PERFORMING PROBABILISTIC COST-EFFECTIVENESS ANALYSIS VIA DECISION TREE MODELING IN STATA: THE MANANTIAL COMMAND

Manuel García Goñi & Ricardo Mora

UCM & UC3M

Madrid, Oct 2017

Outline

1. Decision Trees & Stata
2. Case Study: First-episode psychosis
3. Manantial

Decision Trees & Stata

Decision Trees

- A Decision Tree is a tree-like structure in which
  - each **non-terminal node** represents a "test"
  - each **branch** represents the outcome of the test
  - each **terminal node** represents a class label (decision taken after computing all attributes)
- Probabilities into each branch are represented by a Markov-chain process
- Useful as an analytical decision tool when standard statistical techniques are not implementable (for lacking of data)
- Also used to simulate unconditional probabilities of each category
- The tree structure and the conditional distributions of all parameters are known

Stata & Decision Trees

- Decision Trees can be seen as decision rules with the form:

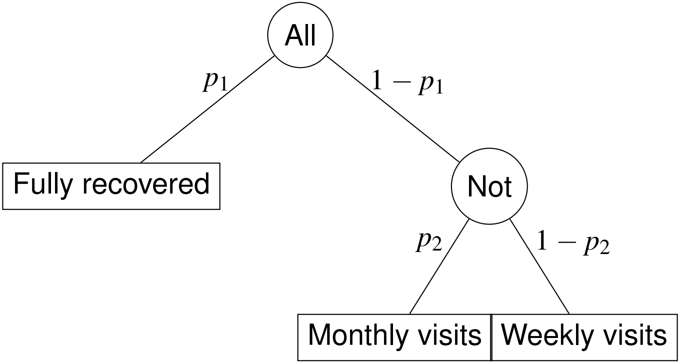
  if condition1 and condition2 and condition3 then outcome.

- Hence, for a given tree, it is actually not difficult to program its simulation in Stata using standard Stata commands
  - However, each change in the dimensions of the tree substantially changes the code
  - Bootstrapping and other sensitivity techniques are better performed using a given comand

- We present a new comand, <manantial>, that implements the Decision Tree algorithm

A simple example

$$\text{Health Status} = \begin{cases} \text{Fully recovered with prob. } p_1 \\ \text{If not } (1 - p_1)\text{, monthly visits with prob. } p_2 \\ \text{Weekly visits with prob. } 1 - p_2 \end{cases}$$



Case Study: First-episode psychosis

The potential application

- Goal: to analyze the cost-effectiveness of **early intervention for first-episode of psychosis** in young individuals in Madrid.
- Related case studies: EPPIC (Australia), OPUS (Denmark), or LEO (London)
  - They find lower severity, better health quality and lower hospital admissions and health expenditures for patients with early interventions.
- An early intervention program consists of a multidisciplinary team going to see and accompanying the patient in his/her itinerary through the health system.
- The expected impact is related to whether patients get back to the educational system or the labor market, reducing the productivity loss in their working life, compared to the standard treatment.

Two policies

- Standard treatment
  - After the first-episode of psychosis, the patiet goes to the formal healthcare provision (emergency, primary care, psychiatric, hospital).
  - Usually it ends up taking a strong pharmacological treatment after visiting the specialist.
- Early intervention (Intensive treatment)
  - Two persons in the multidisciplinary team visit the patient within 48 hours of the first-episode.
  - They talk to the patient and family to find the root of the psychotic problem.
  - They go with him to the doctor or specialist or hospital.
  - They follow the health status of the patient in the following weeks.
  - They promote social activities with family and friends and often reduce the use of pharmacological treatments.

The literature: McCrone et al (2009)

- Decision tree model through a Markov process to estimate probabilities.
- Ranges for the costs for the model of admissions, readmissions and use of community services were obtained from the literature

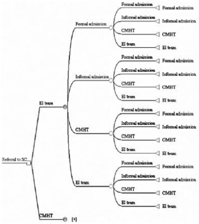


FIGURE 1. Early intervention sub-tree. CMHT, community mental health team (SC, standard care); EI, early intervention.

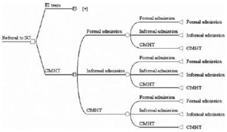
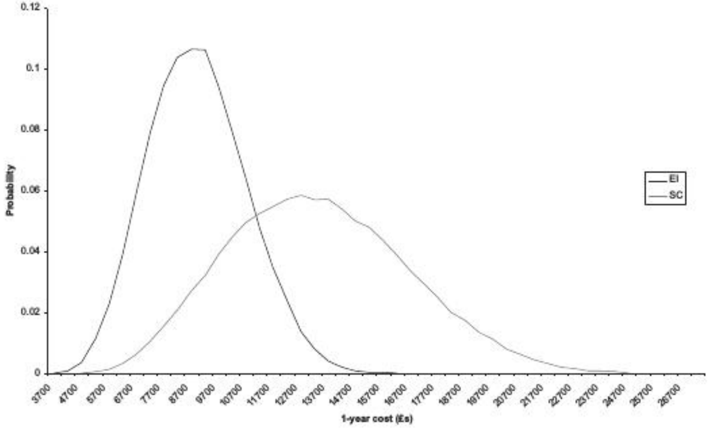


FIGURE 2. Standard care sub-tree. CMHT, community mental health team (SC, standard care); EI, early intervention.

FIGURE 5. Probabilistic sensitivity analyses of 1-year costs (2006/2007 £s). EI, early intervention; SC, standard care.



The Literature: La Park et al. (2016)

- Markov process to measure impact on employment and education
- It compares the costs and outcomes of EI services with standard care (SC) using TreeAge Data Pro 4.0 software

FIGURE 1. Decision analytical model for employment and education.

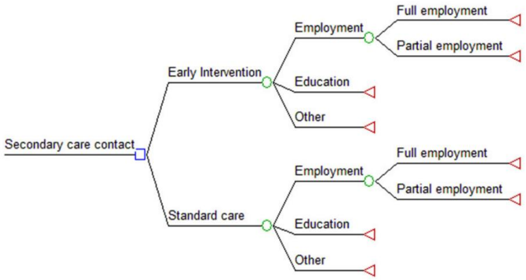


TABLE 3. Net savings per person for early intervention services over time (2009 £)

	Year 1	Years 1–3	Years 4–10
Health care services	5360		
Employment		2087	
Homicide			480
Suicide			5742
Total	5360	2087	6222

Our goals

- In the future: Measure the impact of Early Intervention (EI) Program in the health, employment and education status of young individuals with first-episode psychosis in Madrid
  - We will model the costs of both EI and Standard Care
  - We will model the outcomes in terms of employment and education

- Today: Develop a Decision Tree tool for Stata in which the Cost-Effectiveness Analysis is performed in an efficient way using mata

Manantial

What the ado file does

- Basic algorithm:
  - Simulate random shocks that represent individual idiosyncracies (the shocks that determine whether to go left or right in the tree)
  - Simulate random shocks that represent researcher's limited knowledge of decision parameters (the thresholds in each node to which the individual shocks must be compared)
  - Classify individual using tree structure
- The basic algorithm is computed for N individual shocks. We obtain:
  - A sample distribution of individuals along all categories
  - An average profit of the tree structure
- We replicate the distribution and the expected profit R times
- The command
  - displays summary statistics of the distribution for each estimated unconditional probability
  - displays summary statistics of the distribution of the expected profits
  - replaces the current data set with the replications data (optional)

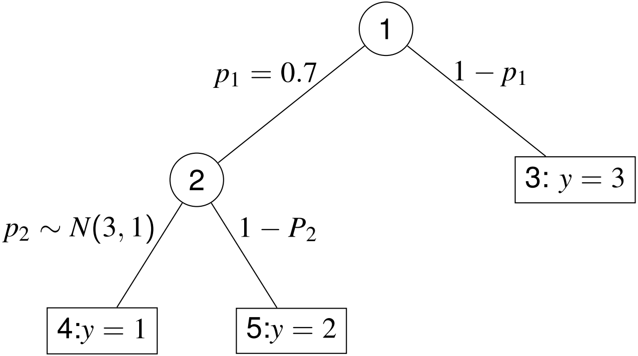
Inputs

- Required:
  - A matrix describing the tree structure and the statistical properties of the parameters of the tree and the individual shocks
  - A matrix describing the statistical properties of the profits associated with each category
  - A matrix describing the covariances between the individual shocks
- Optional:
  - a number (seed) to control pseudo-random simulation
  - the number of observations within each replication to estimate the class probabilities and the expected profits
  - the number of replications to estimate the estatistical properties of the class probabilities and the expected profits
  - a name to describe the distribution of the individual shocks
  - an indication whether the current data set should be replaced

Syntax

- **manantial** *[newvarlist]*, **t**ree(string) **p**rofits(string) [**seed**(real) **n**oobs(real) **r**eplications(real) **c**ovariance(string) **d**istribution(string) **replace**]
    - newvarlist: names of new variables corresponding to each category if data set is to be replaced
    - tree*(matname)*: name of matrix where tree information is stored
    - profits(*matname*): string is name of matrix where statistical properties of the net profits associated with each category are stored
    - seed(*#*): seed for the random number generator
    - noobs(*#*): sample size for the shocks in each replication
    - replications(*#*): number of bootstrap replications
    - covariance(*matname*): matrix name for variance -covariance matrix of shocks
    - distribution(*name*): name of distribution of shocks: uniform (default) or normal
    - replace: current dataset to be replaced with Montecarlo results

The tree matrix



1	2	0	0	0	0.7	0
2	4	1	0	2	3	1
3	0	1	3	9	0	0
4	0	2	1	9	0	0
5	0	2	2	9	0	0

The profits matrix

- In the previous example, ther are three terminal nodes, and three categories: 1, 2, and 3
- Suppose:
  - $\pi_1 = 0.5$
  - $\pi_2 \sim U[0.1, 0.7]$
  - $\pi_3 \sim N(0.2, 0.04)$
- Then, the profits matrix would be

$$\begin{array}{cccc} 1 & 0 & 0.5 & 0 \\ 2 & 1 & 0.1 & 0.7 \\ 3 & 2 & 0.2 & 0.04 \end{array}$$

`manantial FT PT Other, t(A) c(C) p(P) n(10000) r(500) replace`

- runs 500 bootstrap replications for estimates using samples of 10,000 observations
- detailed descriptive statistics for the distribution of the probability estimates of the categories and of the expeceted returns are shown
- the montecarlo dataset replaces the current data set
  - contains 500 observations
  - variables `FT`, `PT`, and `Other` contains the 500 unconditional probability estimates of the three categories (full time, part time, other)
  - variable `profits`

Extensions

- inputing the information into matrices is useful when programming complex simulations, but an alternative graphical & user-friendly method would be desirable
- integration with tree estimation using the CART algorithm
- ...

Thank you