Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

11 – Models and early estimates

Roberto Gismondi

Istat, Rome, Italy; gismondi@istat.it

Loredana De Gaetano

Istat, Rome, Italy; degaetan@istat.it

INE Instituto Nacional de Estadística eurostat
Madrid, May 31 - June 3

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Summary

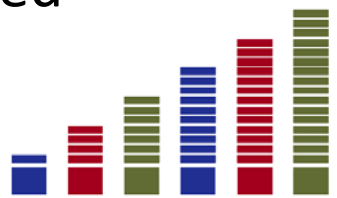
- Quality issues for Italian crop statistics
- Administrative data on land use for agricultural purposes
- New estimation strategy for crop early estimates
- Conclusions

Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Quality issues for Italian crop statistics

- ✓ ISTAT survey “Crop statistics”: agriculture surface and production from regional estimates
- ✓ “Estimative” technique mostly used, a few administrative data, no direct survey or satellite photos. Quality issues and lack of capability of local experts have to be tackled
- ✓ ISTAT survey “Early estimates” on surface use in the coming year: sample of 12.000 farmers, CATI technique
- ✓ Comparisons of year-to-year data for the same holding often difficult; problems with significant zeroes; sampling variance quite changeable depending on domains
- ✓ For both surveys some improvements are presented

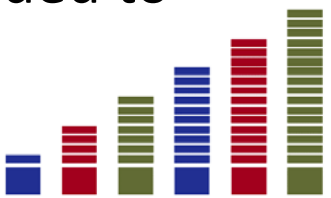


Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Administrative data on land use

- ✓ IACS (*Integrated Administration and Control System*)
- ✓ Management of subsidies to the EU agricultural holdings
- ✓ The logic underlying the IACS register is based on self-declarations as regards area used for agricultural purposes
- ✓ Potential risks:
  - missing declaration from holders
  - mistakes due to producers' declarations;
  - duplications derived from double counting of some productions
- ✓ Common "population coverage" problems which must be tackled whenever an administrative source is intended to be used for statistical purposes

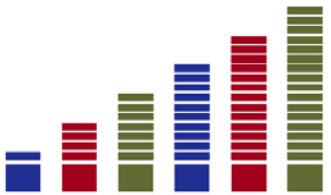


Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Administrative data on land use

- ✓ Limitations to the use of IACS data within current crop statistics mainly derive from:
  - periodicity of declarations (data are available after 6 months from the end of the reference year)
  - the need to manage properly and gradually the overlapping between this data source and estimates carried out by Italian Regions
  - scarce use of administrative data in agriculture
- ✓ Further effort for achieving deeper comparison between concepts and definitions adopted within the IACS and the ISTAT current crops statistics frameworks
- ✓ In the next slides some first comparative results…



Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Administrative data on land use

- ✓ Comparison among IACS data and the ISTAT crop statistics as regards 2014
- ✓ IACS data compared with ISTAT crop statistics and FSS (Farm Structure Survey) data
- ✓ Main outcomes have been resumed in table 1
- ✓ The kind of cultivations analyzed cover the 20% of Italian agricultural area: they are rice, olives, grapes, fruit and citrus fruit
- ✓ As regards fruit, additional details are presented in table 2
- ✓ The main outcome is that IACS data are aligned with crop statistics and are not systematically higher or lower, both at the whole Italy and at the geographical area levels

Q 2016
European Conference on Quality in Official Statistics

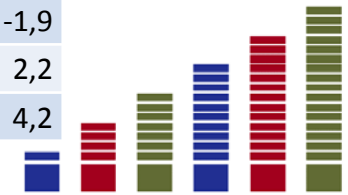
Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Administrative data on land use

Table 1 – *Agricultural surface use in 2014 - Comparison among sources (hectares)*

Source/Cultivation	Rice	Olives	Grapes	Citrus fruit	Fruit	Total
IACS						
Italy	234.813	1.119.633	653.697	106.476	377.557	2.492.176
North	229.981	17.879	253.983	17	159.437	661.298
Centre	422	176.959	101.243	313	62.238	341.175
South	4.410	924.795	298.471	106.145	155.883	1.489.703
CROP STATISTICS						
Italy	219.532	1.125.183	682.183	142.011	372.582	2.541.491
North	215.342	23.343	230.959	55	133.559	603.258
Centre	378	201.986	107.984	653	37.893	348.894
South	3.812	899.854	343.240	141.303	201.130	1.589.339
FSS 2013						
Italy	212.238	1.073.324	635.979	129.155	388.808	2.439.504
North	209.960	20.121	246.962	16	164.886	641.945
Centre	0	182.122	103.056	2.286	51.834	339.298
South	1.834	871.081	285.961	126.853	172.088	1.457.817
% DIFFERENCE (ITALY)						
IACS vs crop statistics	7,0	-0,5	-4,2	-25,0	1,3	-1,9
IACS vs FSS 2013	10,6	4,3	2,8	-17,6	-2,9	2,2
Crop statistics vs FSS 2013	3,4	4,8	7,3	10,0	-4,2	4,2

Source: elaboration on ISTAT and IACS data.



Q 2016
European Conference on Quality in Official Statistics

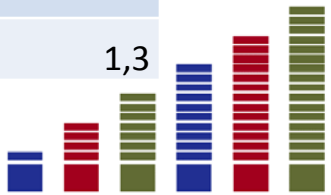
Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Administrative data on land use

Table 2 – *Fruit surface use in 2014 - Comparison among sources (hectares)*

Source/Cultivation	Nuts*	Peers	Peaches	Other fruit	Fruit
IACS					
Italy	136.531	28.278	59.141	153.607	377.557
North	21.191	26.098	24.323	87.825	159.437
Centre	32.346	576	2.829	26.487	62.238
South	82.995	1.604	31.988	39.295	155.883
CROP STATISTICS					
Italy	125.558	30.145	63.733	153.146	372.582
North	15.598	23.756	20.823	73.382	133.559
Centre	19.665	907	4.088	13.233	37.893
South	90.295	5.482	38.822	66.531	201.130
% DIFFERENCE (ITALY)					
IACS vs crop statistics	8,7	-6,2	-7,2	0,3	1,3

Source: elaboration on ISTAT and IACS data. *Hazelnut, almond, pistachio.



Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

New estimation strategy for crop early estimates

- ✓ "Crop early estimates survey" (*Cees*): carried out between November 2015 and January 2016 (CATI technique)
- ✓ Sample of 12.000 agricultural holdings for collecting early estimates regarding land use for agricultural purposes in the agrarian year (*ay*) 2015-16 (year-to-year % changes)
- ✓ 5 surface categories (EU Regulation 543/2009): common wheat, durum wheat, rye, barley, rape and turnip rape
- ✓ Stratified random sample selected from the 2010 agriculture census list using last updates form administrative sources
- ✓ Design-based Horvitz-Thompson estimator, with sampling weights adjusted for non responses
- ✓ Potential improvements regarding the sample selection and the estimation procedure

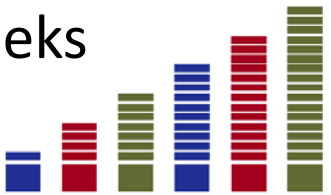
Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

New estimation strategy for crop early estimates

Sampling

- ✓ Switch from probabilistic to deterministic sampling
- ✓ Two sub-samples including 6.000 units drawn from the subsets of respondents in the following surveys: *Cees* 2015 and FSS 2013
- ✓ The samples were selected choosing the largest holders in each Italian Region which guaranteed at least the 80% of agricultural area surveyed in *Cees* 2015 and FSS 2013
- ✓ Simpler link between each sampling unit and its certified electronic postal address
- ✓ The *Cees* 2016 response rate was 74,5%, against the 65,8% obtained in the *Cees* 2015
- ✓ Shorter time needed for data editing process: 4 weeks (2016) against 6 weeks (2015)



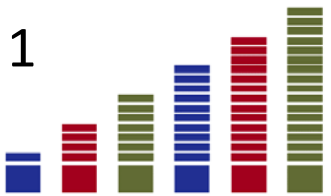
Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

New estimation strategy for crop early estimates

Definitions (two consecutive years 1 and 2)

- ✓ *Y*: surface used for a certain cultivation
- ✓ *m*: sample size at time 2. Respondents provided also surface at time 1, so *m* is the sample size at time 1 as well
- ✓ $n\lambda$: number of units which declared positive surface at both times
- ✓ $n(1-\lambda)$: number of units with positive surface at time 2 and surface equal to zero at time 1; therefore, the overall number of units which declared positive surface at time 2 is *n*
- ✓ *m-n*: number of units which declared surface equal to zero at time 2
- ✓ $\bar{y}_1$: sample mean (1) of the *m* units observed at time 1
- ✓ $\bar{y}_1'$: mean (1) of the $n\lambda$ units with positive surface at both times
- ✓ $\bar{y}_2'$: mean (2) of the $n\lambda$ units with positive surface at both times
- ✓ $\bar{y}_2''$: mean (2) of the $n(1-\lambda)$ units with surface=0 at time 1



Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

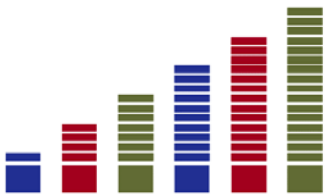
New estimation strategy for crop early estimates

Methodology

- ✓ According to double sampling, the regression estimator of the total at time 2 is $\hat{Y}_{2r} = N[\bar{y}_2' + \hat{\beta}(\bar{y}_1 - \bar{y}_1')]$, where $\hat{\beta}$ is calculated on the $n\lambda$ units with not null surfaces at both times. Under given conditions the *BLUP* of β is the ratio estimator: $\hat{\beta}^* = \bar{y}_2' / \bar{y}_1'$
- ✓ The final unbiased combined estimator given by:

$$\hat{Y}_{2c} = \phi \hat{Y}_{2r} + (1-\phi) t_2'' \quad (1)$$

- ✓ In model (1) we can put $t_2'' = \bar{y}_2''$. Used in Small Area Estimation (Rao, 2003, 2010) and estimation from multiple frame survey (Lohr and Rao, 2006)
- ✓ Optimal factor ϕ_0 depends on estimators variances



Administrative Data and Model Based Estimation in Italian Agriculture Statistics

New estimation strategy for crop early estimates

Methodology

- ✓ Alternative approach: different model as regards the $n(1-\lambda)$ agricultural holdings which declared zero surface at time 1

$$y_{2i} = \gamma\, z_i + \delta_i \qquad \text{where:} \quad \begin{cases} E_\varphi(\delta_i) = 0 & \forall i \\ V_\varphi(\delta_i) = \theta^2\, z_i & \forall i \\ Cov_\varphi(\delta_i, \delta_j) = 0 & \text{if} \quad i \neq j \end{cases}$$

- ✓ We can calculate the estimator:

$$\hat{t}_2'' = N[\bar{y}_2'' + \hat{\gamma}(\bar{z} - \bar{z}'')] \qquad \text{where} \quad \hat{\gamma}^* = \bar{y}_2'' / \bar{z}'' \quad (2)$$

where z is given by agricultural surface referred to 2010 as from the last agriculture census

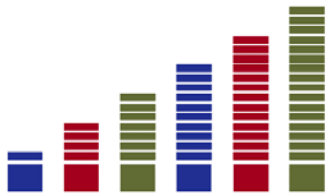
Administrative Data and Model Based Estimation in Italian Agriculture Statistics

New estimation strategy for crop early estimates

- ✓ The table 3 resumes the 5 estimations strategies compared
- ✓ Strategy (I) used until the *Cees* 2015. Strategy (IV) applied in *Cees* 2016

Table 3 – *Compared estimation strategies for crop early estimates*

Code	Methodology	Estimator time 1	Estimator time 2
(I)	Sample mean expansion	$N\,\bar{y}_1$	$N\,\bar{y}_2$
(II)	Sample mean expansion using only units with positive surfaces at both times	$N\,\bar{y}_1'$	$N\,\bar{y}_2'$
(III)	Use of (1) where $\phi = 1$	Crop statistics	$N[\bar{y}_2' + \hat{\beta}(\bar{y}_1 - \bar{y}_1')]$
(IV)	Use of (1) where $t_2'' = \bar{y}_2''$, $\phi = \phi_0$	Crop statistics	$\phi_0 \hat{Y}_{2r} + (1-\phi_0)\,\bar{y}_2''$
(V)	Use of (1) where t_2'' is calculated as defined in (2), $\phi = \phi_0$	Crop statistics	$\phi_0 \hat{Y}_{2r} + (1-\phi_0)\,\hat{t}_2''$



Administrative Data and Model Based Estimation in Italian Agriculture Statistics

New estimation strategy for crop early estimates

✓ On average strategy IV is characterized by the smallest MSE

Table 4 - *Main results of compared estimation strategies (agrarian year 2015-16) – Agricultural surfaces % changes and coefficient of variation (Cv) of estimates*

Strategy	Arable land	Common wheat	Durum wheat	Barley	Oat	Grain Maize	Sum of 5 crops
(I)	-0,3	-1,6	-0,5	2,1	7,4	-3,0	-0,8
	(3,6)	*(8,9)*	*(15,7)*	*(14,5)*	*(12,6)*	*(17,7)*	*(7,9)*
(II)	0,9	2,5	2,3	3,3	9,1	-5,1	1,0
	(4,4)	*(9,5)*	*(11,6)*	*(15,0)*	*(11,9)*	*(17,5)*	*(7,3)*
(III)	0,5	1,5	0,7	0,8	4,2	-2,5	0,3
	(4,8)	*(9,5)*	*(14,8)*	*(15,3)*	*(15,1)*	*(16,7)*	*(7,9)*
(IV)	**2,4**	**5,6**	**6,2**	**6,9**	**11,2**	**-3,9**	**3,8**
	(2,7)	***(7,8)***	***(9,2)***	***(9,5)***	***(8,4)***	***(13,4)***	***(5,4)***
(V)	2,9	6,2	7,1	9,5	10,0	-4,3	4,6
	(2,8)	*(8,3)*	*(10,1)*	*(9,3)*	*(9,0)*	*(15,5)*	*(5,8)*

Source: elaboration on ISTAT data. CVs are into squared brackets.

Q 2016
European Conference on Quality in Official Statistics

Administrative Data and Model Based Estimation in Italian Agriculture Statistics

Conclusions

Crop statistics

- ✓ Administrative data collected by the Italian agency for payment in agriculture can be used for statistical purposes
- ✓ Further work should concern:
  - extension of the database to 2015 and to other cultivations
  - deeper comparison between classification criteria adopted by IACS and in the current crop statistics context

Early estimates

- ✓ Sampling design based on a deterministic approach coupled with a model based estimation technique
- ✓ The presence of many zeroes implies the use of specific models whenever the traditional regression model may fail