Ph.D. Program in Transportation Systems

Transport Demand Modeling

Filipe Moura

Session 1

Presentation, Objectives and evaluation methods
What is a model? Application limits and interpretation

(Aknowledgements to Prof. João Abreu e Silva who prepared these slides)

Objectives

- Be capable of understanding and applying the quantitative transport modeling techniques
- Know how and when to apply the different modeling techniques
- Being able to analyze and communicate the results obtained by each modeling technique
- Being able to apply the results obtained in the elaboration and design of transport public policies and corresponding evaluation

Learning Materials

- The learning materials are available in the course webpage:

  https://fenix.tecnico.ulisboa.pt/disciplinas/MPTra9/2016-2017/1-semestre

The following elements will be available

- Lecture slides
- Recommended Readings
- Examples and Problems Databases
- Evaluation rules

Evaluation process

- **Continuous evaluation (60%):**
  - Resolution of 5 exercises (one for each topic) in groups of two students;
    - Exercise 1 – Linear Regression;
    - Exercise 2 – Generalized Linear Models;
    - Exercise 3 – Panel Models;
    - Exercise 4 – Discrete Choice Models.

- **Final exam (40%):**
  - Covering all the subjects. Main focus on theoretical, methodological and conceptual aspects and also in interpretation and communication of results obtained using modeling techniques. Minimum grade in the exam is 8.0/20.0.

Lectures: Detailed Program

Day	starting hour	duration (h)	Subjects	Faculty
Oct. 2nd	9h-11h	2	Presentation, Objectives and evaluation methods. What is a model? Application Limits and intepretation	Filipe Moura (IST)
Oct. 2nd	11h-13h	2	Sampling and statistical tests	Filipe Moura (IST)
Oct. 9th	9h-13h	4	Multiple Linear Regression	Filipe Moura (IST)
Oct. 16th	10h-13h	3	Factor Analysis	Filipe Moura (IST)
Oct. 22nd	10h-13h	3	Cluster Analysis	Filipe Moura (IST)
Oct. 29th	10h-13h	3	Generalized Linear Models	Filipe Moura (IST)
Oct. 31st	10h-13h	3	Generalized Linear Models	Filipe Moura (IST)
Nov. 2nd	9h-13h	4	Panel Models	João Zeferino (FCTUC)
Nov. 2nd	14h-16h	2	Panel Models	João Zeferino (FCTUC)
Nov. 6th	9h-13h	4	Discrete Choice Models	Filipe Moura (IST)
Nov. 9th	14h-18h	4	Discrete Choice Models	Filipe Moura (IST)
Nov. 30th	9h-13h	4	Hazard-Based Duration Models	Filipe Moura (IST)
Dec. 5th.	9h-13h	4	Ordered Choice Models	Filipe Moura (IST)

Faculty

- Faculties:
  - **Filipe Moura (coordinator)** (ISTUL) - fmoura@tecnico.ulisboa.pt
    - https://fenix.tecnico.ulisboa.pt/homepage/ist14188
  - João Zeferino (FCTUC) - zeferino@dec.uc.pt
    - https://apps.uc.pt/mypage/faculty/uc41591

Software needed

- Microsoft Office
- Other Softwares:
  - SPSS: https://delta.ist.utl.pt/software/spss.php
  - NLOGIT/Limdep: cd rom available on request

Final Notes

- Classes include a pratical component where faculties will present softwares and case studies for the home assignment exercises.
- Therefore, students should:
  - Bring their laptos (at least 1 per group of 2 students)
  - Install the softwares before lectures and verify that they are operating properly
  - Donwload the databases required for the exercises before classes.

What is a model? Application limits and interpretation

What is a model? Definition

- A model could be defined as a simplified representation of the real world.
  - It concentrates several elements considered relevant for its analysis according to a specific point of view.

- Another definition
  - Formal expression of one theory, of a causal or associative relationship between variables which is assumed by the analyst as being generated by the observed data.

Essencial properties

- Mapping – A model is based in one original reality.
  - In mathematics, it is an operation that associates each element of a given set with one or more elements of a second set.
- Reduction – The model reflects only a selection (expectably relevant) of the original properties.
- Pragmatism – A model has to be useful for a specific purpose.

Models and theory (I)

- Modeling is an important part of decision processes
- Intuitively we all have mental models on how the world works based on which we make our decisions
- We simplify the reality and make it more abstract and simpler so we can make our decisions
- A model could be descriptive but is commonly used with the objectives of looking for links between causes and effects

Models and theory (II)

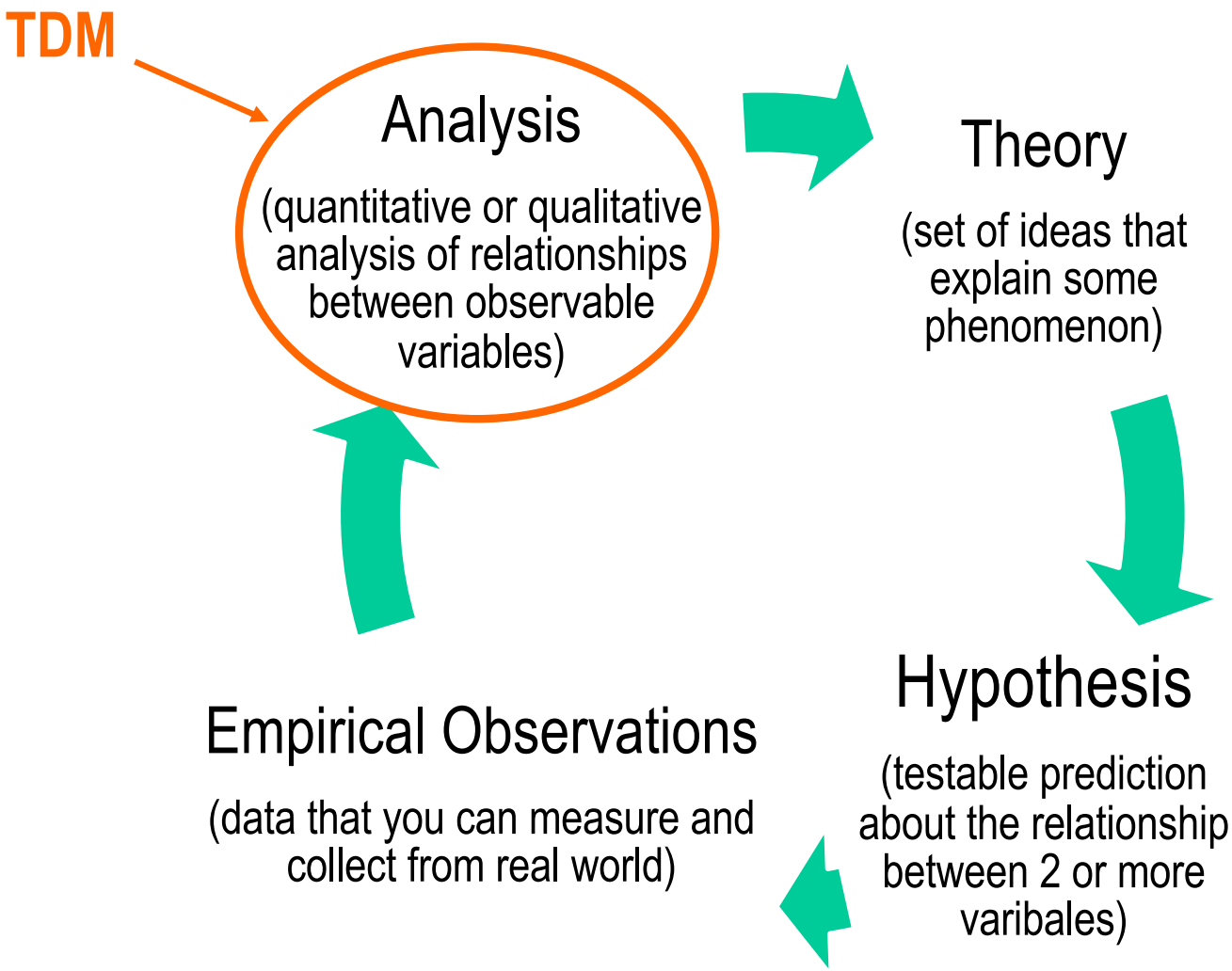
- Most models are based on the premises that by observing individuals and systems behavior – past and present – we can infer the rules that determine that behavior.

- These rules could then be used to predict behavior in similar situations (either in the future or in different places)

Models and theory (III)

- The validity of a model depends on the context for which it was conceived.
- A theory could be built using two different approaches:
  - Inductive – you begin with a research question and the collection of empirical data, which are used to generate an hypothesis and theory.
    - More common in social sciences
  - Deductive – you begin with a theory-driven hypothesis, which guide data collection and analysis.
    - Natural sciences
- Data is central
  - Data availability (quantity and quality) restricts the type of modeling approaches.
  - Avoid "Garbage in, Garbage out", by performing data analysis before starting modeling

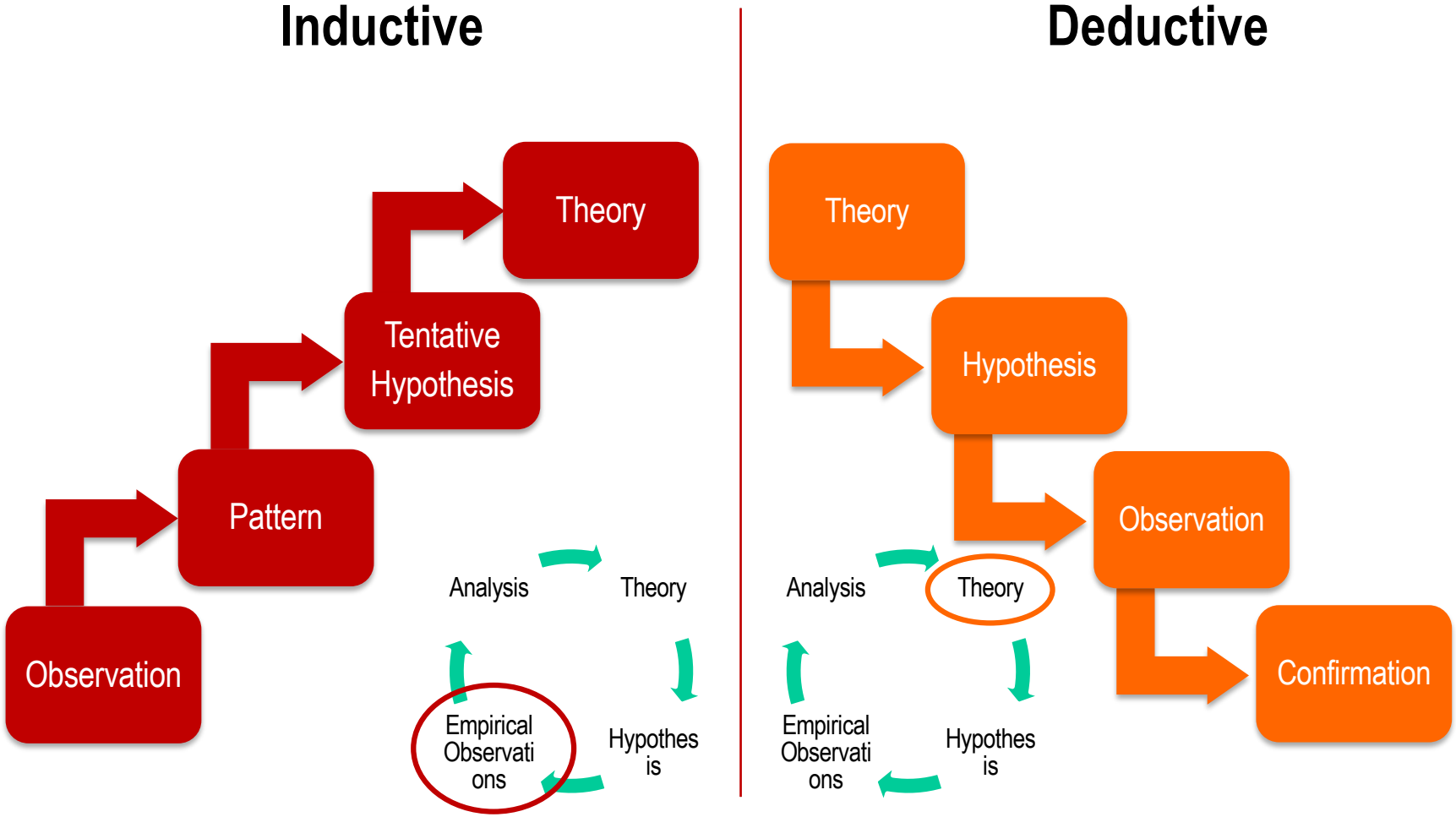
Research Circle



FEUP

Inductive versus deductive research (I)

- Depends mostly where you start in the research circle



Inductive versus deductive research (II)

EXAMPLES

- **Inductive research**
  - **How will students adopt the universities' bike-sharing system? What do they perceive as their main barriers to adoption?**
  - You begin with a **survey**, **analyze** data and identify patterns, build an tentative **hypothesis** (review it, possibly), and formulate a **theory**
- **Deductive research**
  - **Does 'Mobility-as-a-service' (MaaS) induce modal shift away from private car use in urban areas, towards a more sustainable urban mobility system?**
  - You begin with a **theory** of urban mobility, whereby MaaS is seen as the next disruptive concept in urban mobility management, formulate an **hypothesis**, collect **data**, and **confirm** your hypothesis.

FEUP

Models and Variables

- A model is constituted by exogenous (independent) and endogenous (dependent) variables.

$$y = f(x_1, x_2, x_3, \ldots, x_n)$$

- Policy variables must be the ones that could be controlled by the decision maker, if your model aims to support decision-making
  - In transportation systems, it is often the case.

Data aggregation

- How many different population segments do we need to have a good representation of reality and understand it?
- What is the level of detail for measuring a specific variable? (in order for it to replicate adequately a specific phenomena)
- Space – spatial aggregation – level of detail?

Parsimony

- Philosophical principle used in scientific research and mathematic modeling.

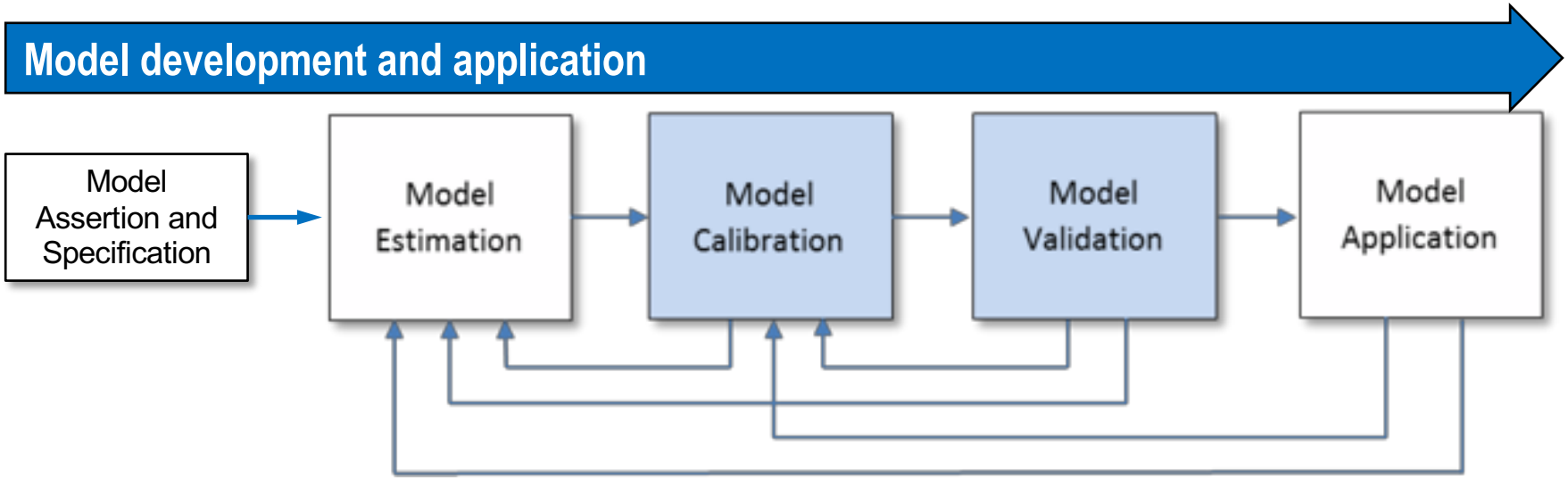
- Proposed by William Occam in the XVII century (Occam´s Razor).

Entities are not to be multiplied without necessity

Pluralities ought not be posited (assumed) without necessity

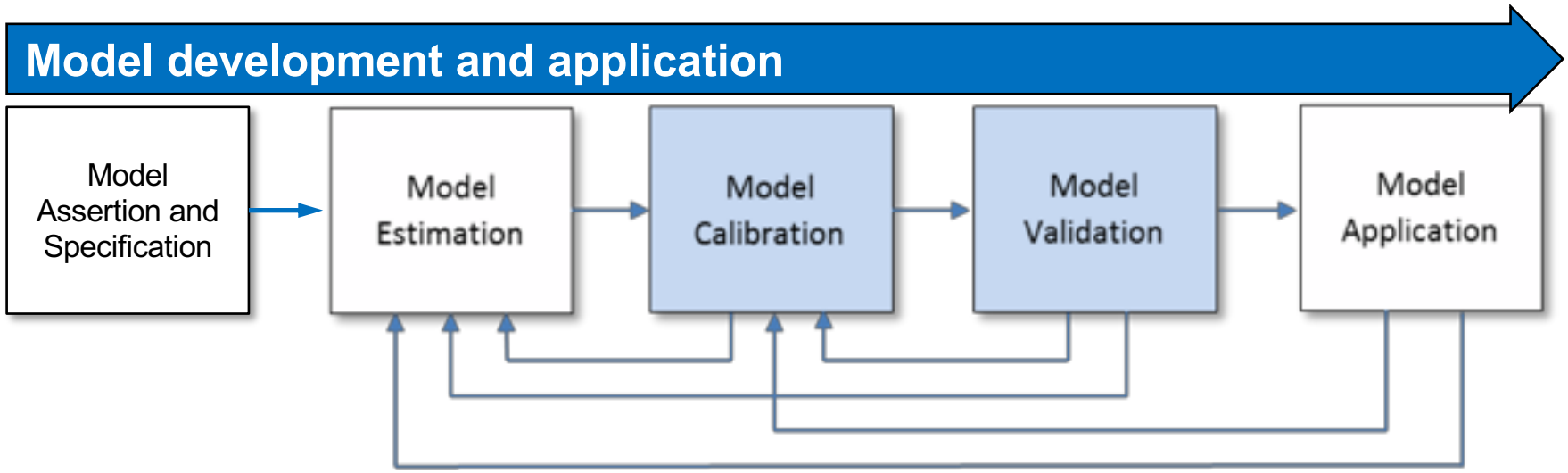
- A model able to explain more variation with a simpler form is the preferred one – Parsimony principle.

Model estimation and application process (I)



- **Model assertion** is the declaration of model forms or parameters without the use of statistical analysis of observed data. Based on this assertion, the model is then specified.
- **Model estimation** is the use of statistical analysis techniques and observed data to develop model parameters or coefficients (estimation may include calibration)
- **Model calibration** is the adjustment of constants and other model parameters *("coefficient tweaking")* in estimated or asserted models in an effort to make the models replicate observed data for a base year or otherwise produce more reasonable results

Model development and application process (II)



- **Model validation** is the application of the calibrated models and comparison of the results against observed data
  - ✓ Ideally, the observed data are not the same data used for model estimation or calibration (**reserve 20% of the sample data for validation!**)
- **Model sensitivity testing** is the application of the models and the model set using alternative input data or assumptions to determine if the model results are plausible and reasonable

Model specification

- **Model specification** – which rules should be included in the model:
  - **Model structure** – Simple versus complex structure;
  - **Functional form** - Linear? Non-linear?
    - Trade-off between realism/precision and ability to calibrate the algebraic relation between the dependent and independent variables;
  - **Variables specification** – How are variables are entered in the model?
    - Should these be transformed?

Model estimation

Model estimation

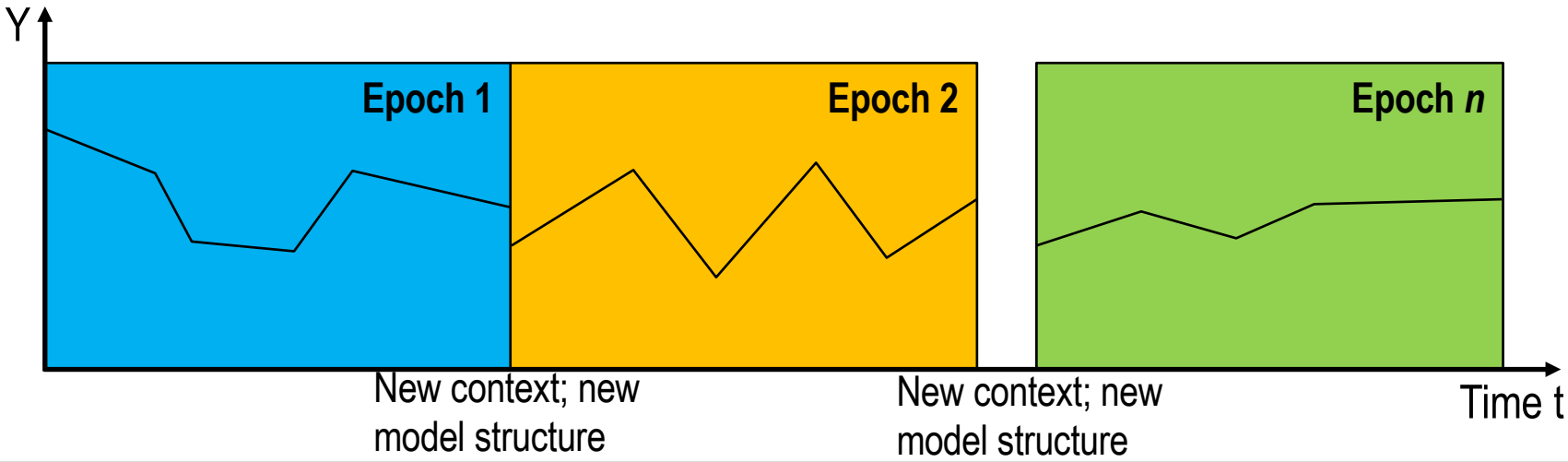
- Estimating parameter values that increase the models probability of reproducing the observed data, that is to minimize/maximize a function of the prediction errors of the model compared to some pre-specified targets (observed values).
- When the model is acceptable (statistically speaking), the parameters should be considered as statistically significant.
- One parameter not statistically significant has a strong probability of being equal to zero.

Model validation

- Validation means achieving a good estimation or calibration and compare the model results against data not used in these processes.
  - For example, saving 20% of data sample for validation
- Validation also implies that the parameters are in accordance with theory (or mental model):
  - Magnitude (comparing with similar models);
  - Effects direction (algebraic sign).

Forecasts

- Forecasting is to predict or estimate (a future event or trend).
- Models forecasts are conditional, since it produces estimates based on a group of forecasted independent variables:
  - Relative to the values of the policy variables;
  - Relative to the values assumed for the other independent variables.
- In general, forecasts are “epoch” specific, i.e. no radical changes/ breakthroughs are expected to occur within the forecasting horizon, compared with the period of observed data.



What is a good model (I)?

- **Precise** – higher levels of precision imply more costs.
  - Caution against apparent higher precision which is generally spurious – **overfitting**
  - When overfitting, a statistical model describes random error or noise instead of the underlying relationship.
  - It occurs when a model is excessively complex, such as having too many parameters relative to the number of observations.
  - A model that has been overfit has poor predictive performance, as it overreacts to minor fluctuations in the training data.
- **Economic** in the use of resources (data and computation capacity)

What is a good model (II)?

- Capacity to produce **relevant indicators** with an adequate disaggregation level
- Capacity to reproduce the **relevant interaction processes**
- **Adequate** in terms of geographical and temporal scopes
- **Transparent** and **friendly**
  - Allowing a more precise evaluation about its results plausibility.

Model evaluation (I)

- How to evaluate a model?
  - By its capacity to increase our understanding and help explaining behavior?
  - By its capacity to generate good forecasts?

- Good predictions could be the result of chance.
  - Be critical on the underlying phenomenon you are modelling by analyzing the validated variables and parameters
- There is always a pressure to have good calibration results (sometimes overfitting) and only **afterwards** rationalize (understand) the relations present in the model

Model evaluation (II)

- There are very few *ex-post* analyses about models predictions
  - The ones that exist tend not to show good results…
  - This doesn´t happen only in the public sector.
- Why?
  - Incorrect estimations for future values of independent variables
    - Out of epoch of observed data
  - Prediction capacity is sometimes viewed as more relevant than causality
  - Incomplete information of future context variables (e.g., new road construction)

Recommended readings

- Juan de Dios Ortúzar, Luis G. Willumsen (2001) “Modeling Transport (3rd edition)”, Wiley and Sons - Chapter 1
- Kenneth Button, David Hensher (2000) “Handbook of Transport Modelling”, Pergamon Press – Chapter 1
- Kuhne, Thomas, (2005), What is a model?, Dagstuhl Seminar Proceedings 04101.