



Editors

Roberto Benedetti

Marco Bee

Giuseppe Espa

Federica Piersimoni

Agricultural Survey Methods

 WILEY

AGRICULTURAL SURVEY METHODS

Edited by

Roberto Benedetti

'G. d'Annunzio' University, Chieti-Pescara, Italy

Marco Bee

University of Trento, Italy

Giuseppe Espa

University of Trento, Italy

Federica Pierseferoni

National Institute of Statistics (ISTAT), Rome, Italy



 **WILEY**

A John Wiley and Sons, Ltd., Publication

This edition first published 2010
© 2010 John Wiley & Sons Ltd

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com. The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloguing-in-Publication Data

Benedetti, Roberto, 1964-

Agricultural survey methods / Roberto Benedetti, Marco Bee, Giuseppe Espa, Federica Piersimoni.

p. cm.

Based on papers presented at the 1998, 2001, 2004 and 2007 International Conferences on Agricultural Statistics.

Includes bibliographical references and index.

ISBN 978-0-470-74371-3 (cloth)

I. Agriculture—Statistical methods. I. Bee, Marco. II. Piersimoni, Federica. III. Title.

S566.55.B46 2010

338.1021—dc22

2010000132

A catalogue record for this book is available from the British Library.

ISBN: 978-0-470-74371-3 (H/B)

Typeset in 10/12 Times-Roman by Laserwords Private Limited, Chennai, India

Printed and bound in the United Kingdom by CPI Antony Rowe, Chippenham, Wiltshire.

Agricultural Survey Methods

To Agnese and Giulio

Roberto Benedetti

To Chiara and Annalisa

Marco Bee

To Massimo, Guido and Maria Rita

Giuseppe Espa

To Renata and Oscar

Federica Piersimoni

List of Contributors

Luca Alinovi

Agricultural Development Economics
Division (ESA)
FAO of the United Nations
Rome, Italy.
Luca.Alinovi@fao.org

Dale Atkinson

Research Division
National Agricultural Statistics
Service (NASS)/USDA
Fairfax, VA, USA.
Dale.Atkinson@usda.gov

Bettina Baruth

IPSC-MARS, JRC
Ispra, Italy.
Bettina.baruth@jrc.ec.europa.eu

Marco Bee

Department of Economics
University of Trento
Italy.
marco.bee@unitn.it

Roberto Benedetti

Department of Business, Statistical,
Technological and Environmental
Sciences (DASTA)
University 'G. d'Annunzio' of
Chieti-Pescara
Pescara, Italy.
benedett@unich.it

Gianni Betti

Department of Quantitative Methods
University of Siena
Italy.
betti2@unisi.it

Andrea Carfagna

Department of Statistical Sciences
University of Bologna
Italy.
andreacarfagna@virgilio.it

Elisabetta Carfagna

Department of Statistical Sciences
University of Bologna
Italy.
elisabetta.carfagna@unibo.it

Raymond L. Chambers

School of Mathematics and Applied
Statistics
University of Wollongong
Australia.
ray@uow.edu.au

Denis Chartrand

Agriculture Division
Statistics Canada
Ottawa, Canada.
Denis.Chartrand@statcan.gc.ca

Arijit Chaudhuri

Applied Statistics Unit
Indian Statistical Institute
Kolkata, India.
arijitchaudhuri@rediffmail.com

Nhu Che

Australian Bureau of Agriculture
and Research Economics (ABARE)
Canberra, Australia.
NChe@abare.gov.au

Jim Cotter

National Agricultural Statistics
Service (NASS)/USDA, retired.

Carrie Davies

Area Frame Section of National
Agricultural Statistics Service
(NASS)/USDA
Fairfax, VA, USA.
Carrie_Davis@nass.usda.gov

Jacques Delincé

Institute for Prospective
Technological Studies
Seville, Spain.
jacques.delince@ec.europa.eu

Marcelle Dion

Agriculture, Technology and
Transportation Statistics Branch
Statistics Canada
Ottawa, Canada.
Marcelle.Dion@statcan.gc.ca

Giuseppe Espa

Department of Economics
University of Trento
Italy.
giuseppe.espa@economia.unitn.it

Pieter Everaers

Eurostat, Directorate D
External Cooperation, Communication
and Key Indicators
Luxembourg.
Pieter.EVERAERS@ec.europa.eu

Danila Filipponi

Istat
National Institute of Statistics
Rome, Italy.
danila.filipponi@istat.it

Javier Gallego

IPSC-MARS, JRC
Ispra, Italy.
javier.gallego@jrc.ec.europa.eu

Giulio Ghellini

Department of Quantitative Methods
University of Siena
Italy.
ghellini@unisi.it

Antonio Giusti

Department of Statistics
University of Florence
Firenze, Italy.
giusti@ds.unifi.it

Carol C. House

Agricultural Statistics Board
National Agricultural Statistics
Service
Washington, DC, USA.
Carol_House@nass.usda.gov

Philip N. Kokic

Commonwealth Scientific and Industrial
Research Organisation (CSIRO)
Canberra, Australia.
phil.kokic@csiro.au

Ulf Jorner

Statistics Sweden
Stockholm, Sweden.
Ulf.Jorner@scb.se

Claude Julien

Business Survey Methods Division
Statistics Canada
Ottawa, Canada.
Claude.Julien@statcan.gc.ca

Erdgin Mane

Agricultural Development Economics
Division (ESA)
FAO of the United Nations
Rome, Italy.
Erdgin.Mane@fao.org

Paul Murray

Agriculture Division
Statistics Canada
Ottawa, Canada.
Paul.Murray@statcan.gc.ca

Jack Nealon

National Agricultural Statistics
Service (NASS)/USDA
Fairfax, VA, USA.
Jack.Nealon@usda.gov

Jeroen Pannekoek

Department of Methodology
Statistics Netherlands
The Hague, Netherlands.
j.pannekoek@cbs.nl (and jpnk@cbs.nl)

Federica Piersimoni

Istat
National Institute of Statistics
Rome, Italy.
piersimo@istat.it

Paolo Postiglione

Department of Business, Statistical,
Technological and Environmental
Sciences (DASTA)
University 'G. d'Annunzio' of
Chieti-Pescara
Pescara, Italy.
postigli@unich.it

J.N.K. Rao

School of Mathematics and Statistics
Carleton University
Ottawa, Canada.
jr Rao@math.carleton.ca

Ray Roberts

National Agricultural Statistics
Service (NASS)/USDA
Fairfax, VA, USA.
Ray.Roberts@usda.gov

Donato Romano

Department of Agricultural and
Resource Economics
University of Florence
Firenze, Italy.
donato.romano@unifi.it

Vijay Verma

Department of Quantitative Methods
University of Siena
Italy.
verma@unisi.it

Frederic A. Vogel

The World Bank
Washington, DC, USA.
fvogel@worldbank.org

Ton de Waal

Department of Methodology
Statistics Netherlands
The Hague, Netherlands.
t.dewaal@cbs.nl

Anders Wallgren

Örebro University and
BA Statistiksysteem AB
Vintrosa, Sweden.
ba.statistik@telia.com

Britt Wallgren

Örebro University and
BA Statistiksysteem AB
Vintrosa, Sweden.
ba.statistik@telia.com

Introduction

The importance of comprehensive, reliable and timely information on agricultural resources is now more than ever recognized in various practical situations arising in economic, social and environmental studies. These renewable, dynamic natural resources are necessary for many countries where the growing population pressure implies the need for increased agricultural production. To improve the management of these resources it is necessary to know at least their quality, quantity and location.

In Western countries, agriculture is assuming a more and more marginal economic role in terms of its percentage contribution to GNP, but recent radical economic and social transformations have caused a renewed interest in this sector. Such interest is due not only to economic factors but also to issues related to the quality of life and to the protection of the public health. Today the food industry suffers from phenomena that originate in the primary sector, such as diseases on farms or the production of genetically modified organisms. The growing attention of consumers to the quality of food products has strongly reinforced the need to look at a single agro-food product as the result of a chain of processes linked together. In this approach, agriculture represents not only an economic sector but also the origin of the food chain, and because of this role it deserves special attention. These aspects, together with the related recovery and protection of the environment, have led to deep modifications in the data provided in this sector.

Agricultural surveys are thus conducted all over the world in order to gather a large amount of information on the classic crops, yields, livestock and other related agricultural resources. As a result, the statistics produced are so strongly conditioned by this largely diversified demand that many countries, in order to be able to comply with these requests, began to set up a complex system of surveys based on a harmonized and integrated set of information whose design, implementation and maintenance require a strong methodological effort.

Apart from the difficulties typical of business data, such as the quantitative nature of many variables and their high concentration, agricultural surveys are indeed characterized by some additional peculiarities that often make it impossible or inefficient to make use of classical solutions proposed in the literature. In particular we refer to the following:

- (a) The definition of the statistical units to be surveyed is neither obvious nor unique, because the list of possible options is quite large (family, agricultural holding, household, parcel of land, point, etc.) and its choice depends not only on the

phenomenon for which we are interested in collecting the data, but also on the availability of a frame of units of sufficient quality.

- (b) Typological classifications of the statistical units are very important tools to define the estimation domains and to design an efficient survey. However, harmonized hierarchical nomenclatures are usually not available for a certain definition of statistical unit, or they do exist but are so subjective that they cannot be considered as standard.
- (c) The concentration of many variables is often even higher than in other business surveys.
- (d) In many countries the use of the so-called 'census frames' is considered an ordinal procedure, with the obvious consequence that the list of sampling units is not updated and is very far from the target population. This has evident implications in terms of non-sampling errors due to under- or, less dangerous, over-coverage of the list.
- (e) When designing a sample, the theory suggests two classical ways of using a size measure existing in the frame: a scheme with inclusion probabilities proportional to the size of the unit, and a stratification obtained through the definition of a set of threshold levels on the size variable. Nonetheless, a typical frame of agricultural units has a large amount of auxiliaries and the size of each unit is usually multivariate. In this case the solutions proposed by the methodological literature are much more complex.
- (f) At least when using geographically referred units, there often exists a particular auxiliary variable requiring ad hoc procedures to be used in a sampling environment: the remotely sensed data. Remote sensing is nothing but a tool to get information about an object without being in physical contact with the object itself, and it is usually represented by digital images sensed from a satellite or an aircraft.

As far as we know, in the current literature there exists no comprehensive source of information regarding the use of modern survey methods adapted to these distinctive features of agricultural surveys.

However, the successful series of conferences on agricultural statistics, known as ICAS (International Conference on Agricultural Statistics), demonstrates that there is a broad and recognizable interest in methods and techniques for collecting and processing agricultural data. In our opinion, the remarkable number of high-quality methodological papers presented in these conferences may serve to fill this gap.

This book originates from a selection of the methodological papers presented at this set of conferences held in Washington, DC (1998), Rome (2001), Cancún (2004) and Beijing (2007). The declared aim was to develop an information network of individuals and institutions involved in the use and production of agricultural statistics.

These conferences were organized by the national statistical offices of the host-ing countries – the National Agricultural Statistics Service (NASS) and United States Department of Agriculture (USDA); Italian National Statistical Institute (Istat), Agri-food and Fishery Information Service, Mexico (SIAP) and National Bureau of Statistics of China (NBS), in collaboration with international institutes such as Eurostat, FAO, OECD, UN/ECE, and ISI.

This book is an attempt to bring together the competences of academics and of experts from national statistical offices to increase the dissemination of the most recent survey methods in the agricultural sector. With this ambition in mind, the authors were asked to extend and update their research and the project was completed by some chapters on specialized topics.

Although the present book can serve as a supplementary text in graduate seminars in survey methodology, the primary audience is constituted by researchers having at least some prior training in sampling methods. Since it contains a number of review chapters on several specific themes in survey research, it will be useful to researchers actively engaged in organizing, managing and conducting agricultural surveys who are looking for an introduction to advanced techniques from both a practical and a methodological perspective.

Another aim of this book is to stimulate research in this field and, for this reason, we are aware that it cannot be considered as a comprehensive and definitive reference on the methods that can be used in agricultural surveys, since many topics were intentionally omitted. However, it reflects, to the best of our judgement, the state of the art on several crucial issues.

The volume contains 22 chapters of which the first one can be considered as an introductory chapter reviewing the current status of agricultural statistics, and the remaining 21 are divided into five parts:

- I. Census, frames, registers and administrative data (Chapters 2–5). These chapters provide an overview of the basic tools used in agricultural surveys, including some practical and theoretical considerations regarding the definitions of the statistical units. Attention is then focused on the use of administrative data that in the last few years have evolved from a simple backup source to the main element in ensuring the coverage of a list of units. The opportunity to reduce census and survey costs implies growing interest, among statistical agencies, in the use of administrative registers for statistical purposes. However it requires attitudes, methods and terms that are not yet typical in the statistical tradition. The keyword is the harmonization of the registers in such a way that information from different sources and observed data should be consistent and coherent. In particular, the expensive agricultural census activities conducted periodically in every country of the world should benefit from such a radical innovation.
- II. Sample design, weighting and estimation (Chapters 6–9). These chapters review advanced methods and techniques recently developed in the sampling literature as applied to agricultural units, in the attempt to address the distinctive features (c)–(e) described above. Some interesting proposals arise from the field of small-area estimation, which has received a lot of attention in recent years due to the growing demand for reliable small-area statistics needed for formulating policies and programmes. An appraisal is provided of indirect estimates, both traditional and model-based, that are used because direct area-specific estimates may not be reliable due to small-area-specific sample sizes.
- III. GIS and remote sensing (Chapters 10–13). These chapters describe the use of the Geographic Information System technology as a tool to manage area- and point-based surveys. These devices are applied to carry out a wide range of operations on spatial information retrieved from many kinds of mixed sources. They

provide a detailed description of the procedures currently used in the European Union and United States to develop and sample area frames for agricultural surveys. Additionally, the usefulness of remotely sensed data as the main auxiliary variables for geographically coded units is assessed through empirical evidence, and some techniques to increase the performance of their use are proposed.

- IV. Data editing and quality assurance (Chapters 14–17). These chapters deal with the classical problem of error handling, localization and correction. This is strictly connected with the issue of guaranteeing data quality, which obviously plays a central role within any statistical institute, both in strategic decisions and in daily operations. In this framework, it is evident that quality is not as much concerned with the individual data sets as with the whole set of procedures used. Some approaches to ensure data quality in collecting, compiling, analysing and disseminating agriculture data are described.
- V. Data dissemination and survey data analysis (Chapters 18–22). These chapters examine some experiences in the use of statistical methods to analyse agricultural survey data. In particular, regression analysis (or some of its generalizations) is quite often applied to survey microdata to estimate, validate or forecast models formalized within agricultural economics theory. The purpose is to take into account the nature of the data analysed, as observed through complex sampling designs, and to consider how, when and if statistical methods may be formulated and used appropriately to model agricultural survey data. Web tools and techniques to assist the users to access statistical figures online are then described, for complete, safe and adequate remote statistical analyses and dissemination.

We would like to thank Daniel Berze of the International Statistical Institute for useful advice and suggestions during the starting phase of the project. Thanks are also due to Susan Barclay, Heather Kay and Richard Davies of John Wiley & Sons, Ltd for editorial assistance, and to Alistair Smith of Sunrise Setting Ltd for assistance with LaTeX.

Finally, we are grateful to the chapter authors for their diligence and support for the goal of providing an overview of such an active research field. We are confident that their competence will lead to new insights into the dynamics of agricultural surveys methods, to new techniques to increase the efficiency of the estimates, and to innovative tools to improve the timeliness and comprehensiveness of agricultural statistics.

Roberto Benedetti

Marco Bee

Giuseppe Espa

Federica Piersimoni

Pescara, Trento and Rome

November 2009

Contents

List of Contributors	xvii
Introduction	xxi
1 The present state of agricultural statistics in developed countries: situation and challenges	1
1.1 Introduction	1
1.2 Current state and political and methodological context	4
1.2.1 General	4
1.2.2 Specific agricultural statistics in the UNECE region	6
1.3 Governance and horizontal issues	15
1.3.1 The governance of agricultural statistics	15
1.3.2 Horizontal issues in the methodology of agricultural statistics	16
1.4 Development in the demand for agricultural statistics	20
1.5 Conclusions	22
Acknowledgements	23
Reference	24
 Part I Census, Frames, Registers and Administrative Data	 25
2 Using administrative registers for agricultural statistics	27
2.1 Introduction	27
2.2 Registers, register systems and methodological issues	28
2.3 Using registers for agricultural statistics	29
2.3.1 One source	29
2.3.2 Use in a farm register system	30
2.3.3 Use in a system for agricultural statistics linked with the business register	30
2.4 Creating a farm register: the population	34
2.5 Creating a farm register: the statistical units	38
2.6 Creating a farm register: the variables	42

2.7	Conclusions	44
	References	44
3	Alternative sampling frames and administrative data. What is the best data source for agricultural statistics?	45
3.1	Introduction	45
3.2	Administrative data	46
3.3	Administrative data versus sample surveys	46
3.4	Direct tabulation of administrative data	46
3.4.1	Disadvantages of direct tabulation of administrative data	47
3.5	Errors in administrative registers	48
3.5.1	Coverage of administrative registers	48
3.6	Errors in administrative data	49
3.6.1	Quality control of the IACS data	49
3.6.2	An estimate of errors of commission and omission in the IACS data	50
3.7	Alternatives to direct tabulation	51
3.7.1	Matching different registers	51
3.7.2	Integrating surveys and administrative data	52
3.7.3	Taking advantage of administrative data for censuses	52
3.7.4	Updating area or point sampling frames with administrative data	53
3.8	Calibration and small-area estimators	53
3.9	Combined use of different frames	54
3.9.1	Estimation of a total	55
3.9.2	Accuracy of estimates	55
3.9.3	Complex sample designs	56
3.10	Area frames	57
3.10.1	Combining a list and an area frame	57
3.11	Conclusions	58
	Acknowledgements	59
	References	60
4	Statistical aspects of a census	63
4.1	Introduction	63
4.2	Frame	64
4.2.1	Coverage	64
4.2.2	Classification	64
4.2.3	Duplication	65
4.3	Sampling	65
4.4	Non-sampling error	66
4.4.1	Response error	66
4.4.2	Non-response	67
4.5	Post-collection processing	68
4.6	Weighting	68
4.7	Modelling	69
4.8	Disclosure avoidance	69
4.9	Dissemination	70

4.10	Conclusions	71
	References	71
5	Using administrative data for census coverage	73
5.1	Introduction	73
5.2	Statistics Canada's agriculture statistics programme	74
5.3	1996 Census	75
5.4	Strategy to add farms to the farm register	75
5.4.1	Step 1: Match data from E to M	76
5.4.2	Step 2: Identify potential farm operations among the unmatched records from E	76
5.4.3	Step 3: Search for the potential farms from E on M	76
5.4.4	Step 4: Collect information on the potential farms	77
5.4.5	Step 5: Search for the potential farms with the updated key identifiers	77
5.5	2001 Census	77
5.5.1	2001 Farm Coverage Follow-up	77
5.5.2	2001 Coverage Evaluation Study	77
5.6	2006 Census	78
5.6.1	2006 Missing Farms Follow-up	79
5.6.2	2006 Coverage Evaluation Study	80
5.7	Towards the 2011 Census	81
5.8	Conclusions	81
	Acknowledgements	83
	References	83

Part II Sample Design, Weighting and Estimation 85

6	Area sampling for small-scale economic units	87
6.1	Introduction	87
6.2	Similarities and differences from household survey design	88
6.2.1	Probability proportional to size selection of area units	88
6.2.2	Heterogeneity	90
6.2.3	Uneven distribution	90
6.2.4	Integrated versus separate sectoral surveys	90
6.2.5	Sampling different types of units in an integrated design	91
6.3	Description of the basic design	91
6.4	Evaluation criterion: the effect of weights on sampling precision	93
6.4.1	The effect of 'random' weights	93
6.4.2	Computation of D^2 from the frame	94
6.4.3	Meeting sample size requirements	94
6.5	Constructing and using 'strata of concentration'	95
6.5.1	Concept and notation	95
6.5.2	Data by StrCon and sector (aggregated over areas)	95
6.5.3	Using StrCon for determining the sampling rates: a basic model	97
6.6	Numerical illustrations and more flexible models	97

6.6.1	Numerical illustrations	97
6.6.2	More flexible models: an empirical approach	100
6.7	Conclusions	104
	Acknowledgements	105
	References	105
7	On the use of auxiliary variables in agricultural survey design	107
7.1	Introduction	107
7.2	Stratification	109
7.3	Probability proportional to size sampling	113
7.4	Balanced sampling	116
7.5	Calibration weighting	118
7.6	Combining <i>ex ante</i> and <i>ex post</i> auxiliary information: a simulated approach	124
7.7	Conclusions	128
	References	129
8	Estimation with inadequate frames	133
8.1	Introduction	133
8.2	Estimation procedure	133
8.2.1	Network sampling	133
8.2.2	Adaptive sampling	135
	References	138
9	Small-area estimation with applications to agriculture	139
9.1	Introduction	139
9.2	Design issues	140
9.3	Synthetic and composite estimates	140
9.3.1	Synthetic estimates	141
9.3.2	Composite estimates	141
9.4	Area-level models	142
9.5	Unit-level models	144
9.6	Conclusions	146
	References	147

Part III GIS and Remote Sensing 149

10	The European land use and cover area-frame statistical survey	151
10.1	Introduction	151
10.2	Integrating agricultural and environmental information with LUCAS	154
10.3	LUCAS 2001–2003: Target region, sample design and results	155
10.4	The transect survey in LUCAS 2001–2003	156
10.5	LUCAS 2006: a two-phase sampling plan of unclustered points	158
10.6	Stratified systematic sampling with a common pattern of replicates	159
10.7	Ground work and check survey	159
10.8	Variance estimation and some results in LUCAS 2006	160
10.9	Relative efficiency of the LUCAS 2006 sampling plan	161

10.10	Expected accuracy of area estimates with the LUCAS 2006 scheme	163
10.11	Non-sampling errors in LUCAS 2006	164
10.11.1	Identification errors	164
10.11.2	Excluded areas	164
10.12	Conclusions	165
	Acknowledgements	166
	References	166
11	Area frame design for agricultural surveys	169
11.1	Introduction	169
11.1.1	Brief history	170
11.1.2	Advantages of using an area frame	171
11.1.3	Disadvantages of using an area frame	171
11.1.4	How the NASS uses an area frame	172
11.2	Pre-construction analysis	173
11.3	Land-use stratification	176
11.4	Sub-stratification	178
11.5	Replicated sampling	180
11.6	Sample allocation	183
11.7	Selection probabilities	185
11.7.1	Equal probability of selection	186
11.7.2	Unequal probability of selection	187
11.8	Sample selection	188
11.8.1	Equal probability of selection	188
11.8.2	Unequal probability of selection	188
11.9	Sample rotation	189
11.10	Sample estimation	190
11.11	Conclusions	192
12	Accuracy, objectivity and efficiency of remote sensing for agricultural statistics	193
12.1	Introduction	193
12.2	Satellites and sensors	194
12.3	Accuracy, objectivity and cost-efficiency	195
12.4	Main approaches to using EO for crop area estimation	196
12.5	Bias and subjectivity in pixel counting	197
12.6	Simple correction of bias with a confusion matrix	197
12.7	Calibration and regression estimators	197
12.8	Examples of crop area estimation with remote sensing in large regions	199
12.8.1	US Department of Agriculture	199
12.8.2	Monitoring agriculture with remote sensing	200
12.8.3	India	200
12.9	The GEOSS best practices document on EO for crop area estimation	200
12.10	Sub-pixel analysis	201
12.11	Accuracy assessment of classified images and land cover maps	201
12.12	General data and methods for yield estimation	203
12.13	Forecasting yields	203