Time Series Forecasting

Henrika Langen

Central European University
Department of Economics and Business

Winter 2024

Structure of the Lecture

1. Introduction to time series
2. Modeling time series and evaluating forecasting performance
3. Exponential smoothing models
4. ARIMA models
5. **Regression models and advanced forecasting models**
6. Machine Learning and Deep Learning methods

Time Series Regression Models

Topics

- Linear regression models
- Dynamic regression models
- Complex seasonality
- State space models
- Prophet model

Time Series Regression Models

- Time series regression models allow to integrate information from other series x_t into the modeling of the series of interest y_t.
- The underlying assumption is that y_t and x_t have a linear relationship.
- The x_t can measure other phenomena related to y_t but also deterministic series that help to model seasonality, calendrical effect and trends.

Linear Time Series Regression Models

A **simple linear regression model** is given by the following equation

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$

where ε_t is the error term, i.e., the deviation from the linear model in t.

- The coefficients can be interpreted just as in simple linear regression with cross-sectional data: An increase in x_t by one unit leads to a change in y_t by β_1 units.
- The variables y_t and/or x_t can be a non-linear transformation (e.g. logarithmic) of the original variables if the relationship between the transformed series is linear, while that of the original series is not.

Linear Time Series Regression Models

A **multiple linear regression model** is given by the following equation

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t ,$$

where ε_t is the error term, i.e., the deviation from the linear model in t.

- The coefficients can be interpreted just as in multiple linear regression with cross-sectional data: The coefficients $\beta_1, \ldots, \beta_k$ measure how y_t changes in response to a change in one predictor after taking into account the influence of the other predictors.
- The variables y_t and/or $x_{1,t}, \ldots, x_{k,t}$ can be a non-linear transformation (e.g. logarithmic) of the original variables if the relationship between the transformed series is linear, while that of the original series is not.

Underlying assumptions

The models build on the following assumptions about the errors $(\varepsilon_1, ..., \varepsilon_T)$:

- They have mean zero: $E(\varepsilon_t) = 0$
- They are not autocorrelated.
- They are unrelated to the predictors in the model.
- In order to easily produce prediction intervals, they should be normally distributed and have constant variance σ^2.

$\Rightarrow$ These assumptions are very similar to the assumptions underlying linear regression models for cross-sectional data. Only difference: Here, autocorrelation is an issue and for many time series it's not plausible to assume there's none $\Rightarrow$ in these cases, the error could be modeled with an ARIMA model (see section on dynamic regression models)

Estimation

Estimation based on Least Squares, i.e., by choosing $\beta_0, \ldots, \beta_k$ that minimize the squared error:

$$\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \ldots - \beta_k x_{k,t})^2$$

Like in regressions with cross-sectional data, common measures for goodness-of-fit can be employed like e.g.:

$$R^2 = \frac{\sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}, \qquad (0 \leq R^2 \leq 1)$$

measuring the proportion of variation in y_t that is captured by the model.

$\Rightarrow$ R^2 should be calculated in test data to avoid overfitting.

Model evaluation

Certain checks should be run to evaluate if the model assumptions hold (again similar to linear regression with cross-sectional data), namely:

- Create a histogram of the residuals $e_t = y_t - \hat{y}_t$: Are the e_t approximately normally distributed?
- Plot the e_t against predictors: If there is a pattern in any plot, the relationship between y_t and the respective predictor may be non-linear.

Model evaluation:
Additionally, certain checks particular to time series regression should be employed.
Check for **autocorrelation in the residuals**:

- Plot the residual ACF.
- Run formal tests to test for autocorrelation.

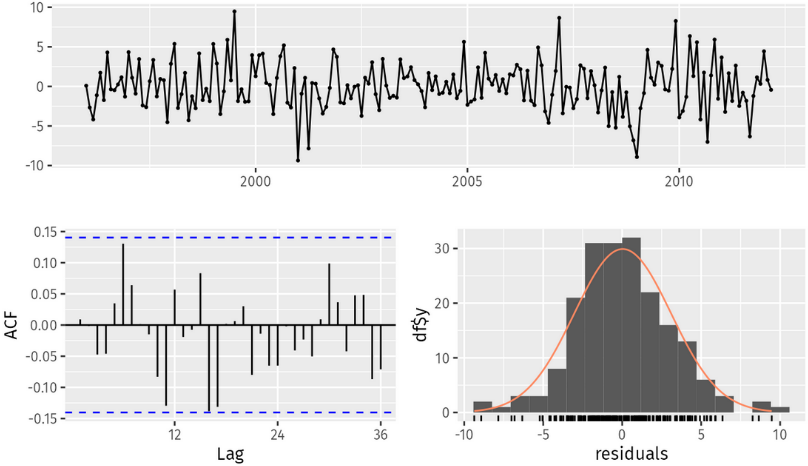
$\Rightarrow$ Autocorrelation in the residuals indicates that the model is inefficient as it doesn't account for all information available. The model is unbiased but the prediction intervals are unnecessarily wide.

Check for **time-dependent heteroscedasticity**:

- Create a residual time plot: Is the residual variance changing over time?

$\Rightarrow$ In case of heteroscedasticity, the model is unbiased but the prediction intervals may be wrong.

Time Series Regression Models



Residual diagnostics for a model fit to a time series of electrical equipment orders.

Source: FPP

Spurious regression

- Spurious relations are apparent connections/correlations between variables (or time series) that are actually coincidental and not indicative of any true relationship.
- In time series regression, there is a particularly high risk of modeling spurious relationships when dealing with non-stationary, and especially with trending, series because technically all series with a similar trend are highly correlated whether there is an underlying theoretical relationship or not.

⇒ Important to consider whether a relationship between y_t and potential predictors is theoretically plausible. Additionally, high R^2 combined with high autocorrelation in residuals can be an indicator of spurious relations.

Potential predictors

- a **time series z_t that is correlated with y_t**, e.g. to model ice-cream sales y_t one can include weather z_t as predictor
- lagged values of a series, e.g. if z_t measures the advertisement expenditures, $x_{i,t} = z_{t-1}, x_{i+1,t} = z_{t-2}, \ldots$ can be included in the model to capture the **longer-term effect** of advertisement
- a variable $x_{i,t} = t$ with $t = 1, \ldots, T$ can be included to capture a **linear trend** in the series
- a series of **seasonal dummies**, e.g. for a quarterly series: $d_{Q1,t}, d_{Q2,t}, d_{Q3,t}, d_{Q4,t}$, where $d_{Q1,t} = 1$ if t falls into the first quarter and 0 otherwise, $d_{Q2,t} = 1$ if t falls in the first quarter and 0 otherwise,...

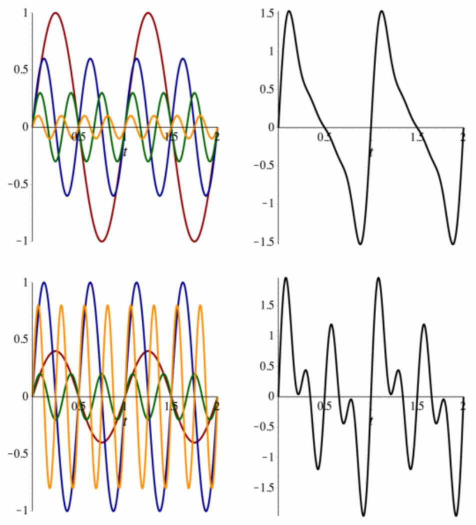
Potential predictors

- a series of a dummy variable that e.g. ...
  - ... measures whether a day t is a working day (=1) or not (=0)
  - ... accounts for outliers. It takes on the value 1 for observations that coincide with a special event like a lockdown or strike period and 0 elsewhere
  - ... captures the effect of an intervention (e.g. marketing campaigns). This variable can either only equal 1 during the time of the intervention, or - if permanent effect of the intervention is assumed - be 0 before and 1 after the intervention.
  - ... captures if a certain holiday/event (e.g. Easter) falls into t.
- a series of the number of trading/business days in t

Potential predictors

- For high seasonal frequency (e.g. $m = 52$ weekly or $m = 365$ daily data), one can include Fourier terms to capture seasonal patterns instead of including a dummy for each season.
- Fourier series are a tool for expressing an arbitrary periodic function (like the seasonality pattern of a time series) as a weighted sum of sine and cosine terms.

Potential predictors - Fourier terms



Source: https://www.physics.uoguelph.ca/chapter-7-fourier-series

Potential predictors - Fourier terms

- The Fourier terms are given by:

$$x_{1,t} = sin\left(\frac{2\pi t}{m}\right), \quad x_{2,t} = cos\left(\frac{2\pi t}{m}\right), \quad x_{3,t} = sin\left(\frac{4\pi t}{m}\right),$$
$$x_{4,t} = cos\left(\frac{4\pi t}{m}\right), \quad x_{5,t} = sin\left(\frac{6\pi t}{m}\right), \quad x_{6,t} = cos\left(\frac{6\pi t}{m}\right), \ldots$$

- The terms can be included in pairs of one sine and one cosine term, i.e., either the first two, the first four,..., up to the first m.
- Usually a few Fourier terms are sufficient to capture the seasonality of a time series, making them particularly advantageous for series with large m.

Selecting predictors

Once all potential predictors are identified, we must identify the set of predictors that yields the highest prediction accuracy when included in the model.

There are two commonly used approaches to screen through all possible sets of predictors:

- **Best subset regression**: compare the prediction accuracy of all possible models $\Rightarrow$ if computationally not to challenging, this approach should be preferred.
- **Backwards step-wise regression**: Start with a model containing all potential predictors, remove one predictor at a time and keep the model with the highest predictive accuracy.

Selecting predictors

Measures/approaches to compare candidate models with different sets of predictors:

- The **adjusted** R^2, denoted as $\overline{R}^2$, is a modified version of R^2 that penalizes for the number of predictors k, i.e. it accounts for the model's complexity and prevents overfitting:

$$\overline{R}^2 = 1 - (1 - R^2)\frac{T-1}{T-k-1}$$

- The information criteria AIC, AIC$_c$ and BIC can be used as discussed in Session 2.
- **Cross validation** as discussed in Session 2.

Forecasting

To forecast y_{T+h} with linear time series regression models, we have to plug the values for $x_{1,T+h},\dots,x_{k,T+h}$ into the estimated model.

- For $x_{i,t}$ that capture deterministic features like calendrical specificities (number of business days, day-of-the-week dummies) or seasonality (Fourier terms or seasonality dummies), these values are known for any $T+h$.
- For $x_{i,t}$ that measure lagged values of a variable, future values are known as long as the h is not larger than the lag.
- For $x_{i,t}$ that measure other variables, we can simply set $x_{i,T+H}$ to a likely values, run through different scenarios, use available forecasts (for weather, economic growth,...) or predict the predictors using any suitable time series forecasting approach.

Forecasting - Prediction intervals

The width of the prediction intervals for the forecast $\hat{y}_{T+h}$ depends on the residual variance and on how far away the predictors are from their average.

E.g. for a simple linear regression model, the 95%-prediction interval for $\hat{y}_{T+h}$ is given by:

$$\hat{y}_{T+h} \pm 1.96\hat{\sigma}_e \sqrt{1 + \frac{1}{T} + \frac{(x_{T+h} - \overline{x})^2}{(T-1)s_x^2}}$$

where $\hat{\sigma}_e = \sqrt{\frac{1}{T-k-1} \sum_{t=1}^{T} e_t^2}$ is the standard error of the regression and s_x^2 is the standard deviation of x.

Careful: the prediction interval does not take into account possible uncertainty in predicting the predictors.

Dynamic Regression Models

- Dynamic regression models extent ARIMA models by including predictors x_t, or, put differently, dynamic regression models model the potential autoregression in the errors of a regression model.
- E.g.: A regression model where the error term η_t follows an ARIMA(1,1,1) model is given by:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{1,t} + \ldots + \beta_k x_{k,t} + \eta_t \\ (1-\phi_1 B)(1-B)\eta_t &= (1+\theta_1 B)\varepsilon_t, \end{aligned}$$

  where ε_t is white noise.
- If the above model was the true model underlying a time series, simply estimating the regression model while ignoring the autoregressive error term would result in ..
  - ... the model not being the best option given the available information and
  - ... in incorrect prediction intervals.

Estimation:

- To estimate a regression model with ARIMA errors, y_t and all $x_{1,t}, \ldots, x_{k,t}$ must first be made stationary
  $\Rightarrow$ Apply the same transformations to all series to maintain the relationship between variables and ensure interpretability of the results.
- To determine the optimal set of predictors, compare the AIC$_c$ of all candidate models.

Forecasting

- Forecast the regression part and the ARMA part of the model separately and combine it.
- If the predictors are unknown in $T + h$, they may be forecasted before being plugged into the regression model or set to specific values (Careful: the prediction intervals do not take into account the uncertainty in predicting the predictors)

Dynamic Regression Models

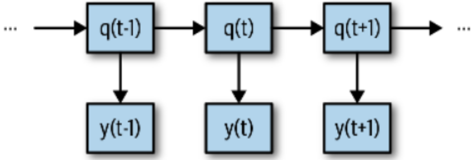
Advantages and disadvantages of dynamic regression models

+ Dynamic regression models allow us to account for autocorrelation in the data and at the same time take advantage of relationships between y_t and other series x_t.
+ They are very flexible in capturing seasonality, calendar peculiarities and events.
- Dynamic regression models only capture linear relationships.

Underlying Idea

- There is an unobservable system that determines the development of the observed time series due to internal dynamics and external forces that affect the system.
- In every t, the system is in a certain state but we cannot observe the true state directly - we observe the state with some noise added
- The aim of state space modeling is to model the dynamic system producing the data we observe rather than the noisy data itself.
    - State space models are particularly suitable when we have strong theories/knowledge about how the underlying system works.
- The model coefficients can change over time as the model is constantly updated based on the most recent data points.

Underlying Idea



Source: PTSA

Example: Hidden Markov Model

- **Assumption about the underlying process:**
  - The states we observe are generated by a **Markov process that jumps between discrete states**.
  - Markov processes are memory-less, i.e., the state of the process in $t+1$ depends only on the state in t.

  $\Rightarrow$ The dynamics within a Markov process can be fully described based on the transition probabilities between any t and $t+1$, which are usually represented in matrix form. E.g.:

$$\begin{array}{ccc} & A & B \\ A & 0.6 & 0.4 \\ B & 0.3 & 0.7 \end{array}$$

- **Assumption about the observed series:** We cannot observe which state the system is in but the observations give clues about it.
- **Potential Applications:** Modeling an economic time series to capture underlying economic states and transitions, or financial series to identify regime shifts in the financial market.

Example: Hidden Markov Model - Fitting the model

Fitting a Hidden Markov Model to a time series is an unsupervised learning task (we have to identify the hidden states without knowing any of them a priori).

Procedure:

1. Make assumptions about the number of underlying states and the relationship between the true states $q_1, ...q_T$ and the observed series $y_1, ...y_T$.

   E.g.: 2 states produce a normally distributed output (the values we observe, y_t). Each state q_i is characterized by mean μ_i and standard deviation σ_i. The probability to change from one state to the other is p.

2. Set the parameters that characterize the states randomly (i.e., μ_i, σ_i, p in the example).
3. For each t, determine which state is most likely to have produced y_t.
4. Update the parameters based on the identified series of states.
5. Repeat steps 2 and 3 until convergence is achieved.

Complex Seasonality

Time series with high frequency often exhibit complex seasonal patterns:

- They exhibit more than one seasonal pattern. E.g.: Daily data may have a weekly and a yearly seasonal pattern.
- They may have a seasonal pattern with non-integer m like e.g. for weekly data with annual seasonality $m = 365.25/7 \approx 52.179$

Dealing with complex seasonality - (TBATS)

- Automated modeling framework that allows for a slowly changing seasonality pattern
- Procedure: The time series is Box-Cox transformed to decrease its variability, and is then modeled as a linear combination of an exponentially smoothed trend, a seasonal component (modeled with Fourier terms) and an ARMA remainder component.

Dealing with complex seasonality - Fourier Terms

- Include pairs of sine and cosine terms for each seasonality pattern in a dynamic harmonic regression model. E.g. For daily data:

$$sin\left(\frac{2\pi t}{52}\right), \quad cos\left(\frac{2\pi t}{52}\right), \quad sin\left(\frac{2\pi t}{365}\right), \quad cos\left(\frac{2\pi t}{365}\right),$$
$$sin\left(\frac{4\pi t}{52}\right), \quad cos\left(\frac{4\pi t}{52}\right), \quad sin\left(\frac{4\pi t}{365}\right), \quad cos\left(\frac{4\pi t}{365}\right),$$
$$, \ldots$$

- Fourier terms also allow for non-integer m

Prophet Model

- The Prophet Model is an automated time series forecasting package developed by Meta based on additive decomposition.
- The trend is modeled with a non-parametric regression model $\Rightarrow$ The model makes no assumptions about the form of the trend curve, and in particular does not impose linearity.
- The model captures yearly, weekly and daily seasonal patterns via Fourier terms, and can account for holidays.
- The model integrates mechanism to automatically handle missing data.
- It allows for the integration of additional predictors x_t as in time series regression.

Prophet Model

Advantages and disadvantages of the Prophet model and what data to use it for

+ Prophet is user-friendly and easy to interpret.
+ The model works particularly well with time series that exhibit strong seasonal patterns (also if different seasonal patterns occur simultaneously).
+ The model is relatively robust to outliers and automatically deals with missing values.
+ It can model non-linear trends and changing points well.
- The model needs a long time series to be trained on.
- Prophet does not account for autocorrelation in residuals $\Rightarrow$ In many studies it is outperformed by ARIMA models.
- The types of seasonality captured are limited to yearly, weekly and daily seasonal patterns. Furthermore, the model was originally developed for and works best for daily data. With other types of data, the model‘s prediction accuracy might suffer.