Evidence-Based Transportation Demand Analysis

Evidence, Model & Parameter Estimation

Khandker Nurul Habib, PhD. PEng

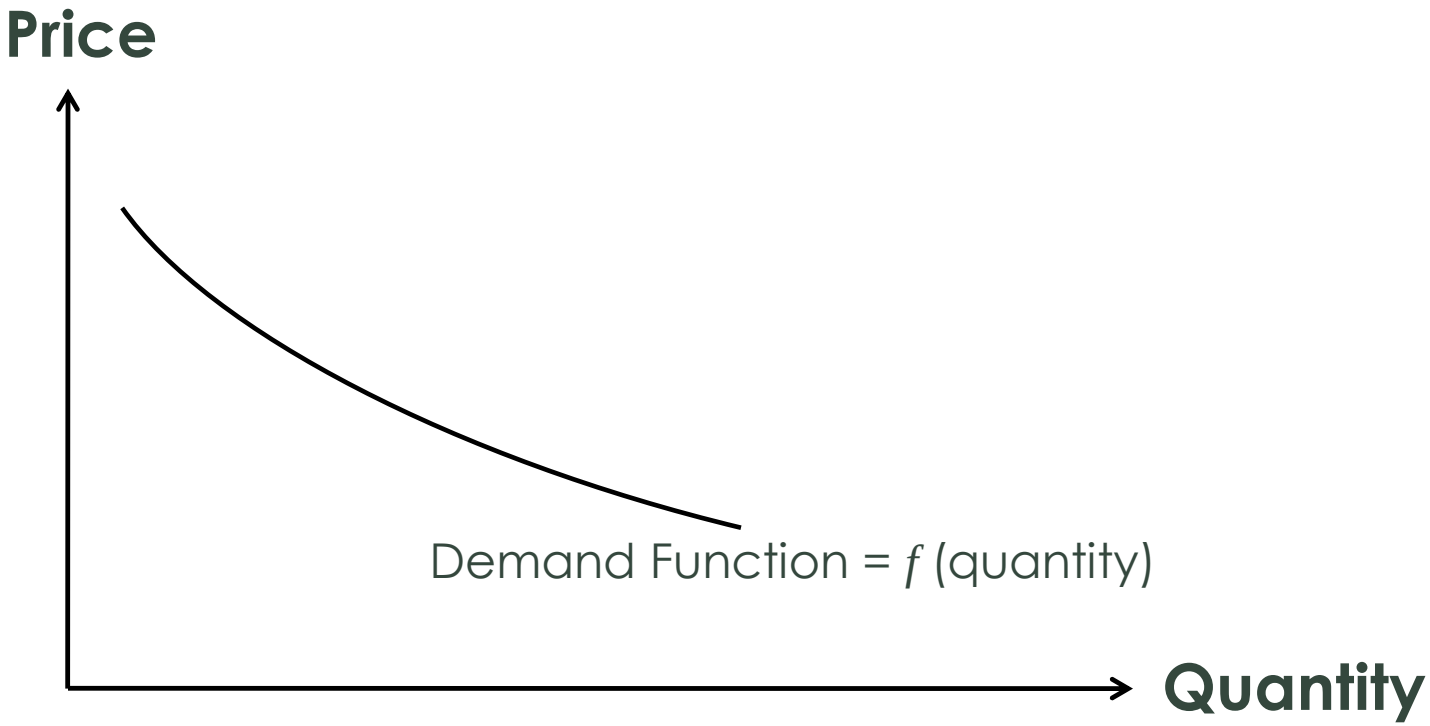
Outline

- Economic concept of transportation
- Data/Evidence and errors
- Sample versus population
- Econometric model

Economics of Transportation Demand

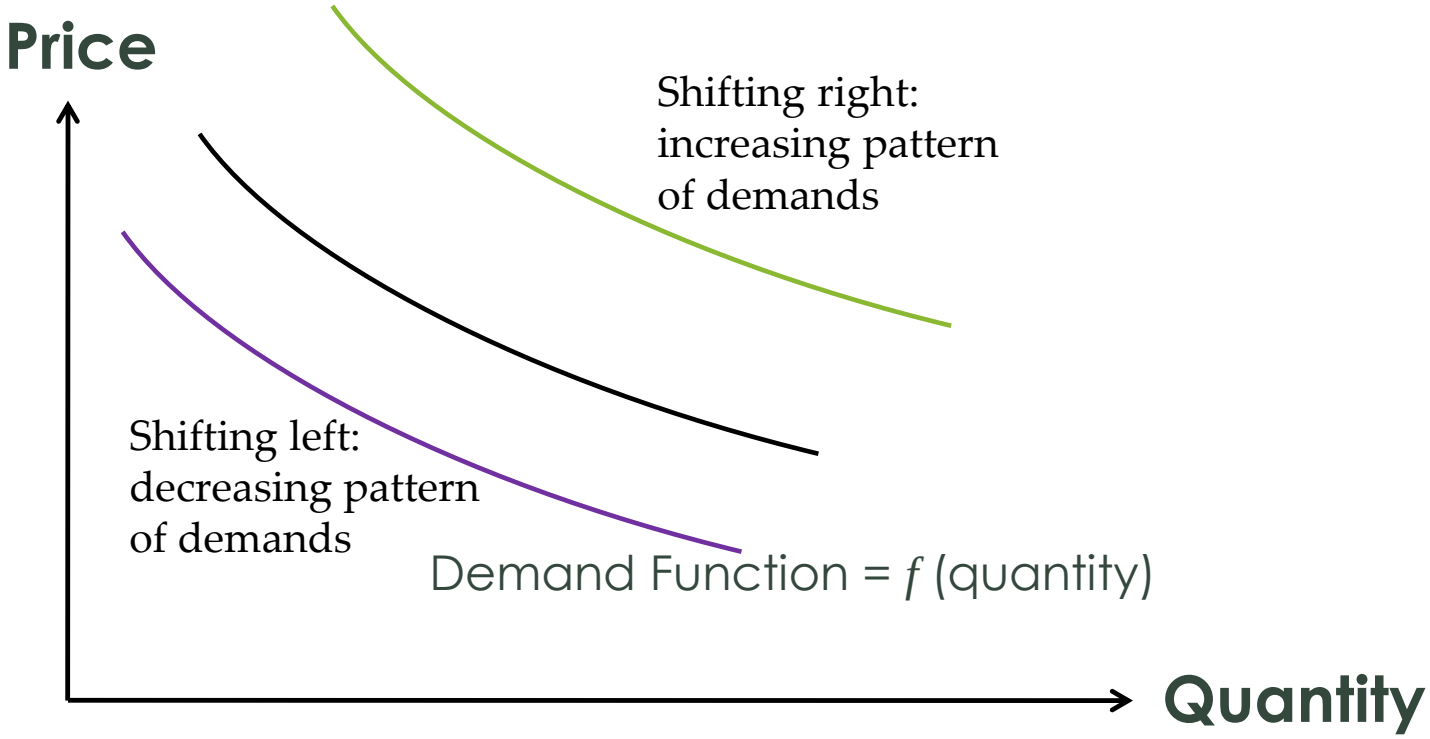
- Key concepts:
  - Demands for transportation: People generates such demand that are derived from the need to engage in activities at different locations
  - Transportation system enables the movements and thereby supply the means of transportation at the cost of travel time, fare/fuel cost, charges (toll, parking, etc.) and other externalities
  - Demand-Supply interactions:
    - Price/cost equilibrium
    - Externalities: congestions/delays, (in)convenience, (un)reliability, collisions/accidents, emission, pollutions, etc.
    - Consumer surplus/Social welfare

Demand Curves



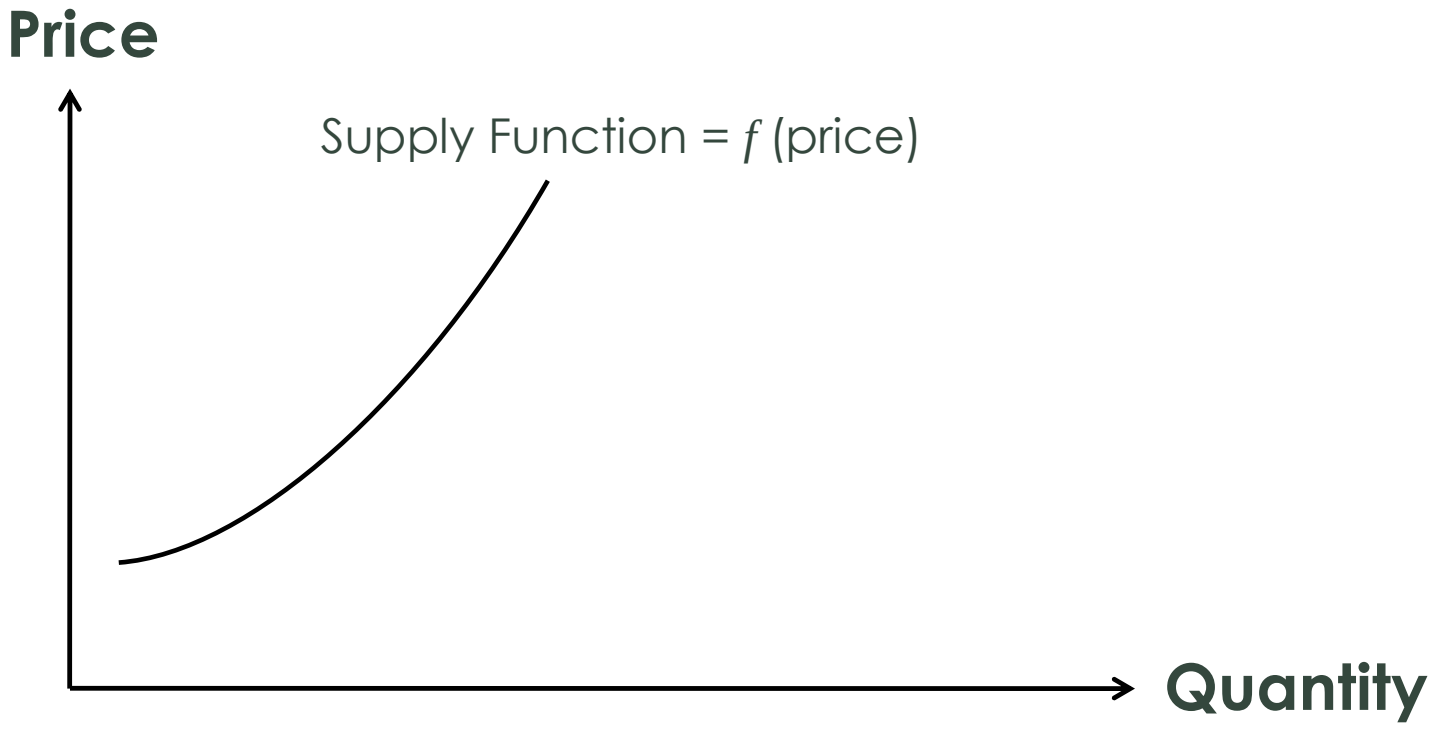
- Price/Cost
  - Travel time, cost, safety, comfort, convenience, etc.
  - Externalities
- Quantity of Demand
  - Total number of trips per day
  - Trips by different purposes
  - Trips at different destinations
- Demand Curve
  - Demand function
  - Cost/Price sensitivity of total demand

Changing Demand Patterns



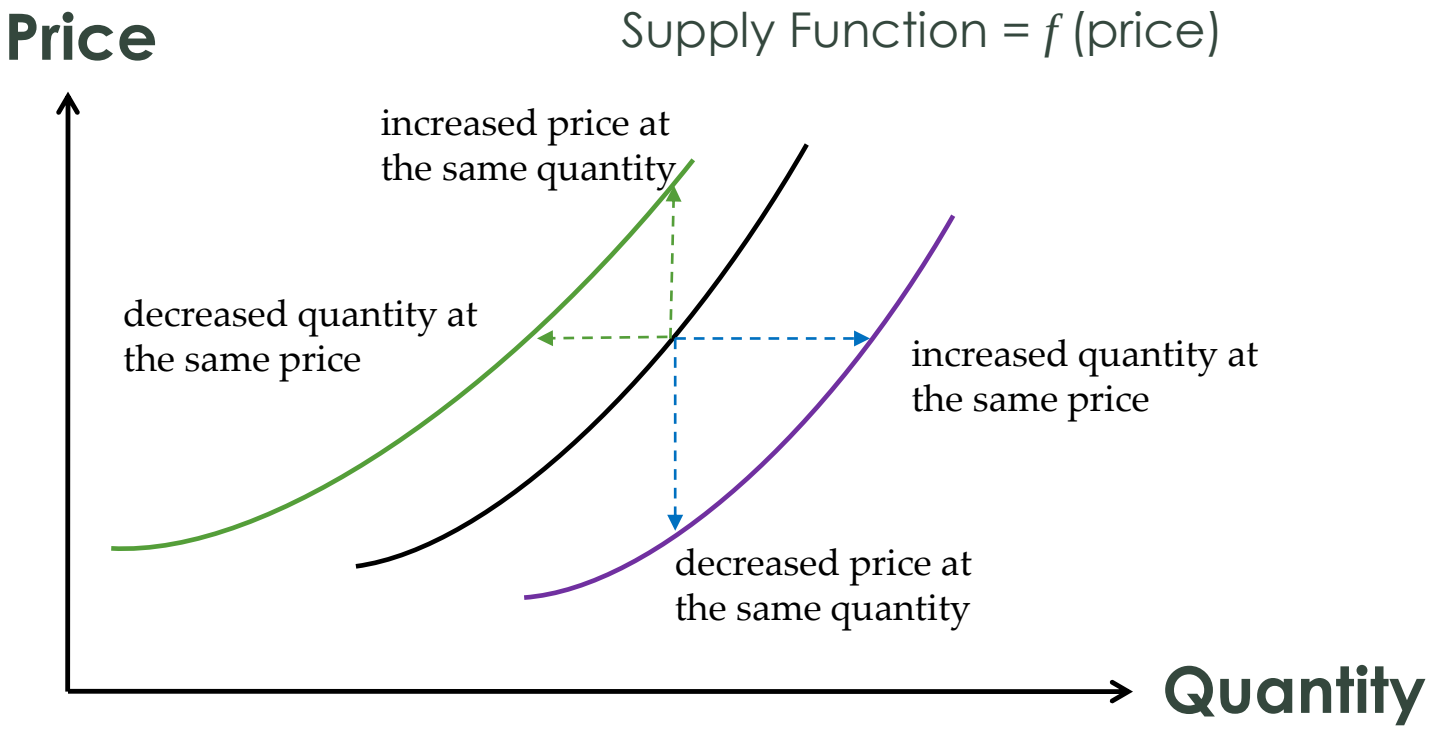
- Price sensitivity: Changes in quantity of travel due to changes in other variables than perceived price/cost of travel:
  - Income
  - Land use and urban form
  - Lifestyle, technology
  - Population and regional economy

Supply Curve



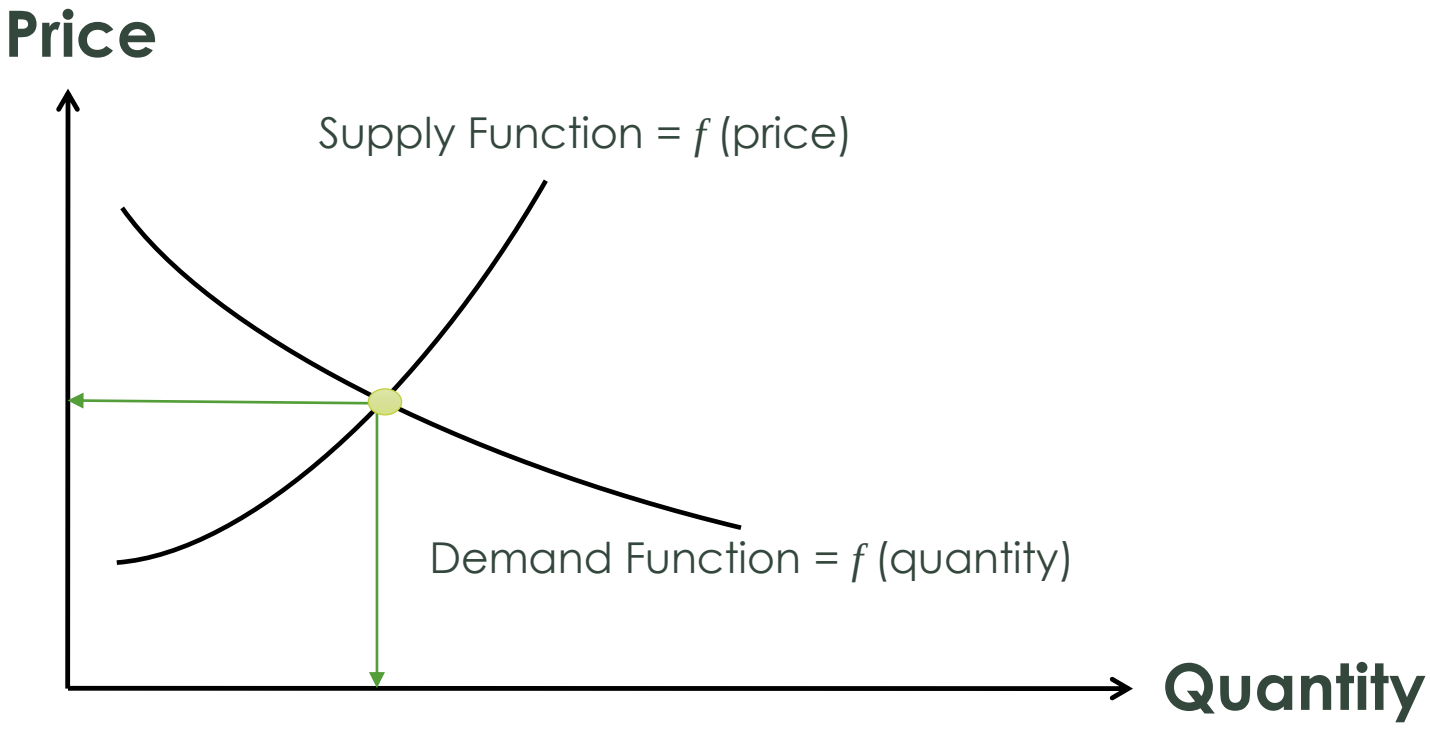
- Price
  - Average travel time, congestion delay, externalities, etc.
- Quantity of demands
  - Flow, density, speed
- Supply curve
  - Performance function

Changing Supply



- Increase in system Capacity: shifting to the right
  - Adding new element
  - Efficient operation and maintenance
- Decrease in system Capacity: shifting to the left
  - Reducing elements
  - Deterioration

Demand-Supply Equilibrium



- Equilibrium: the observed system
  - Observed link flows, speed, average travel time, average travel cost etc.
  - Observed total number of trips by zones, households, persons etc.
  - Observed trip distribution
  - Observed modal splits

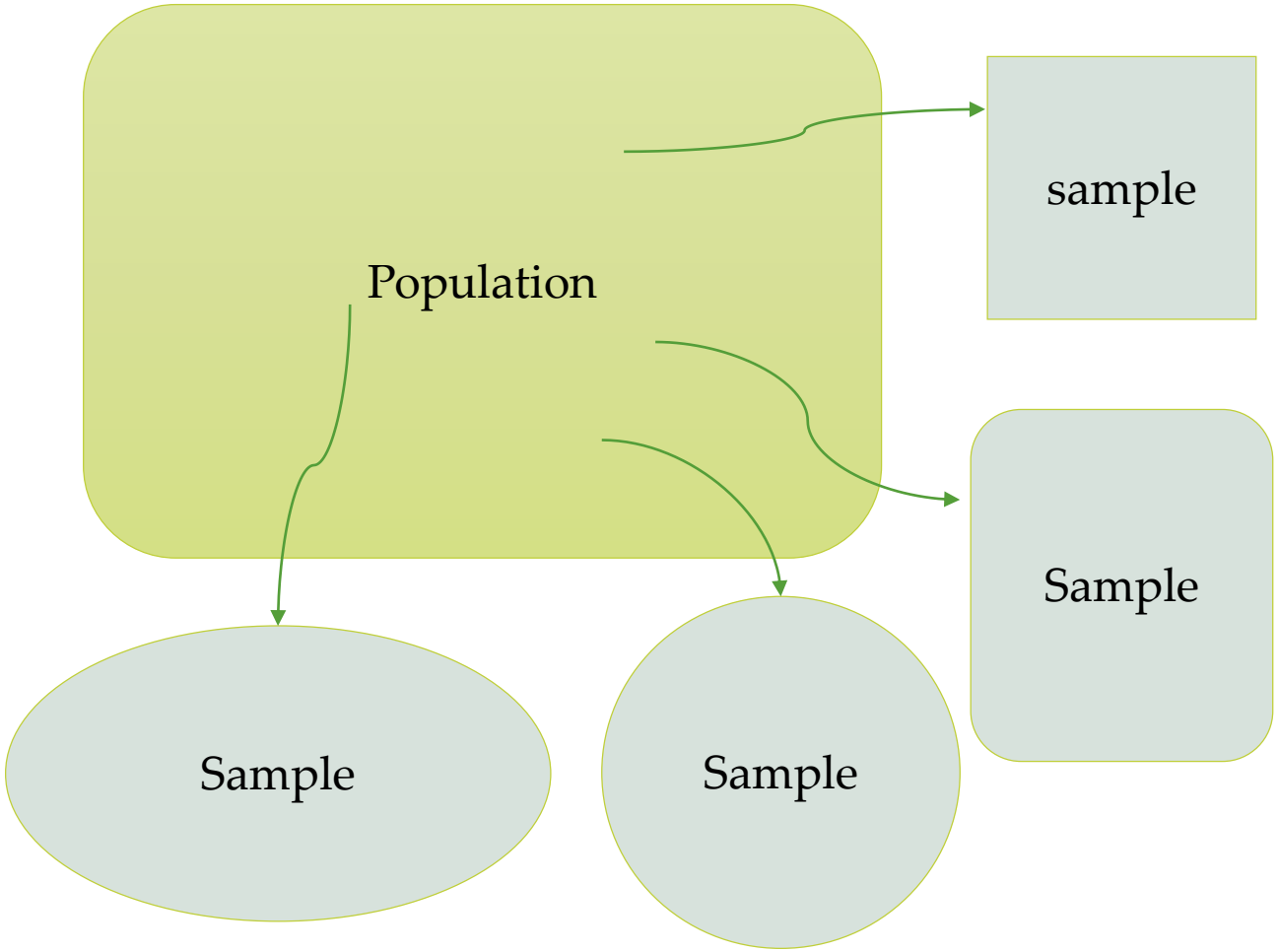
Evidence of Transportation Demand

- Evidence = Observed reality or Stated opinion/choice/preferences
- Transportation system observation (supply) has to have boundaries:
  - Observation time frame: daily observation, weekly observation, etc.
  - Observation extant: system wide observation, observation of key links, routes, corridors, etc.
- Transportation demand observation has to have boundary:
  - Observation of whole population of travellers (missing people not travelling) on a particular time-of-day, on a particular day
  - Observation of a sample of population (sample of households including those not generating any trips) on a particular day, week, or longer

Evidence of Transportation Demand

- Constraints in evidence generation (observation)
  - **Temporal constraint**:
    - Impossible to observe for indefinite time period
    - So, even a comprehensive observation within a finite timeframe is a sample of infinite temporal dynamics
  - **Spatial constraints**: Urban space and network need to be categorized for meaningful data presentation:
    - Zones, planning district etc. are sample of urban space
    - Highways, arterials, corridors etc. are sample of network
  - **Population/Socio-economic constraints**:
    - Even a census is a sample of continuous socio-economics and population dynamics
    - Majority of time, we rely on sample of households or individuals

Evidences are Samples of Reality

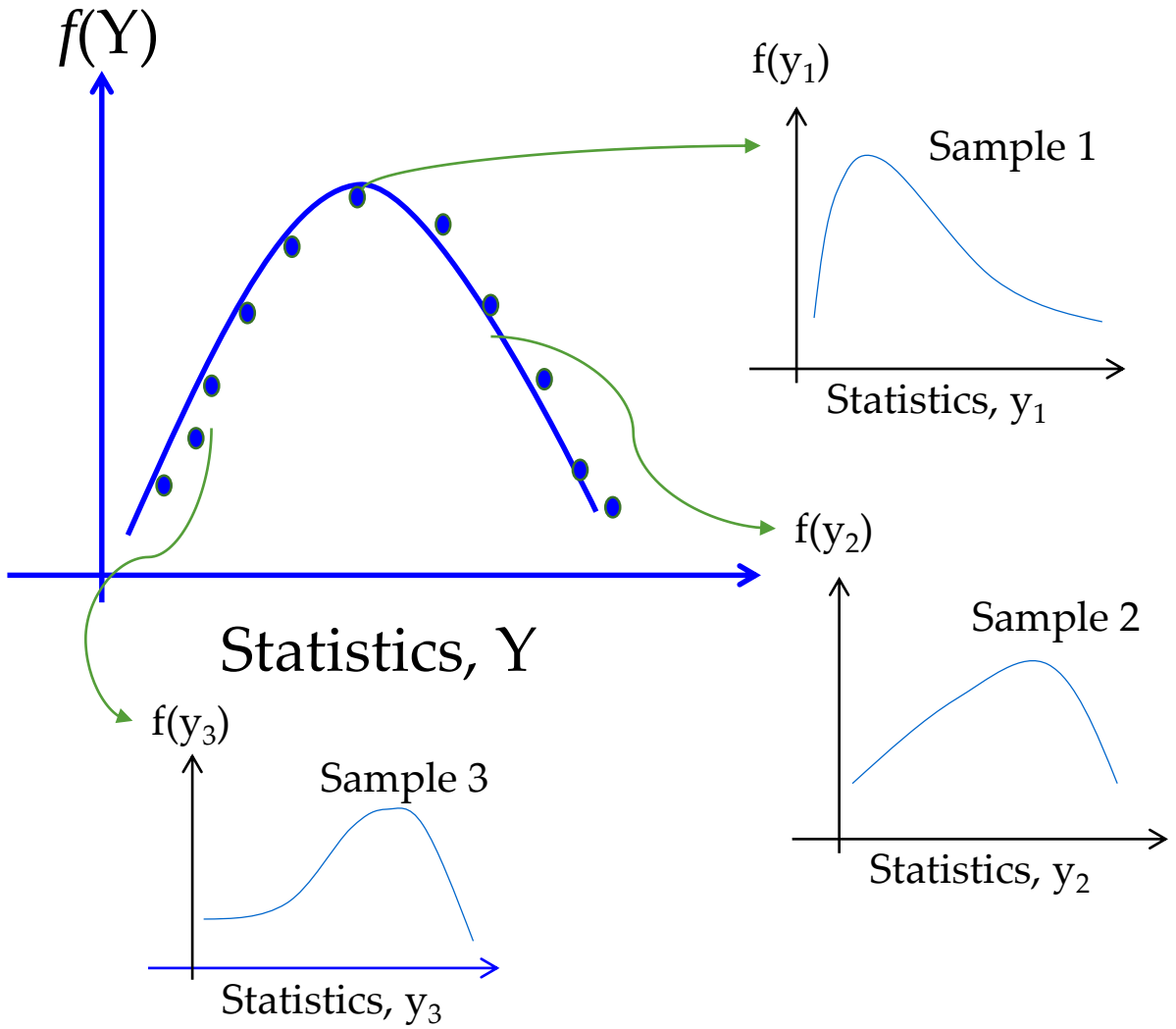


- **(Population of) Reality** : a general concept referring to the complete and defined set of possible evidences
  - Of all possible subjects (people, households, etc.) of the study area
  - Of all possible days, weeks, months, years, decades, etc.
  - Of all possible components of transportation network
- **Sample Evidence**: information of a subset of population or of a subset of possible contexts
  - Can vary by size, context and extant

Samples have Sampling Error

- Two types of **sampling error**:
  - ✓ Sampling bias: arises when specific aspects of actual information are either missing or overlooked
  - ✓ Sampling variance : it rests on the notion of inherent variations of transportation demand resulting from dynamics of demand-supply interactions
- Sampling bias can be overcome by properly specifying study context recognizing all elements of demand generation processes.
- Sample variance cannot me eliminated, but can be minimized by consider large amount of evidences (large sample size)

Samples have Sampling Error



- **Random sample**: Any representative set of evidences
  - Key statistics may vary with sample size, sample compositions, etc.
- **Sampling Distribution**: Distribution of sample statistics. Such as distribution of sample mean, sample variance etc.
  - Sample of evidences needs to be verified against the (population of) reality
  - To have confidence on sample evidences
- **Application of statistics (Statistical Inference) is unavoidable**

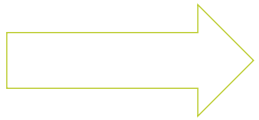
Evidences are Collection of Measurements

- Evidences are generated by measuring different aspects of transportation demands and supply:
  - Average speed, flow, density, congestion delay etc. are measurement of network performances
  - Trip generation, trip distribution, modal split, route choice etc. are measurements of transportation demands
- We measure transportation demands by specifying variables:
  - Quantitative variables
  - Qualitative variables
- Measuring of transportation demands through variable specification has inherent measurement errors

Construct - Measurement

- Construct: The underlying information that are to be measured is called construct
  - A single construct can be measured by specifying different variables (measurements): Each of such measurement (specified by variables) have inherent measurement error

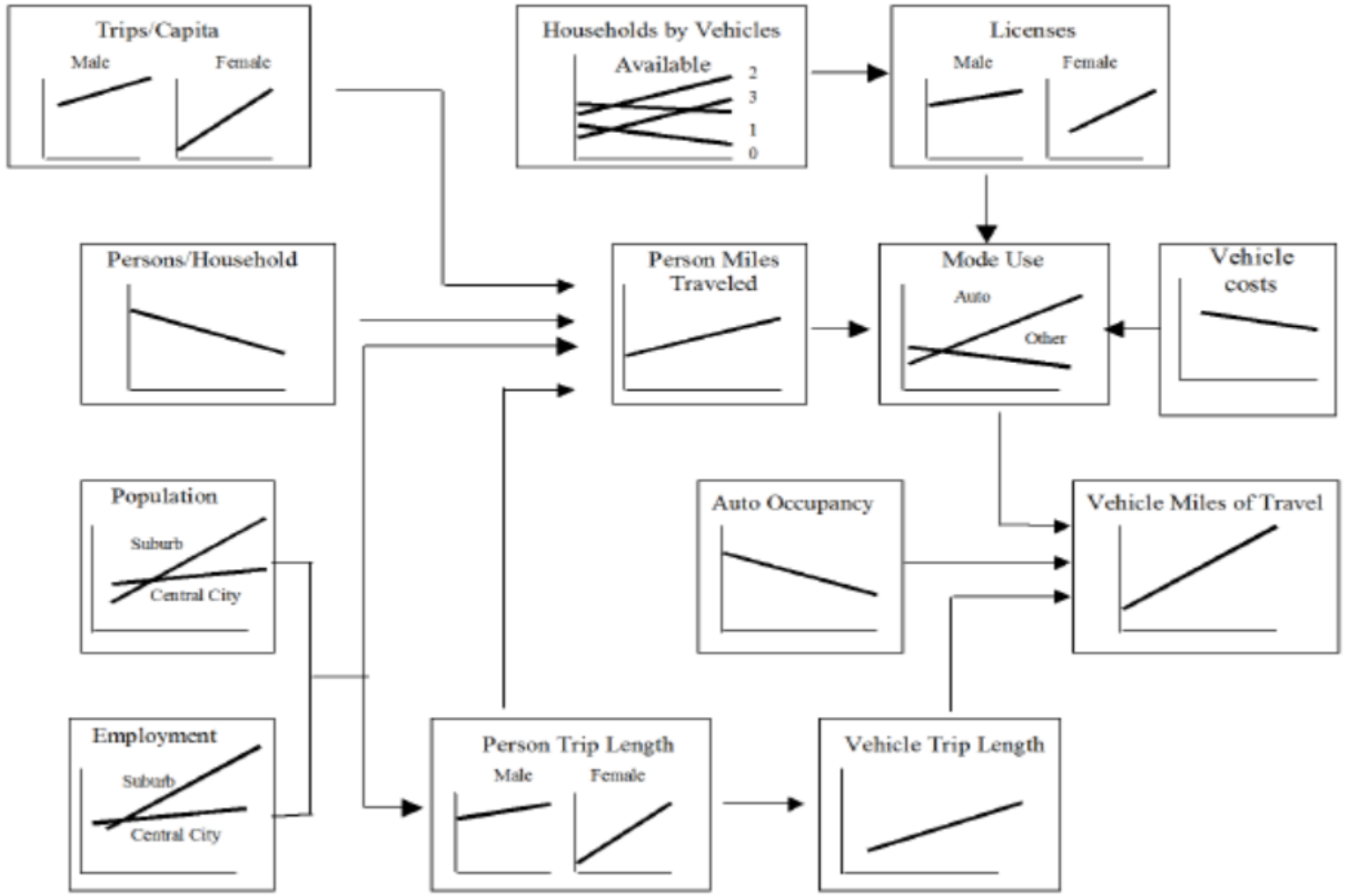
Construct: Trip Generation Tendency



Measurements of Trip Generation:
- Trips per day
- Trips by destination purpose
- Trips by mode of transportation
- Trips by time-of-day

  - Defining construct underlying evidences (variables) of travel demand requires consideration of travel behaviour
  - **Application of behavioural theory (e.g. Microeconomic) for transportation demand analysis is unavoidable**

Complexity in Travel Demand



- Multiple variables influence a single aspect of travel demand
- All aspects of travel demand are also correlated
- Correlations among the variables complicates the errors in measurement
- Correlation among individual people and household further complicates sampling and measurement errors

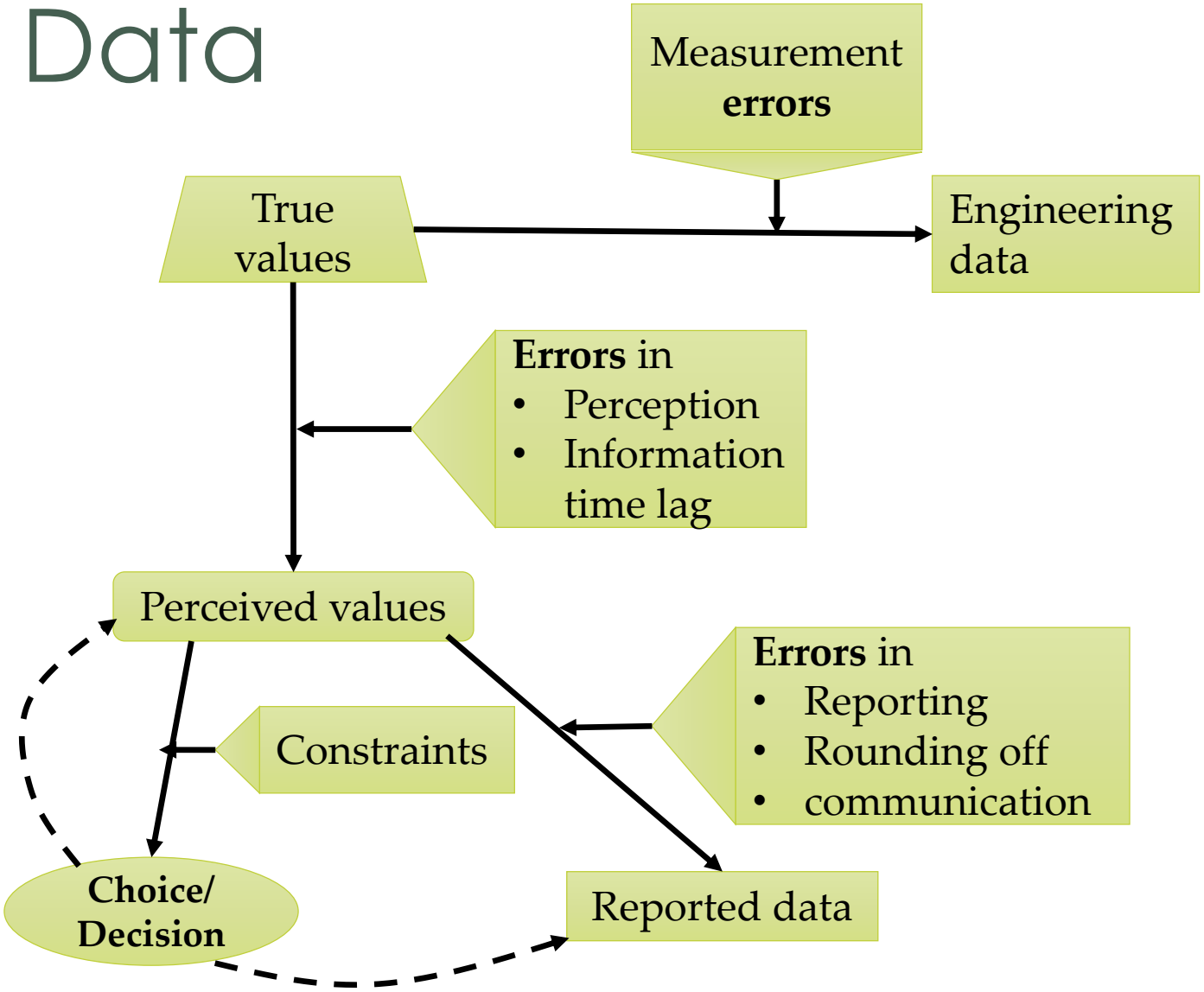
Sampling Error, measurement Error and Evidence-based Analysis

- A collection of information (observation of variable values) does not necessarily give the appropriate evidences of transportation demands
- Appropriate evidence-based transpiration demand analysis requires
  - Observed information set (variables)
  - Definition of sample characteristics of the observed set of information
  - Identifying appropriate construct underlying the measurements presented in the observation set
  - Estimating confidence on the evidence presented in the observations set through measurements of constructs of interest
- Use of behavioural theory to specify travel demand constructs representing in a set of measured observations and estimating confidence on those measurements is **econometrics**

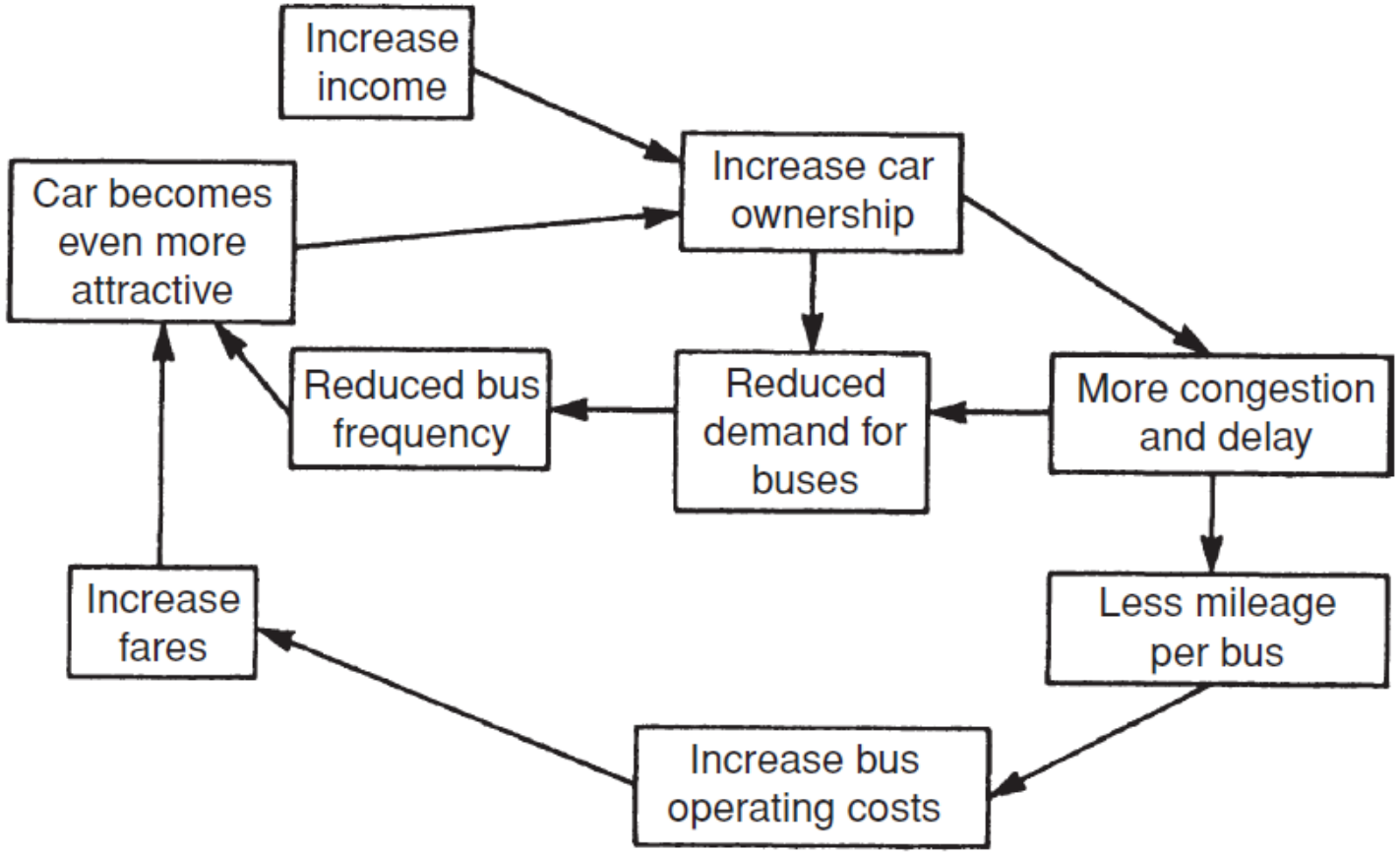
Econometrics for Evidence-Based Analysis

- **Data (variable)**: Observed set of information necessary to understand aspects of travel demand
  - Trip generation, mode choice, route choice, car ownership, home location choice etc.
- **Defining Construct**: Behavioural Theory of Travel Demand
  - Errors are nothing by unexplained variations of actual behaviour (construct) underlying measurements
  - Errors are random across the population and has specific distribution types
- **Model specification**: Specifying functional forms:
  - Systematic component of the construct that is explained by measured variables
  - Random errors components of construction with specific (appropriate) distribution
- **Model estimation**: Estimate weights (coefficients) of different variables as measured in the dataset and confidence on these estimates

Errors in Data



Model for Investigation



Vicious Cycle of Car and Public Transit

- Requires definition of key element/variable/phenomena of interest: Objective variable/Dependent variable (y)
- Requires identification key Explanatory/independent variables (x)
- Requires specification of logical/mathematical relationship between dependent and independent variables

$$y = f(x, \theta)$$

- Recognize the error in model specification and variable measurement

$$y = f(x, \theta) + \varepsilon$$

The Concept of Econometric Models

- Expressing measurement of a construction (Dependent Variable) as a function of other measurements (variables):
  - Dependent variable, y: The objective variable of interest
  - Independent variables, x: The explanatory variables
  - Weights/coefficients, θ of x in explaining the systematic elements of measuring , y
  - Functional form, f(x, θ) of the systematic explanation of y
  - Random error, ε in measuring y

$$y = f(x, \theta) + \varepsilon$$

- Model estimation: estimating the values as well as confidence limit of the parameters, θ, by using a set of observed data

- Specification of variables y as well as the distribution of ε define the estimation technique

Specification of Variables

- Possible types of variables from measurement points of view
  - Qualitative variable: Categorical/Discrete variables
    - Nominal scale: sex (male, female), urban location (CBD, suburb), employment status (employed, not employed), etc.
    - Ordinal scale: raking/ordered scale of measurement, e.g. income (low, medium, high), satisfaction (low, medium, high), reliability (low, medium, high), etc.
  - Quantitative variables: Continuous variables
    - Interval/cardinal scale: Likert scale satisfaction, numerical ranking scale etc.
    - Ratio scale: continues numbers e.g. age, travel distance, cost, etc.

Specification of Variables

- Possible types of (dependent) variables from modelling points of view

  - Continuous variable → Continuous Econometric Model
    - Pure continuous numbers including 0 as a possible value
    - Log-transformed continuous number for positive values only
    - Other non-linear transformation: box-cox transformation

  - Discrete variable → Discrete Econometric Model
    - Binary variable (1 or 0)
    - Nominal variables with more than 2 possible values or ordinal variables can be further specified as a set of binary variables for each category

Model Parameter Estimation

- Least-Square Estimation: Method of Moments
  - Dependent variable is a continuous variable

$$y = f(x,\theta) + \varepsilon \qquad y_{prediction} = f(x,\theta) + \varepsilon$$

  - For a set of observed y and corresponding x values, there will be error/residual due to the presence of ε

$$E = y_{observed} - y_{prediction} = y_{observed} - f(x,\theta)$$

  - Minimize the (E) for the optimum values of θ

- Least square estimation technique is unbiased, consistent and efficient.
- Microsoft excel has functions to conduct least-square estimation.
- Other software: R, Stata, LIMDEP, SAS, etc.

Model Estimation

- Maximum likelihood estimation
  - Specify the distribution of the random error component (ε)

$$y_{prediction} = f(x, \theta) + \varepsilon$$

  - Such specification makes the model ($y_{prediction}$) a random variable of the same distribution of (ε) and so, one can define the likelihood (probability) of an observed value ($y_{observed}$) of a record in the dataset i, L_i

$$likelihood\ of\ \ an\ observation, L_i = \Pr(y_{observed})$$

  - Considering that the observed dataset has N number of observation and each observations are independent, the sample likelihood becomes the multiplication of individual likelihoods

$$Sample\ likelihood, L = L_1 \times L_2 \times L_3 \times \ldots\ldots\ldots\ldots\ldots \times L_N$$

  - Maximize L (minimize log of L) to find out the optimum values of θ
- A non-linear optimization and standard numerical methods are available
- Max Likelihood estimation technique is unbiased, consistent and efficient
- Software: R, Stata, LIMDEP, SAS, etc.

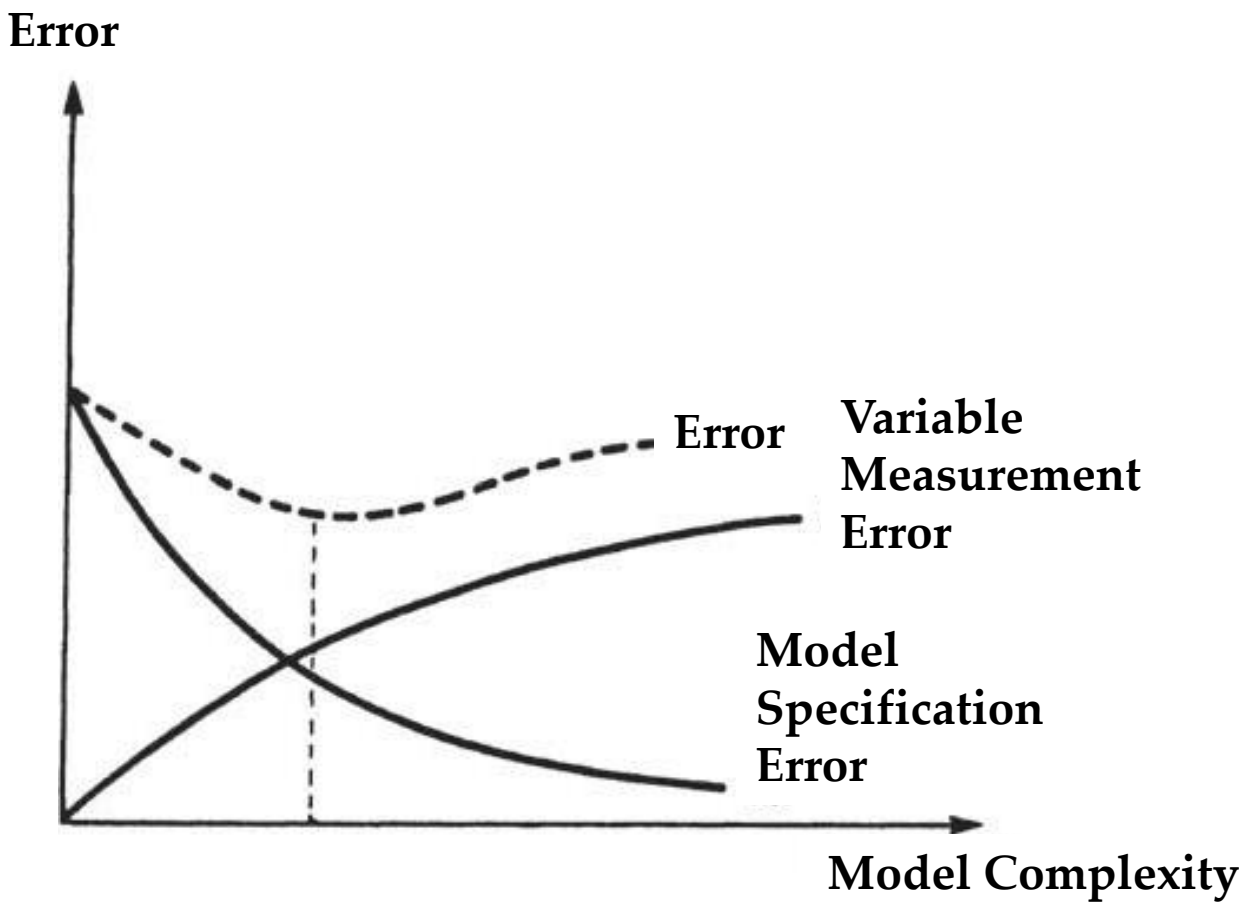
Notes on Model Estimation

- Least-Square estimation is feasible for mostly if the dependent variable is continuous in nature
- Maximum likelihood estimation can be applied for both discrete and continuous variable models
- These estimation methods have necessary statistical properties:
  - Requires a sample of observation (dataset) with dependent as well as independent variables
  - Estimations process gives the mean value and standard errors of the parameters estimates (θ): So, statistical significance test can be done
  - Estimation process allows testing the goodness-of-fit of the observed values (R-squared value, Rho-squared values)

Notes on Model Estimation

- Statistical significance of estimated parameters gives the confidence on the estimated effects of an independent variable (x) on the dependent variable (y)
  - Ratio of estimated mean and the standard error (mean of θ/ st. err of θ) gives t-statistics
  - For 95% confidence limit a t value of 1.64 confirms that the corresponding estimated parameter (θ) value is statistically significant if the sign of θ is known
  - For 95% confidence limit a t value of 1.96 confirms that the corresponding estimated parameter (θ) value is statistically significant if the sign of θ is unknown
- A 95% confidence on estimated parameter value means, for the 95 out of 100 random sample of observation dataset, the estimated value will be very close to actual estimated value.

Errors in Measurement and Implications



- Errors in modelling (prediction)
- Modelling complexity:
  - Increasing the number of influential variables
  - Comprehensive functional forms
- Variable measurement error
- If data are of poor quality, it may be safer to use simpler model specification

Key Concepts of Demand Investigation

- Marginal effects
- Demand elasticity
- Substitution patterns
- Complementary versus supplementary relationships
- Income/budget effect
- Lagged effects
- Demand/Preference Heterogeneity
- Heteroskedasticity
- Choice versus Demand

Thank You