Food and Agriculture Organization
of the United Nations

MEASURING FOOD LOSSES

Session 6:
Loss assessment through modelling

2018

Objectives of the presentation

- Provide guidance on conducting an assessment of losses through modelling
- Present the different models and issues

Outline

Introduction

1) Concepts and types of data

2) Regression analysis: general linear regression model

3) Example of Pakistan

Introduction

- Modelling approaches can be used to quantify PHL of certain crops or commodities at various stages of the supply chain

- The objective is to try to estimate what the researcher thinks are the determinants of post-harvest losses at various levels

- A model is a set of assumptions that describes the behaviour of a phenomenon

- It consists of :
  - A set of equations describing the behaviour
  - A statement about the errors in the observed values of the variables
  - A specification of the probability distribution of the disturbances

1 Concepts and types of data

1. Concepts and types of data

- The researchers or investigators creating the model should aim to make it representative
- It needs to be seen as continuous work of the other methods: variables can come from a survey
- It should contain the salient features of the phenomena under study
- Variables (parameters) thought to be relevant to explain the phenomenon should be explicitly included in the model
- The choice of the variables is the most important part:
  - **Expert judgements**
  - **Factor analysis**
  - **Machine learning methods (random forest, deep learning etc.)**

1. Concepts and types of data

- **Times series data:** Information about the numerical values of variables collected over time from period to period

- **Cross-section data:** Information on the variables concerning individual agents in the supply chain at a given point of time

- **Panel data:** Data from a repeated survey of a single (cross-section) sample in different periods of time

- **Dummy variable data:** When the variables are qualitative in nature, then data are recorded in the form of an indicator function

2 Regression analysis: general linear regression model

2. Regression analysis: general linear regression model

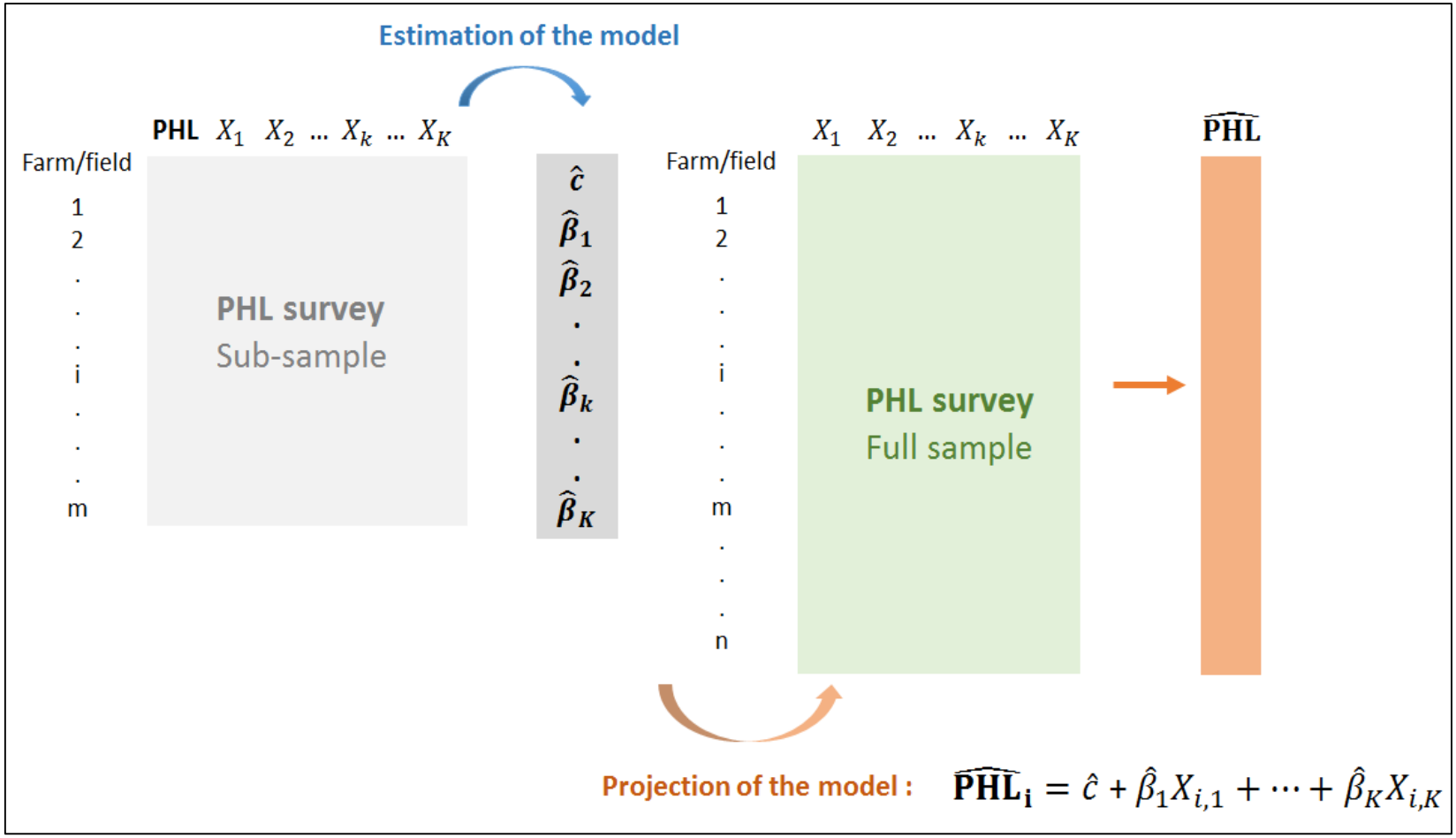
- Statistical relationship between two or more variables so that one variable (quantitative) can be predicted from the other(s)

- In the case of PHL, Neter & Wasserman (1985) propose:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,\, p-1} + \varepsilon_i$$

- As independent variables are used for the loss assessment, one can choose, for example, the **type of seed used, the area planted, the agricultural practices for harvesting (mechanical or traditional), etc**. In the case of storage, the dependent variable could be the type of pesticide used, the storage facilities etc.

2. Regression analysis: general linear regression model



3 Example of Pakistan

3. Example of Pakistan

Ahmed et al.

- **Goal:** Quantity PHL of Kinnow (citrus fruit) at various stages of the supply chain (farm, wholesale market and retail levels)

- Fruit but illustrate the modelling aspect

- **Data sources:** Survey from one district in Pakistan using all the sampling procedures

3. Example of Pakistan

Model at the farm level

$$\text{Ln } L1 = \beta 0 + \beta 1 \text{ Ln } X1 + \beta 2 \text{ Ln } X2 + \beta 3 \text{ Ln } X3 + \beta 4\ D1 + \beta 5\ D2 + \varepsilon$$

Where

- L1 = Post-harvest losses of Kinnow in kg
- X1 = Education in years
- X2 = Experience in years
- X3 = Orchard size in Acres
- D1 = Dummy variable for picking time (1 if picking time is morning, 0 if picking method is evening)
- D2 = Dummy variables for picking method (1 if picked with scissor, 0 if picked manually)
- ε = Disturbance term

3. Example of Pakistan

Model at the farm level

Variables	Coefficients	Std. Error	t-value	Sig.	Overall fitness
(Constant)	3.839	0.590	6.507	0.000	R2 = 0.406, Adjusted R2 = 0.315, F-value = 4.5 at 5% degree of freedom
Ln X1 (education in years)	-0.211	0.138	-1.526	0.137	
Ln X2 (experience in years)	-0.222	0.108	-2.057	0.048	
Ln X3 (orchard size in acres)	0.214	0.074	2.878	0.007	
D1 (dummy for picking time)	-0.276	0.143	-1.936	0.061	
D2 (dummy for picking method)	-0.477	0.218	-2.187	0.036	

3. Example of Pakistan

Model at the wholesale level

$$\text{Ln } L2 = \alpha 0 + \alpha 1 \text{ Ln } Y1 + \alpha 2 \text{ Ln } Y2 + \alpha 3\ D3 + \alpha 4\ D4 + \alpha 5\ D5 + \mu$$

Where

- L2 = Quantity of post-harvest losses in kg
- Y1 = Education in years
- Y2 = Experience in years
- D3 = Dummy variables for Infrastructure of transportation (1 if roads were metalled, 0 if roads were non-metalled)
- D4 = Dummy variable for loading method (1 if produce was loaded in boxes, 0 if produce was openly loaded)
- D5 = Dummy variable for storage place (1 if cold storage, 0 if normal storage)
- μ = Disturbance term

3. Example of Pakistan

Model at wholesale level

Variables	Coefficients	Std. Error	t-value	Sig.	Overall fitness
(Constant)	4.808	0.461	10.437	0.000	R2 = 0.68, Adjusted R2 = 0.63, F-value = 14.48 at 5% degree of freedom
Ln Y1 (education in years)	-0.154	0.115	-1.335	0.191	
Ln Y2 (experience in years)	-0.272	0.140	-1.944	0.060	
D3 (dummy for Infrastructure of transportation)	-0.593	0.390	-1.521	0.137	
D4 (dummy for loading method)	-0.555	0.273	-2.031	0.050	
D5 (dummy for storage place)	-0.562	0.293	-1.916	0.064	

3. Example of Pakistan

Model at retail level

$$\text{Ln } L3 = \gamma 0 + \gamma 1 \text{ Ln } Z1 + \gamma 2 \text{ Ln } Z2 + \gamma 3 \text{ } D6 + e$$

Where

- L3 = Quantity of post-harvest losses in Kg
- Z1= Experience in years
- Z2 = Unsold quantity on daily basis
- D6 = Dummy variable for type of retailor (1 if respondent was a shopkeeper, 0, if respondent was a hawker)
- e = Disturbance term

3. Example of Pakistan

Model at retail level

Variables	Coefficients	Std. Error	t-value	Sig.	Overall fitness
(Constant)	0.453	0.317	1.429	0.162	R2 = 0.62 Adjusted R2 = 0.59 F-value = 18.74
Ln Z1 (experience in years)	-0.080	0.108	-0.738	0.466	
Ln Z2 (unsold quantity in kgs)	0.259	0.135	1.921	0.063	
D6 (dummy variable for type of retailer)	-1.320	0.266	-4.959	0.000	

Conclusion

- This presentation described the main approaches and strategies to model losses

- The most important aspect in modelling is the choice of the variables used to explain the phenomenon

- A model can be designed for all the stages of the supply chain if data are available

References

- **Ahmad, T. Sud, U.C., Rai, A., Sahoo, P.M., Jha, S.N. & Vishwakarma,** *R.K. Sampling methodology for estimation of harvest and post-harvest losses of major crops and commodities*. Paper presented during the Global Strategy Outreach Workshop on Agricultural Statistics, FAO Headquarters, Rome. 24–25 October 2016

- **Neter, J., Wasserman, W. & Kutner, M.H.** 1985. *Applied Linear Statistical Models Regression, Analysis of Variance, and Experimental Design*. 2nd edn. Richard D. Irwin, Inc.: Homewood, IL, USA.

20

Thank You