parameters of the model),**

R – square statistic (0, 1): $R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$

Adjusted R – square: $R^2_{adj} = 1 - \dfrac{s^2}{\frac{1}{T-1}\sum_{t=1}^{n}(y_t - \bar{y})^2}$

Akaike information criterion: $AIC = \ln(\dfrac{s^2}{T}) + \dfrac{2p}{T}$ **,**

Schwarz information criterion: $SIC = \ln(\dfrac{s^2}{T}) + \dfrac{p\ln(T)}{T}$ **,**

Corrected (consistent) AICC: $AICC = \ln(\dfrac{s^2}{T}) + \dfrac{2T(p+1)}{T-p-2}$ **.**

SIC, AICC – consistent criterions of model selection (detect true model as *T* gets large), AIC – asymptotically consistent (approaches the true model as fast as any other criterion).

• The basis for time series forecasting - the theory of stochastic processes: identification (of type), estimation, verification and optimal extrapolation - for assumed criterion function (variance of error).

*** The first models have been based on the formula:**

$$Y_t = f(t) + \varepsilon_t; \quad t = 1, ...,$$

Y_t **– forecasted series (random variables),**

$f(t)$ **– some (deterministic) function of time** t**,**

ε_t **– (Gaussian) white noise** $N(0\ , \sigma_\varepsilon^2)$**;**

Stochastic assumptions can be significantly weaker.

• The basis of the model – the representation of the stochastic process Y_t **as the sum of (uncorrelated) deterministic** D_t **and stochastic** Z_t **components (Wold, Cramer):**

$$Y_t = D_t + Z_t \quad (t = 1, ...,); \quad Y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}; \quad Y_t = \sum_{j=0}^{\infty} c_j^2 < \infty.$$

D_t **- deterministic,** Z_t **- purely non-deterministic,** c_j^2 **- parameters,** D_t, Z_t **- uncorrelated.**

*** Examples:**

$$f(t) = b_0 + b_1 t + ... + b_k t^k, \quad f(t) = b_0 + b_1 \sin\tfrac{2\pi t}{12} + b_2 \cos\tfrac{2\pi t}{12}.$$

• The estimation determines the optimal estimates of parameters of the models, the verification examines accuracy of the obtained form, the extrapolation determines optimal forecasts, i.e. minimizing assumed criterion, e.g. variance of forecast error (optimal predictor - the conditional expected value of the forecasted variable, for given information set).

• The estimation applies statistical methods: the least squares, maximal likelihood, non-parametric methods, robust methods.

• The verification is usually based on statistical tests, which verify: adequacy of the model and time series (e.g. R^2), significance of its parameters (e.g. *t*-Student) and properties of forecasts (e.g. Chi-square). The verification can confirm the model or suggest its modification: change of a class, analytical form or parameters (e.g. an order of auto-regression).

• The extrapolation generates optimal forecasts and measures of their precision. Main forms of the forecasts are: point ($\hat{Y}_{t+h}$) and interval (general form of linear models: $\hat{Y}_{t+h} \mp t_{\alpha/2,T-p}\, const(\hat{\sigma}_{\varepsilon})$, with confidence coefficient (from *t*-Student distribution) $(1-\alpha)$), for horizon, $h \geq 1$.

*** Simple example of a forecast and its error:**

$$\hat{Y}_{T+h} = f(T+h); \quad h \geq 1,$$

$$Var(\hat{Y}_{T+h} - Y_{T+h}) = Var(Y_{T+h}) + Var(\hat{Y}_{T+h}).$$

• The process of the model (stochastic process) building can be not easy-to-operate; typical steps include (P.H. Franses, D. van Dijk):

- **calculate certain statistics indicating a type of a model (ARIMA: autocovariances functions),**
- **comparing the statistics with theoretical values, if the type of the model is adequate,**
- **estimate the parameters of the model, suggested in previous steps, on the basis of information set,**
- **evaluate the model using diagnostic measures,**
- **re-specify the model if necessary,**
- **use the model for forecasting (or analytic) purposes.**

• Final model selection is typically realized by comparison of different forms of models, using statistical tests and other measures, e.g. AIC or BIC criterion.

The model, which satisfies verification requirements is reliable and useful in practice!

All models are wrong, some are useful.
G.E.P. Box

2. Main directions of development of forecasting methods

• Simple models, like trend, have good (simple currently) theory: estimation, verification and forecasting, but are applicable under restrictive assumptions, especially:

*** simple, constant form of a trend and**

*** restrictive assumptions about the form of random variables of stochastic components (constant parameters, independency).**

Therefore can be applied to simple phenomena.

• The next models, developed in last decades, have eliminated these disadvantages in the following, main, directions (overlapping classification):

- **relaxing the assumptions about stability of a structure and parameters of the model,**
- **allowing for complex form of stochastic dependencies between variables of the stochastic process,**
- **advancing analytical form of the model.**

• Important feature of the models – parsimonious parametrization (methods efficient for moderate number of observations).

Directions for univariate time series

• Simple trend models

$$Y_t = f(t) + \varepsilon_t; \quad t = 1, ..., \ (\varepsilon_t \sim i.i.d., \ N(0, \sigma_\varepsilon^2))$$

$$f(t) = b_0 + b_1 t \ + ... + b_k t^k,$$

$$f(t) = b_0 + b_1 \sin\tfrac{2\pi t}{12} + b_2 \cos\tfrac{2\pi t}{12};$$

Estimation of the polynomial trend, k=1:

$$\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ ..... & \\ 1 & T \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_T \end{bmatrix},$$

$$E(\hat{\mathbf{b}}) = \mathbf{b} \quad \mathbf{V} = \mathbf{E}[(\hat{\mathbf{b}} - \mathbf{b})(\hat{\mathbf{b}} - \mathbf{b})'] = \mathbf{G}^{-1}\sigma_\varepsilon^2,$$

$$\mathbf{G} = \begin{bmatrix} T & \frac{T(T+1)}{2} \\ \frac{T(T+1)}{2} & \frac{T(T+1)(2T+1)}{6} \end{bmatrix}$$

Estimates equivalent to ML (unbiased, consistent, minimal variance, asymptotic Gaussian distributions).

$$\hat{\sigma}_\varepsilon^2 = \frac{\varepsilon'\varepsilon}{n-k}, \quad \varepsilon = \mathbf{y} - \hat{\mathbf{y}}, \quad E(\hat{\sigma}_\varepsilon^2) = \sigma_\varepsilon^2,$$

$$Var(\hat{b}_0) = \frac{2(2T+1)}{T(T-1)}\sigma_\varepsilon^2, \quad Var(\hat{b}_1) = \frac{12}{T(T^2-1)}\sigma_\varepsilon^2,$$

$$Cov(\hat{b}_0, \hat{b}_1) = \frac{-6}{T(T-1)}\sigma_\varepsilon^2.$$

Verification of the model:

- verification of identical, independent, Gaussian distribution of ε_t using appropriate statistical tests,

- significance of b_1 the hypothesis: $H_0 : b_1 = 0, \; H_1 : b_1 \neq 0 \; (>,<),$ test *t*-Student,

- significance of the model: test F-Snedecor (under Gaussian assumption).

Forecasts:

Point: $\hat{y}_{T+h} = \mathbf{z}'_{T+h}\hat{\mathbf{b}}'$

$$Var[e_h(T+h)] = \sigma_\varepsilon^2 + \frac{2}{T(T^2-1)}[(2T-1)(T-1)+6h(T+h-1)]\sigma_\varepsilon^2$$

Interval (symmetric): $100(1-\alpha)\%$

$$\hat{y}_{t+h} \pm t_{\alpha/2,n-k}[\hat{\sigma}_\varepsilon^2(1+\mathbf{z}'_{T+h}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}'_{T+h}]^{1/2}.$$

Main properties of forecasts:

- unbiasedness, minimal variances (optimal precision), Gaussian distributions with known parameters.

Computations: EXCEL, MINITAB, TSP, STATISTICA, SPSS

If you have to forecast, forecast often.
E.R. Fiedler

• Relaxing the assumptions about stability of parameters of the model; some developments:

- **<u>exponential smoothing</u> (Brown, Holt-Winters) – the idea: re-estimating the model parameters each period in order to incorporate the most recent periods data (with weights decreasing exponentially);**

  the simplest - constant model, first order exponential smoothing (e.g. prices of a fuel):

$$Y_t = b + \varepsilon_t \quad (t = 1, ..., T),$$

$$\tilde{y}_t = \lambda \sum_{t=0}^{T-1} (1-\lambda)^t \, y_{T-t},$$

$$\tilde{y}_t = \lambda y_t + (1-\lambda)\,\tilde{y}_{t-1},$$

$\lambda \in (0,1)$ **- smoothing constant (typically** $\lambda \in [0{,}1,\ 0{,}4]$**),**

with the variance

$$Var(\tilde{y}_T) = \frac{\lambda}{2-\lambda} Var(y_T);$$

Estimate $\hat{b}_0$ is obtained from minimization

$$\sum_{t=0}^{T-1}(1-\lambda)^t\,(y_t - b_0)^2$$

for large T

$$\hat{b}_0 \cong \lambda \sum_{t=0}^{T-1}(1-\lambda)^t\, y_{T-t}$$

Typical models:

Linear trend model: $Y_t = b_0 + b_1 t + \varepsilon_t$

General trend model: $Y_t = b_0 + b_1 t + \ldots + \frac{b_k}{k!}t^k + \varepsilon_t$

Sinusoidal trend $Y_t = b_0 + b_1 \sin\frac{2\pi t}{d} + b_2 \cos\frac{2\pi t}{d} + \varepsilon_t$.

General approach

- **The theoretical basis (Brown):**

THEOREM OF EXPONENTIAL SMOOTHING

for general *n*th–degree polynomial

$$y_t = b_0 + b_1 t + \cdots + \frac{b_n}{n!} t^n + \varepsilon_t \quad (\varepsilon_t \sim iid.\, N(0, \sigma_\varepsilon^2).$$

However, for $n>2$ the calculations get complicated and ARIMA models can instead be considered.

The process $\{Y_t, t = 1, ...\}$ (constant or trend) is changing slowly (parameters b_k). The system of weights heuristic, but with useful statistical (practical) properties.

The point forecast (constant with constant precision):

$$\hat{y}_{T+h} = \hat{b}_0 = \tilde{y}_T, \qquad Var(\hat{y}_T) = \frac{\lambda}{2-\lambda} Var(y_T) = \frac{\lambda}{2-\lambda}\sigma_\varepsilon^2,$$

$$\hat{\varepsilon}_T = y_T - \hat{y}_T.$$

The interval forecast (constant – often unrealistic): $\hat{y}_T \mp u_{\alpha/2}\hat{\sigma}_e$,
where: $u_{\alpha/2}$ **appropriate percentile of standard Gaussian distribution and** $\hat{\sigma}_e^2 = \frac{1}{T}\sum_{t=1}^{T}(y_{t+1} - \hat{y}_{t+1})^2$, $\hat{y}_{t+1}$ **- one-step-ahead forecast (on historic data).**

Choice of λ **- minimization of:** $SS(\lambda) = \sum_{t=1}^{T}(y_t - \hat{y}_t)^2$.

Point forecast for linear trend model:

$$\hat{y}_{T+h} = \hat{\beta}_{0,T} + \hat{\beta}_{1,T}(T+h) = \hat{y}_T + \hat{\beta}_{1,T}h,$$

$$\hat{\beta}_{0,T+1} = \lambda(1+\lambda)\, y_{T+1} + (1-\lambda)^2(\hat{\beta}_{0,T} + \hat{\beta}_{1,T}),$$

$$\hat{\beta}_{1,T+1} = \tfrac{\lambda}{2-\lambda}(\hat{\beta}_{0,T+1} - \hat{\beta}_{0,T}) + \tfrac{2(1-\lambda)}{2-\lambda}\hat{\beta}_{1,T}.$$

Interval forecast:

$$(2 + \tfrac{\lambda}{1-\lambda}h)\,\hat{y}_T^{(1)} - (1 + \tfrac{\lambda}{1-\lambda}h)\,\hat{y}_{T-1}^{(2)} \mp u_{\alpha/2}\tfrac{c_h}{c_1}\hat{\sigma}_e$$

$$\hat{y}_T^{(2)} = \lambda\,\hat{y}_T^{(1)} + (1-\lambda)\,\hat{y}_{T-1}^{(2)}$$

$$c_i^2 = 1 + \tfrac{\lambda}{(2-\lambda)^3}[(10 - 14\lambda + 5\lambda^2) + 2i\lambda(4-3\lambda) + 2i^2\lambda^2].$$

Validation of the model

The basis: sample autocorrelation function of one-step-ahead forecasting errors r_k, $k = 1, ..., \kappa T$ $(\kappa \in (0,1))$ **- should be around** 0 **(zero) with standard deviation** $1/\sqrt{T}$ **. If the values lie outside the** $\pm 2/\sqrt{T}$ **limits it require examination.**

Another approach: monitoring and modifying the discount factor λ**, e.g. Trigg and Leach (1967), Chow (1965).**

There exists models for seasonal data: additive and multiplicative.

Summary assessment

• The models are optimal (minimize mean square error) for some ARIMA(0, *k*, *k*) processes, the forecasts typically not (similarly as estimates of parameters); the benefit – simple model and computations.

• The features of the methods:

- **allowing some changes of the phenomena forecasted,**
- **possibility of computerization (automation) of model building and forecasting,**
- **short horizon of forecasts,**
- **empirical (typical): not unbiased, not optimal, often correlated errors.**

• Allowing for (more) complex form of stochastic dependencies between variables of the stochastic process

- **ARIMA(*p, d, q*) models (Box, Jenkins)**

$$\Phi(B)(1-B)^d Y_t = \Theta(B)\varepsilon_t$$

where:

$B^d(y_t) = y_{t-d}; \quad d \geq 1,$

$\Phi(B) = (1-\phi_1 B - ... - \phi_p B^p),$

$\Theta(B) = (1-\theta_1 B - ... - \theta_q B^q),$

$\Phi(B), \Theta(B)$ - lag-polynomial operators (with roots outside unit circle – providing stationarity $\Phi(B)$ and invertibility $\Theta(B)$),

d – parameter providing stationary process,

ε_t - Gaussian white noise.

• The (linear) model reflects finite autoregressive and moving average dependencies of variables y_t, with *d* differencing providing (weak) stationarity. There exist optimal estimators and predictors; estimators are non-liner in the case of *q*≥1.

• **The basis:**

THEORY OF ARIMA PROCESSES

The main assumptions:

- **weak stationarity of TS: the expected value of TS** $\mu_y = E(y_t)$ **is constant, not dependent on time,**
- **invertibility of ARMA process: roots of** $\Theta(B)$ **less than one in absolute value (has an infinite AR representation),**
- **the autocovariance function** $\gamma_y(k) = Cov(y_t, y_{t+k})$ $(k = 1, 2, ...,)$ **for any lag** k **is only a function of** k **and not time.**

The form of the ARMA(*p*, *q*) model

$$y_t = \delta + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} \qquad \textbf{or}$$

$$\Phi(B)\, y_t = \delta + \Theta(B)\, \varepsilon_t \,.$$

The main tools for model building (identification, estimation, validation):

- **Identification based on the autocorrelation function and partial autocorrelation function,**
- **Estimation the least squares, maximum likelihood – non-linear for $q>0$,**
- **Validation based on:**

*** residuals (realization of Gaussian white noise):**

$$\hat{\varepsilon}_t = y_t - (\hat{\delta} + \sum_{i=1}^{p} \hat{\phi}_i \, y_{t-i} - \sum_{i=1}^{q} \hat{\theta}_i \, \hat{\varepsilon}_{t-i}),$$

*** statistic based on autocorrelations (chi-square K-p-q):**

$$Q = (T-d)\sum_{k=1}^{K} r_\varepsilon^2(k) \quad \textbf{or} \quad Q = T(T+2)\sum_{k=1}^{K} \frac{1}{T-k} r_\varepsilon^2(k).$$

Optimal predictor:

Point forecast (in time T for time $T+h$, $h \geq 1$) – conditional expectation of y_{T+h} for given $y_T, y_{T-1}, \ldots$:

$$\hat{y}_{T+h} = E[y_{T+h} \mid y_T, y_{T-1}, \ldots],$$

$$e_T(h) = y_{T+h} - \hat{y}_{T+h},$$

$$E(e_T(h)) = 0,$$

$$Var(e_T(h) = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \psi_i^2 = \sigma^2(h),$$

with coefficients ψ_i determined by relationship $\Psi(B) = \Phi(B)^{-1}\Theta(B)$;

The interval forecast: $\hat{y}_{T+h} \pm z_{\alpha/2}\sigma(h)$.

Example ARIMA(1, 1, 1) **process:**

$$(1-\phi B)(1-B)\,y_{t+h} = (1-\theta B)\,\varepsilon_{T+h}\ ,$$

$$\hat{y}_{T+h} = (1-\phi)\,y_{T+h-1} - \phi\,y_{T+h-2} + \varepsilon_{T+h} - \theta\,\varepsilon_{T+h-1},$$

$$\hat{y}_{T+h} = \begin{cases} (1-\phi)\,y_T - \phi\,y_{T-1} - \theta\,e_T(1), & e_T(1) = y_T - \hat{y}_T(T-1), \quad h=1, \\ (1-\phi)\,\hat{y}_{T+h-1} - \phi\,\hat{y}_{T+h-2}, & h \geq 2. \end{cases}$$

The rules of forecasting computations:

- unknown values $y_{T+k},\ k>0$ **are replaced by they forecasts** $\hat{y}_{T+k}$,

- "forecasts" of $y_{T+k},\ k \leq 0$ **are the known values,**

- the optimal forecast of $\varepsilon_{T+k},\ k>0$ **is zero, the "forecasts" of** $\varepsilon_{T+k},\ k \leq 0$ **are known values** ε_{T+k}**.**

Practical approach to model building:

- **To determine "potential" models on the basis of autocorrelation and partial autocorrelation function and to estimate these models,**
- **To choose the best version on the basis of estimates of variance $Var(\varepsilon_t)$, values of criterions AIC, BIC and tests used in validation.**

Sometimes more than one version satisfies criterions used during validation; their forecasts are usually similar.

• The features of the methods:

- **optimal properties of statistical tools: identification, estimation, verification and prediction,**
- **broad area of application,**
- **successful applications,**
- **"rigid" assumption about stable form of the model.**

- Combination of forecasts

• The (unbiased) forecasts can be often obtained on the basis of two or more sources (models). Typically, combination of such forecasts, e.g. weighted average, is more precise (a lower variance), than individual forecasts (Granger). It is analogy to linear combination of estimators; the weights can be determined in optimal way (Rao, Serfling).

• In the case of two forecasts the solution, the optimal weight k_0 ($0<k_0<1$), providing minimum error variance $\sigma^2_{c,0}$, has a form:

$$k_0 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}, \text{ and}$$

$$\hat{y}_{comb} = k_0\,\hat{y}_1 + (1-k_0)\,\hat{y}_2,$$

$$\sigma^2_{c,0} = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2},$$

σ_1, σ_2 - standard deviations of forecasts $\hat{y}_1, \hat{y}_2$, ρ - correlation coefficient of forecasts.

• In practice the weights are determined in many ways, e.g. on the basis of n-1 forecasting errors (Granger) – via ML estimator (under assumption that forecasts have bivariate Gaussian distribution, with known variances and covariance) k_n:

$$k_n = \frac{\sum_{t=n-v}^{n-1} e_t^{(2)2}}{\sum_{t=n-v}^{n-1} (e_t^{(1)2} + e_t^{(2)2})} .$$

• The combining forecasts can be also obtained with the use of artificial neuronal networks or non-parametric approach.

• Typically, combining forecasts exceed significantly individual forecasts, but require more than one forecast (model).

- **The features of the method:**
  - **require versatility of a forecaster,**
  - **typically - the best empirical precision, because make use of positive properties of individual, usually suboptimal, forecasts,**
  - **allow for application of forecasts from different sources (e.g. institutions).**

- Multiple time series models

• Majority the univariate models have multivariate extensions (theory) and applied in practice. The following models have particular importance:

linear

*** ARIMA (ARMAX),**

nonlinear

*** Kalman's filter**

*** ARCH and GARCH,**

*** threshold and Markov switching.**

• The multivariate models allow causality and feedback. Typically, the number of variables is low; exception is Kalman's model.

• The ARMA bivariate (vector) model with feedback (Granger approach):

$$Y_{1,t} = \frac{\omega_1^*(B)}{\delta_1^*(B)} Y_{2,t} + \frac{\theta_1^*(B)}{\varphi_1^*(B)} \eta_{1,t}^*,$$

$$Y_{2,t} = \frac{\omega_2^*(B)}{\delta_2^*(B)} Y_{1,t} + \frac{\theta_2^*(B)}{\varphi_2^*(B)} \eta_{2,t}^*,$$

$(Y_{1,t}, Y_{2,t})$ **- bivariate ARMA (with mean subtracted),**

$(\eta_{1,t}^*, \eta_{2,t}^*)$ **- bivariate white noises, mutually uncorrelated**

$(\omega_2^*(0) \equiv 0)$,

$\omega_1^*(B)$, $\delta_1^*(B)$, $\omega_2^*(B)$, $\delta_2^*(B)$, $\theta_2^*(B)$, $\varphi_2^*(B)$ **- parameters;**

The above relationships result from:

$\Phi(B)\mathbf{Y}_t \equiv \Theta(B)\boldsymbol{\eta}_t$ ($\boldsymbol{\eta}_t$ **- vector white noise)**

$$Y_{i,t} = \sum_{j \neq i} \frac{\omega_{ij}^*(B)}{\delta_{ij}^*(B)} Y_{j,t} + \sum_{j \neq i} \frac{\theta_i^*(B)}{\phi_i^*(B)} \eta_{i,t}, \quad (i = 1, 2)$$

example:

$$Y_{1,t} = \frac{\omega_{1,0}^*}{1 - \delta_{1,1}^* B - \ldots - \delta_{1,4}^* B^4} Y_{2,t} + \frac{1 - \theta_{1,1}^* B - \theta_{1,2}^* B^2}{1 - \varphi_{1,1}^* B - \varphi_{1,2}^* B^2} \eta_{1,t}^*,$$

$$Y_{2,t} = \frac{\omega_{2,1}^* B + \ldots + \omega_{2,3}^* B^3}{1 - \delta_{2,1}^* B} Y_{1,t} + (1 - \theta_{2,1}^* B - \theta_{2,2}^* B^2) \eta_{2,t}^*.$$

The model building steps, simplified (Granger, Newbold):

- **fit the single series models for** $Y_{j,t}$ $(j=1,2)$ $\phi_j(B)Y_{j,t}=\theta_j(B)\varepsilon_{j,t}$ $(j=1,2)$**, calculate residuals and standardized residuals,**
- **calculate the cross-correlogram between the univariate models residuals and use this to identify the transfer functions** $\omega_j(B)/\delta_j(B)$ **of models linking standardized residuals** $(\varepsilon_{1,t},\varepsilon_{2,t})$**, i.e.:**

  $$\varepsilon_{1,t}=\frac{\omega_1(B)}{\delta_1(B)}\varepsilon_{2,t}+\frac{\theta_1(B)}{\varphi_1(B)}\eta_{1,t},\ \ \varepsilon_{2,t}=\frac{\omega_2(B)}{\delta_2(B)}\varepsilon_{1,t}+\frac{\theta_2(B)}{\varphi_2(B)}\eta_{2,t},$$

  (standarized to have unit variances),
- **identify the error structures:** $(\varepsilon_{1,t},\varepsilon_{2,t})$**, i.e. the forms** $\omega_j(B)/\delta_j(B)$ **and estimate them; check the adequacy of fitted model,**
- **amalgamate the bivariate model fitted to residuals** $(\varepsilon_{1,t},\varepsilon_{2,t})$ **with two univariate models to suggest the bivariate model** $(Y_{1,t},Y_{2,t})$**; estimate the model relating the original series,**
- **check the adequacy of the model and, if necessary, reestimate it.**

• The multivariate model building is much more complex and time-consuming. The model can provide dependencies between two variables and forecasts with lower errors than univariate models.

- **Main groups of non-linear methods:**
  - Autoregressive conditional heteroscedastic models (ARCH),
  - Generalized autoregressive conditional heterosceda stic models (GARCH),
  - State space models,
  - Regime-switching models: threshold autoregressive Markov-switching (TAR, SETAR, STAR, MSW),
  - Artificial networks models.

- ARCH & GARCH models (Engle, Bollerslev), (generalized) autoregressive conditional heteroscedasticity (introductory facts)

• Models which are capable of describing (not only) the feature of volatility clustering, but also other properties of financial time series, e.g. excess kurtosis $\hat{K}_{\hat{\varepsilon}} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^4 \Big/ (\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2)^2$ **(greater than** 3**) or fat-tailedness. Consistent with theory (e.g. Capital Asset-Pricing Model) and empirical evidence; some of the properties important at financial market have models with non-linear form of the predictor.**

• ARCH model (capture the volatility of clustering of TS – large shocks tend to be followed by large shocks)

$Y_t = E(Y_t \,|\, \Omega_{t-1}) + \varepsilon_t$ **,**

$\varepsilon_t = z_t \sqrt{h_t}$ **;** $\quad \sigma^2 \equiv E(\varepsilon_t^2) = E(E(\varepsilon_t^2 \,|\, \Omega_{t-1})) = E(h_t)$

z_t **- iid. standard Gaussian random variable,**

$h_t \equiv h_t(\Omega_{t-1})$ **- a nonnegative function,**

Ω_{t-1} **- information set up to and including time** $t-1$

(distribution of ε_t **conditional upon** Ω_{t-1} **is** $\mathrm{N}(0, h_t)$**), unconditional expectation of** h_t **is constant).**

An alternative representation of ARCH(1) model (conditional variance of shock at time *t* is linear function of squares of past shocks):

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 \quad (\omega > 0,\ 0 \le \alpha_1 < 1),$$
$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + v_t ,$$
$$v_t \equiv \varepsilon_t^2 - h_t = h_t (z_t^2 - 1)$$

i.e. AR model for ε_t^2 (stationary for $0 \le \alpha_1 < 1$)**.**

Some features of the model:

$$\sigma^2 \equiv E(\varepsilon_t^2) = \frac{\omega}{1-\alpha_1},$$
$$E(v_t \mid \Omega_{t-1}) = 0,$$
$$\varepsilon_t^2 = \sigma^2 + \alpha_1 (\varepsilon_{t-1}^2 - \sigma^2) + v_t .$$

If ε_{t-1}^2 **is larger (smaller) than its unconditional expected value** σ^2, ε_t^2 **is expected to be larger (smaller) than** σ^2 **as well.**

The kurtosis of ε_t^2 **always exceeds kurtosis of** z_t^2**.**

• GARCH model (linear, example – order (1, 1)) – adding h_{t-1} to the ARCH (1) model:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (\omega > 0,\ \alpha_1 > 0,\ \beta_1 \geq 0,\ \alpha_1 + \beta_1 < 1),$$

or

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 (\omega + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 h_{t-2}) =$$

$$\sum_{i=1}^{\infty} \beta_1^i \omega + \alpha_1 \sum_{i=1}^{\infty} \beta_i^{i-1} \varepsilon_{t-i}^2 .$$

Some properties of the model.

GARCH(1, 1) corresponds to ARCH(∞) model, makes it possible to replace ARCH with large q; it is an ARMA(1, 1) for ε_t^2 and $\sigma^2 = \frac{\omega}{1-\alpha_1-\beta_1}$.

Kurtosis K_ε is always larger than 3 (Gaussian).

• The parameters of the models have to satisfy restrictions, allowing estimation (ML, QML, robust methods) and diagnostic checking. The models are adequate in practice (financial data) and have many extensions, also nonlinear (have additional properties, e.g. different effects of positive and negative shocks). The validation is based on the properties of $\hat{z}_t = \hat{\varepsilon}_t \hat{h}_t^{-\frac{1}{2}}$ (constant variance, lack of serial correlation, etc.), tested with the use of standard tests. Other properties (parameter constancy) – specialized LM tests.

• The form of forecasting formulas – extension of optimal forecasts for ARMA models, especially for GARCH(1, 1):

$$y_t = \phi_1 y_{t-1} + \varepsilon_t ,$$
$$\hat{h}_{T+s} = \omega + \alpha_1 \hat{\varepsilon}^2_{T+s-1} + \beta_1 \hat{h}_{T+s-1}, \qquad (s \geq 1)$$
$$where: \hat{\varepsilon}^2_{T+i} = \hat{h}_{T+i} \quad for \quad i > 0,$$
$$while: \ \hat{\varepsilon}^2_{T+i} = \varepsilon^2_{T+i}, \hat{h}_{T+i} = h_{T+i} \quad for \ i \leq 0,$$

$$\hat{y}_{T+s} = E(y_{t+s} \mid \Omega_t),$$
$$\hat{y}_{T+s} = \phi_1^s y_t ,$$

the conditional squared prediction error:

$$E(e^2_{T+s} \mid \Omega_t) = \sum_{i=1}^{s} \phi_1^{2(s-i)} \sigma^2 + \sum_{i=1}^{s} \phi_1^{2(s-i)} (E(h_{t+i} \mid \Omega_t) - \sigma^2)$$

(conditional variance is varying over time – larger or smaller than homoscedastic errors),

$$\lim_{s \to \infty} E(e^2_{T+s} \mid \Omega_t) = \frac{\sigma^2}{1-\phi^2} \equiv \sigma^2_y .$$

The interval forecast cannot be determined in „conventional” way, because e_{T+s} is not Gaussian.

Conditional expectation of h_{T+s}:

$$\hat{h}_{T+s} = \omega + \alpha_1 \hat{\varepsilon}^2_{T+s-1} + \beta_1 \hat{h}^2_{T+s-1};$$

if $\alpha_1 + \beta_1 < 1$ **(covariance stationarity) then**

$$\hat{h}_{T+s} = \sigma^2 + (\alpha_1 + \beta_1)^{s-1}(h_{t+1} - \sigma^2); \quad \sigma^2 = \omega / (1 - \alpha_1 - \beta_1);$$

Thus, forecasts for the conditional variance are similar to forecasts from AR(1) model with mean σ^2 **and AR parameter** $\alpha_1 + \beta_1$**.**

For $\alpha_1 + \beta_1 = 1$ **it simplifies to**

$$\hat{h}_{T+s} = \omega(s-1) + h_{t+1}.$$

• General basis of the methods:

THEORY OF GENERALIZED AUTOREGESSIVE CONDITIONAL HETEROSCEDASTICY

• The features of the methods:

- **allow modeling of the processes with: high frequency, asymmetric disturbances, clustering of variances, long memory, heavy tails,**
- **not easy estimation (many parameters) and prediction,**
- **valuable description (analytical) of a phenomena,**
- **extended for effects of positive and negative shocks with the use of regime switching models,**
- **empirical features: often biased forecasts, poor volatility forecasts (large MSPE error), with low R^2, e.g. lower 0,1.**

Advancing analytical form of the model

reflecting nonlinearity, different form of the model in different regimes and allow some dynamic of variables

- <u>Regime switching (non-linear) models (Tong), 2 regimes</u>

*** threshold autoregressive AR(1)**

$$Y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \le c, \\ \varphi_{0,2} + \varphi_{1,2} Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} > c. \end{cases}$$

c – **constant (threshold),** ε_t **- Gaussian iid.**

Features of the model:

- **existence of single equilibrium (stable or not), multiple equilibria or not equilibrium, for** ε_t **equal zero (equilibrium** $y^* = F(y^*)$**),**
- **existence of stable equilibriums – attractors** (y^* *is the attractor for* $\bar{y}$ *if* $y_t = \bar{y}$, *and* $y_{t+h} \to y^*$ *as* $h \to \infty$, *if* $\varepsilon_{t+j} = 0$ *for all* $j > 0$), **also in the form of limit cycles** ($k-$*period cycle* $-$*a set of points* $y_1^*, ..., y_k^*$ *such that* $y_j^* = F(y_{j-1}^*)$, $j = 2, ..., k$ *and* $y_1^{(*)} = F(y_k^*)$) **(endogenous dynamics),** ($F(y_{t-1})$ **- conditional expectation of** y_t **at time** t-1**).**
- **conditions of stationarity - for simple models.**

*** smooth (gradual) transition AR**

$$Y_t = (\varphi_{0,1} + \varphi_{1,1} Y_{t-1})(1 - G(y_{t-1}; \gamma, c)) +$$

$$(\varphi_{0,2} + \varphi_{1,2} Y_{t-1}) G(y_{t-1}; \gamma, c) + \varepsilon_t ,$$

$$G(y_{t-1}; \gamma, c) = \frac{1}{1 + \exp(-\gamma(y_{t-1} - c))},$$

$$G(c; \gamma, c) = 0{,}5$$

(logistic transition function, c – threshold,

γ - smoothness of the change between regimes).

*** Markov – switching**

$$Y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} Y_{t-1} + \varepsilon_t & if \ \ s_t = 1, \\ \varphi_{0,2} + \varphi_{1,2} Y_{t-1} + \varepsilon_t & if \ \ s_t = 2. \end{cases}$$

s_t - unobservable process

$$P(s_t = 1 | s_{t-1} = 1) = p_{11}, P(s_t = 2 | s_{t-1} = 1) = p_{12}, \ldots,$$

$$P(s_t = 2 | s_{t-1} = 2) = p_{22}, \quad p_{11} + p_{21} = 1, \quad p_{21} + p_{22} = 1.$$

unconditional probabilities:

$$P(s_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \quad P(s_t = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}.$$

Specification procedure for regime switching models:

(1) specify order of AR process,

(2) test linearity versus non-linearity,

(3) estimate the parameters,

(4) use the diagnostic tests (also: AIC, BIC criterions),

(5) modify the model if necessary,

(6) use the model for analyzing and forecasting purposes.

• There exist optimal estimators, however much more complex, than in simple linear models; some of them are based on "direct" (numeric) optimization. The predictors for multiple step ahead are based also on Monte Carlo and bootstrap methods. The threshold models allow state-dependent dynamic behavior, different in individual time intervals (Tong, Granger).

• Many other switching models – Markovian regression (Quandt), breaking points of the Markovian trend (L. Klukowski Ph.D.).

• The features of the methods:

- **allow different form of the model in different time intervals,**
- **give information about braking points,**
- **complex in estimation and forecasting, but indispensable; some theoretical issues – no conclusive results,**
- **have been developed in purpose to allow for asymmetric effect of positive and negative shocks using GARCH concept, asymmetric nonlinear smooth transition GARCH, etc.,**
- **empirical features: often biased forecasts, errors greater than from linear models; examples of reasons: spurious non-linearity, different precision in different regimes, long series from one regime.**

- Kalman's filter, dynamic linear model:

• The model

$Y_t = X_t \theta_t + \upsilon_t$: $(\upsilon_t \sim N(0, V_t))$ - the observation equation,

$\theta_t = G\,\theta_{t-1} + w_t$: $(w_t \sim N(0,\ W_t))$ - the systems equation,

θ_t - vector of parameters, X_t, G - matrices.

• Optimal predictor - unbiased, minimal variance.

• The model with non-constant parameters, updated in recurrent, Bayesian way.

• The idea applied in estimation of special cases: Harrison-Stevens, Hamilton (Markovian) models.

The basis of the univariate Harrison-Stevens model:

$$Y_t = \mu_t + e_t ,$$
$$\mu_t = \mu_{t-1} + \beta_t + \gamma_t ,$$
$$\beta_t = \beta_{t-1} + \zeta_t .$$
$$\theta_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}, \; X_t = [1, 0], \; G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

e_t , γ_t , ζ_t **- random variables with the variances** $\sigma_e^2, \sigma_\gamma^2, \sigma_\zeta^2$**.**

The model can be in one of the four states:

- **steady state (level and trend constant), with "normal" value** σ_e^2 **and zeroes for remaining variances,**
- **a step change – with** σ_e^2 **normal,** σ_γ^2 **large and** σ_ζ^2 **zero,**
- **a slope change – with** σ_e^2 **normal,** σ_γ^2 **zero and** σ_ζ^2 **large,**
- **an outlier – with** σ_e^2 **large,** σ_γ^2 **zero and** σ_ζ^2 **zero.**

The forecasts are calculated as weighted average of the four states with weights expressing probabilities that the current observation of the process Y_t **is in each state. The probabilities are updated using Bayesian approach, with some a'priori distribution. The model is based on Kalman filter idea and reflects typical features of some TS; in such the case provides precise forecasts for one step ahead.**

- **The basis:**

THE THEORY OF FILTERING OF STOCHASTIC PROCESSES

- **The features of the method:**
  - **allows changes of deterministic and stochastic parameters of the model,**
  - **possible application to multidimensional series,**
  - **long time series not necessary, as a basis for estimation; instead some parameters are necessary.**

- Artificial neural networks (general evaluation)

• The models can detect and provide truly nonlinear dynamic relationships, e.g. bilinear models $(y_t = \beta\, y_{t-2}\varepsilon_{t-1} + \varepsilon_t)$**; in practice it is often superior in comparison to linear time series models. The drawbacks are that: * the form and parameters are difficult to interpret, * the superior in sample fit does not guarantee out-of-sample forecasting, * danger of over-fitting.**

• Basis:

THEORY OF ARTIFICIAL NEURONAL NETWORKS

• The model building:

*** choosing the activation function,**

*** choosing the number of hidden units,**

*** choosing the number of lags to use as input variables,**

The selection criterion: AIC, BIC, genetic algorithm, etc.

• The forecasts, in point and interval form, are easy to obtain and evaluate for one step ahead, for multi-steps – more complicated.

• The simulation experiments show, that the neural models do not capture some types of no-linearity, e.g. Markov-switching or GARCH.

• The features of the methods:

- **current, universal, non-linear, optimal tools based on computer technology,**
- **broad spectrum of application,**
- **not for uncritically application, often fitting to the series much more precise than forecasts.**

3. Some comparisons of empirical forecasts

• The accuracy of empirical forecasts depends on many factors, especially:

- **properties of the model – theoretical and empirical (ability to reflect properties of actual phenomena),**
- **number of elements of time series,**
- **properties of time series (variability, non-linearity outliers),**
- **properties of estimators (precision, robustness),**
- **experience level of researchers.**

• In general the best forecasts result from "superior" models, which are estimated and predicted in optimal way. Such the properties have majority of the models presented: ARIMA, Kalman's filter, ARCH, GARCH, regime switching models (threshold, Markovian), artificial neural networks, multivariate models. However, application of advanced methods is not easy - needs: experience, long series (exception is Kalman's filter) and significant computational effort. Therefore simple, and also heuristic methods, are still in use.
Good compromise ensures combining approach.

4. Summary and conclusions

• The amount of time series models for modeling and forecasting of time series is currently enormous. Majority of them have been developed in last decades.

• The main factors, stimulating the development were: complexity of phenomena, theoretical progress and efficiency of computing systems.

• Further development allows applications in broader area, higher precision and longer distant. Importance of accurate forecasts increases in competitive economy.

• Therefore, it is a field of intensive researches.

Basic literature

Anderson T.W. (1971): *THE STATISTICAL ANALYSIS OF TIME SERIES*. J. Wiley, New York.

Box, G.E.P., Jenkins G.M. (1976): *TIME SERIES: ANALYSIS, FORECASTING AND CONTROL*, San Francisco: Holden Day.

Charemza W.W., Deadman D.F. (1997): *NOWA EKONOMETRIA*. PWE, Warszawa 1997.

Franses, P.H., van Dijk D. (2000): *NON-LINEAR TIME SERIES MODELS IN EMPIRICAL FINANCE*, Cambridge University Press.

Granger, C.W.J., Newbold, P. (1986): *FORECASTING ECONOMIC TIME SERIES*, 2nd ed., London: Academic Press.

Klukowski, L. (1993): *Time series forecasting on the basis of Markovian piecewise linear trend models*. Control and Cybernetics vol. 22, No. ½.

Hamilton, J., D. (1994): *TIME SERIES ANALYSIS*. Princeton University Press.

Montgomery D.C., Jennings L.J., Kulahci M. (2008): *INTRO DUCTION TO TIME SERIES ANALYSIS AND FORECASTING*. J. WILEY-INTERSCIENCE, Hoboken, New Jersey.

Tong, H. (1983): *THRESHOLD MODELS IN NON-LINEAR TIME SERIES ANALYSIS*, Springer, Lecture Notes in Statistics, Vol. 21.

West M., Harrison J. (1997): *BAYESIAN FORECASTING AND DYNAMIC MODELS*. Springer 1997, New York, IInd ed.

Thank you for your attention!