Modeling approaches for the allocation of costs

Regional Training Course on Agricultural Cost of Production Statistics
21– 25 November 2016, Daejeon , Republic of Korea

1 – Introduction and rationale

• **Models** (statistical , econometric, etc.) generally **make a more efficient use of auxiliary information** than rule-based procedures (which are also models, but more rudimentary)

• **They can improve the precision and accuracy** of the estimates if:

- o They are used appropriately
- o The data required for the modeling is available
- o The assumptions are plausible (compare with assumptions made by other allocation methods)

• **Models are** generally **more flexible than rule-based procedures**

⇒ The results are less dependent on the analysts' subjectivity

2 – Types of models

- **Statistical imputation techniques**:
  - Nearest neighbor imputation, interpolation, etc.
  - These methods can be used if a sufficient pool of questionnaires with detailed data on costs by commodity exists.
- **Econometric models**:
  - In general, their objective is to estimate commodity-specific technical coefficients
  - They are based on a certain number of assumptions
  - In the most sophisticated models, some of the most restrictive assumptions can be relaxed

3 - Statistical imputation techniques (1/3)

- **Statistical methods that use data from a sub-set of the sample or from other surveys/censuses** for which detailed activity data is available to estimate commodity-specific CoP
- **Nearest neighbor imputation** is the most common method:
  - Hot-deck imputation: missing records (activity-specific CoP) are estimated from "similar" records from the same survey (hot-deck imputation)
  - Cold-deck or donor-based imputation : when data from other surveys are used
- **The difficulty resides in identifying the matching records** (the nearest neighbors)

3 – Statistical imputation techniques (2/3)

CoP Survey

CoP data	Com 1	Com 2	…	Com p	Farm
Farm 1	X	X	X	X	X
Farm 2	X	X	X	X	X
…	X	X	X	X	X
Farm k	X	X	X	X	X
Farm k+1	na	na	na	na	X
…	na	na	na	na	X
Farm n	na	na	na	na	X

Missing records estimated from records from the same survey (hot-deck)

Other survey

CoP data	Com 1	Com 2	…	Com p	Farm
Farm 1	X	X	X	X	X
Farm 2	X	X	X	X	X
…	X	X	X	X	X
Farm m	X	X	X	X	X

Missing records estimated from records from another survey (cold-deck)

3 - Statistical imputation techniques (3/3)

- Step 1: **Obtain commodity specific data on CoP**

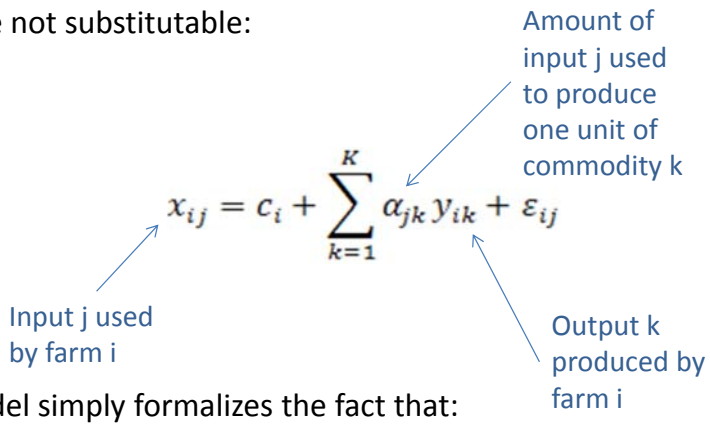
- Step 2: **Identify the variables on which the matching will be realized** (size, region, etc.). They have to be correlated with CoP but not between them

- Step 3: **Identify the distance used to identify nearest neighbors**

- Step4: **Define the procedure (function) used for the imputation**:
  - o Simple average, median
  - oMin, max, etc.

4 – The standard econometric model (1/2)

• The model estimates commodity-specific CoP under the assumption that input use is linearly dependent on the quantities produced and that inputs are not substitutable:



• The model simply formalizes the fact that:

o The **total amount of input consumed by a farm should be equal to the sum of the input uses across all the activities** of the farm, and that:

o **This relationship is true but subject to measurement errors and gaps** in the data

4 – The standard econometric model (2/2)

• **Data requirements:**

o Survey data on farm-level input costs and output by commodity

o Sufficient number of farms and combinations of inputs and outputs is required in order to estimate technical coefficients

• **Estimation techniques** (depending on how the data is structured):

oMCO or Generalized MCO

o Adapted methods for panel data (Between, Within, etc.)

• **Limitations:**

o This technique may lead to obvious errors: negative technical coefficients, estimates outside reasonable bounds, etc.

o The use of more sophisticated models, such as entropy-based approaches, can eliminate some of these errors

5 – Entropy-based regressions (1/2)

- **Rationale**: making use of prior information on:
  - Technical coefficients: minimum and maximum bounds, etc.
  - The existence of constraints: accounting equations, etc.
  - => To improve the quality of the estimations
- **Main assumptions:**
  - The unknown technical coefficients are a random variable
  - Auxiliary information can be used to bound the technical coefficients and provide a set of plausible values
  - Additional prior information is available (accounting equations, non-negativity constraints, etc.)
- **Estimation procedure** (idea)**:** the coefficients minimize the distance between the "true" probability distribution of the technical coefficients and the prior probability distribution

5 – Entropy-based regressions (2/2)

- **Data requirements** (minimal):
  - Survey data on farm-level output and input
  - Information on plausible values for the technical coefficients
- **Advantages**: cost-effective and statistically sound way to estimate commodity CoP
- **Limitations:**
  - Complexity: the implementation of this method requires advanced statistical knowledge and experience
  - *R* is one of the very few software proposing ready-to-use packages for entropy-based estimations

6 – Examples of model use

• **Several references** on model-based methods to estimate commodity CoP :

- Fragoso (2011): estimation of commodity-specific technical coefficients from the 2004 FADN database for a Portuguese region: prior information on the cost structure considerably improved the estimates
- Peeters (2002): similar model to undertake cost-allocation for a small number of dairy-beef farms in Brittany (France).

• **The results clearly show the superiority of these methods** with respect to more rudimentary approaches (rule-based, etc.).

• **However, these modeling tools are very little** used to produce national-level official statistics. This may be due to:

- A lack of data and prior information on the parameters and variables
- A lack of capacity to implement these methods and/or
- A defiance with respect to modeling tools in general

7 – References

• **The FACEPA project**, European Union (7th Framework Program)
The FACEPA project develops tools and methods to analyze production costs in European agriculture using FADN (Farm Accountancy Data Network) data

• Fragoso and Carvalho (2011), **Estimation of Cost Allocation Coefficients at the Farm Level Using and Entropy Approach**

• Desbois (2006)**, Méthodologie d'estimation des coûts de production agricole : comparaison de deux méthodes sur la base du RICA**, Revue MODULAD, num. 35

• Peeters and Surry (2002), **Generalized Cross Entropy Estimation of a Varying-Coefficients Model of Cost Allocation in a Multi-Product Farming,** Symposium on Maximum entropy approaches to modeling agricultural diversity in European agriculture

• **CEoptim**, Cross-entropy R package for optimization