Accuracy and precision of the Brazilian livestock data (Census x Estimation): Bovine, Poultry and Pork

Paula Marques Meyer *
Brazilian Institute of Geography and Statistics (IBGE), Pirassununga/SP, Brazil – paula.meyer@ibge.gov.br

Paulo Henrique Mazza Rodrigues
University of Sao Paulo (USP/FMVZ), Pirassununga/SP, Brazil – pmazza@usp.br

Abstract

Brazil is a leading player in the beef, poultry and pork world markets, ranking as the world's largest meat exporter. It is the largest exporter of beef (home to the world's biggest commercial cattle) and poultry (being the 3rd world's producer), as well as the 4th largest producer and exporter of pork. The Brazilian Institute of Geography and Statistics (IBGE), responsible for the statistical information, performs two surveys for livestock: a) the Municipal Livestock Research (called PPM), by estimation, and b) the Livestock Census. In order to assess the reliability of the surveys throughout the Census years (1975, 1980, 1985, 1995 and 2006), data about bovine (heads), poultry (x 1,000 heads) and pork (heads) were collected by states from Census and PPM database. For the evaluation, accuracy and precision of the surveys were tested. Regarding bovine cattle, it can be stated that accuracy and precision, respectively, have lowered over the years, what can be explained by the increase in the herd size (2006 Census was the least accurate). The estimations overestimated the values, as most biases were positive and also the states with larger cattle sizes were more susceptible to errors, i.e. their biases were positively greater. Concerning poultry, the 1975 and 1980 Censuses were not accurate. Over the years, there was reduction in the estimates precision compared to the Census, especially for 1995 and 2006. The states with larger chicken populations were more susceptible to errors, i.e. their biases were positively greater (1975, 1980 and 1985) and tended to be (1995), or were negatively greater (2006), meaning that estimations were higher or lower than Census data, respectively. Regarding pork, it is possible to state that the pig population declined throughout the years (for both censused and estimated values), unlike bovine and poultry. The 1975, 1980 and 1995 Censuses were not accurate, despite the strong indicative of increasing mean bias in the latest years. The residual errors rose over the years, even without growth of pig population. There is indication that the loss in accuracy and precision over the years was no longer dependent on the pig population of the states. So, caution should be paid for states with larger bovine herds and chicken populations, due to the lower ability of the estimation to be consistent with the censused values. Either the estimation may be overestimating or the Census may be underestimating the values, even by under-declaration by the ranchers.

Keywords: cattle; chicken; pig; sensitivity analysis.

1. Introduction

Brazil is a world-leading player in meat markets (beef, poultry and pork) and grain production, ranking as the world's largest meat exporter. Expressed in numbers, it is the world's largest producer of sugarcane, oranges and coffee, second largest producer of soybeans, behind the United States, and third largest producer of corn (FAO, 2014a). It is the largest exporter of beef (being home to the world's biggest commercial cattle) and poultry (being the third world's producer, only after USA and China) and concerning pork, it is the fourth largest producer and exporter (FAO, 2014b). It is also an international leader in renewable sources of energy such as hydroelectric power and biofuel (sugar-based ethanol) production. The country is also a pioneer for use of ethanol as a motor fuel (Meyer *et al.*, 2013).

The Brazilian Institute of Geography and Statistics (IBGE) is the key agency responsible for the statistical information in Brazil, and performs two methods of surveys for agricultural and livestock production: a) the Municipal Agricultural/Livestock Research, and b) the Agricultural/Livestock Census. The outputs of these surveys differ and the magnitude of the variation also differs according to the years, the states/municipalities and the surveyed variables. In order to assess the reliability of the surveys throughout the available Census years, with regard to the main livestock raised in Brazil, a study was conducted using sensitivity analysis by evaluating the accuracy and precision of the two surveys forementioned.

2. Methods

Data collection

The Brazilian Institute of Geography and Statistics (IBGE) performs two methods of surveys for livestock production in Brazil: a) the Municipal Livestock Research (so called PPM), done annually by estimation, and b) the Livestock Census, done in loco at the ranches, by applying a standard questionnaire to the ranchers. These data are published by regions, states and municipalities. Brazil presents 26 states plus 1 Federal District, grouped in 5 main political regions: Rondônia, Acre, Amazonas, Roraima, Pará, Amapá and Tocantins (North region), Mato Grosso, Mato Grosso do Sul, Goiás and Distrito Federal (Midwest region), Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe and Bahia (North-east region), Minas Gerais, Espírito Santo, Rio de Janeiro and São Paulo (South-east region), and Paraná, Santa Catarina and Rio Grande do Sul (South region).

Data about bovine cattle (heads), poultry (x 1,000 chicken heads) and pork (heads) by states were collected at Digital Database 'Sidra' (IBGE, 2015) for the conducted Census years (1975, 1980, 1985, 1995 and 2006 – years available online), as well as their corresponding data from PPM (estimations).

Evaluation of the surveys

Sensitivity analysis is the study of how the uncertainty in the output of a model or system can be apportioned to different sources of uncertainty in its inputs. So, for this analysis, accuracy, precision and robustness of the models are tested.

For evaluating the surveys of the present study, the accuracy and precision were measured and tested by comparing the predicted values (obtained by PPM) against the observed values (released by the Censuses). The accuracy was assessed by the mean bias. The mean bias is the difference between the predicted and the observed value and represents the average inaccuracy of model prediction across all data (Kohn *et al*., 1998).Therefore, the most accurate model (or survey) is the one with the mean bias closest to zero. The accuracy is measured by the following equation:

$$\text{Mean bias} = \frac{\sum(\text{predicted} - \text{observed})}{\text{number\_of\_observations}}$$

Precision is a dispersion measure between the predicted and the observed values, i.e., the average variability of the distance between the predicted and observed values. It can be evaluated by the root mean square prediction error (RMSPE) or by the residual error. The RMSPE measures how much the predictions fit well to the observed data (Bibby & Toutenburg, 1977) and was calculated by the following equation:

$$\text{RMSPE} = \sqrt{\frac{\sum(\text{predicted} - \text{observed})^2}{\text{number\_of\_observations}}}$$

However, whenever the mean bias is high (lack of accuracy), there will be an overestimation of the lack of precision, i.e. an overestimation of the RMSPE, because the average distance between the predicted and observed values also leads to an increased variability between the predicted and

observed values. Thus, the precision is better evaluated when the RMSPE is corrected for the lack of accuracy, i.e. the residual error, which is defined by the remaining error in the prediction model excluding the error due to the mean bias. It is also referred as the prediction error excluding the mean bias, and is obtained by the following equation:

$$\text{Residual error} = \sqrt{\left[RMSPE^2 - (\text{mean bias})^2\right]}$$

The linear bias was derived from the regression of the residuals (predicted value – observed value) against the observed values and can be used to identify whether or not the magnitude of the bias increases or decreases with the magnitude of the observed values.

For a better comprehension, in this study, the terms "model" will be called "survey", as well as "observed values" as "censused values" and "predicted values" as "estimated values" (obtained from PPM).

Statistical analysis

For comparing the accuracy between the surveys, the mean bias was subjected to analysis of variance (*F*-test) and the comparison between means was done by *Tukey*'s test (5%). For calculating if the mean value was significant, i.e. if it is significantly different from zero, the *T*-test was used for mean equals to zero (mean=0), using the PROC UNIVARIATE (SAS, 2010). For comparing the precision between surveys, the residual error values were subjected to the test of homogeneity of variances (*Hartley*'s test), using PROC TTEST (SAS, 2010), according to Ott (1993). The linear bias was derived from the regression between the bias (estimated – censused value) against censused value by PROC REG (SAS, 2010). The comparison between the slope coefficients was done by the method for comparing two straight lines, using analysis of variance (F-test) and compared pair-wise, by PROC GLM (SAS, 2010).

3. Results and Discussion

In 2006, the bovine herd size revealed by the Census was over 176.1 million heads, while the estimated value by PPM was approximately 205.9 million for the corresponding year. For poultry, the 2006 Census presented over 1.1 billion chicken heads, while the PPM estimated 1.01 billion heads. Regarding pork, the value censused for pig population in 2006 was 31.2 million heads, while the value estimated by PPM was 35.2 heads.

By the data presented in Table 1 (Cattle), Table 2 (Poultry) and Table 3 (Pork), the accuracy and precision varied among the livestock species when comparing the values from the Livestock Census and PPM (estimates) throughout the years. It can be evidenced that all the values estimated from PPM were greater than the ones released by the Census, with only two exceptions for cattle data in 1985 and for poultry in 2006. Nevertheless, for pig population, all the estimated values were greater than the censused ones for all the years. Considering these assumptions, the results obtained for the different livestock species will be presented and discussed separately in this paper.

As results shown in Table 1, concerning bovine cattle, by the mean bias and residual error, it can be stated that accuracy and precision, respectively, have lowered over the years, what can be explained by the increase in the herd size and, therefore, the estimation methods become vulnerable over the years when the bovine herd increases in size. By the mean bias, the 2006 Census was the least accurate, with a bias of over 1 million heads per state, on average. In general, it can be stated that the estimations overestimated the values, as most biases were positive. The increasing residual error over the years indicates reduction in the estimates precision compared to the Census, especially for the latest years (1995 and 2006). By the graphic evaluation (not shown) and the linear biases, it was found that the states with larger herd sizes were more susceptible to errors, i.e. their biases were positively greater, with the exception of 1985 Census with negative bias. In the last Census, carried out in 2006, the largest bovine herds were found in the states of Mato Grosso, Mato Grosso do Sul and Minas

Gerais. So, caution should be paid for states with larger herd sizes, due to the lower ability of the estimation to be consistent with the censused values.

Table 1 – Accuracy (mean bias) and precision (residual error) for cattle data collected from Census and PPM (estimations)

	1975	**1980**	**1985**	**1995**	**2006**
n [1]	24	26	25	27	27
Censused cattle	3,737,176.67	4,541,754.15	4,398,455.16	5,668,825.00	6,523,981.52
Estimated cattle	3,774,574.42	4,575,823.77	4,354,862.24	5,971,405.11	7,625,416.44
Mean bias [2,3]	37,397.75 B	34,069.62 B	-43,592.92 B	302,580.11 B	1,101,434.93 A**
RMSPE [4]	93,292.99	106,790.20	297,472.68	792,943.68	1,422,451.55
Residual error [5]	85,469.24 C	101,209.72 C	294,407.61 B	732,942.67 A	900,116.39 A
	Regression of bias against censused cattle [6]				
Linear bias [7]	0.0136 C	0.0101 C	-0.0080 C	0.0740 B	0.1753 A
R^2 [8]	0.6044	0.2858	0.0222	0.3468	0.7345
Pr > | t | [9]	0.0001	0.0049	0.4774	0.0012	0.0001

[1] Number of states. [2] Means within line, followed by different letters, differ (P<0.05). [3] Mean different from 0: * (P<0.05), ** (P<0.01). [4] Root mean square prediction error. [5] Residual errors within line, followed by different letters, differ (P<0.05). [6] Regression of bias (estimated cattle – censused cattle) against censused cattle. [7] Slope coefficient of regression of bias (estimated cattle – censused cattle) against censused cattle. Values within line, followed by different letters, differ (P<0.05). [8] Coefficient of determination for the linear model of regression of bias against the censused cattle. [9] Statistical probability of considering the linear bias equals to 0.

As results regarding the poultry production (Table 2), by analyzing the mean bias, it can be stated that the 1975 and 1980 Censuses were not accurate, as their positive biases differed from zero, while 2006 Census showed the highest bias, despite being negative and not different from 0. The increasing residual error over the years indicates reduction in the estimates precision compared to the Census, especially for 1995 and 2006 Censuses. By the graphic evaluation (not shown) and the linear biases, it was found that states with larger chicken populations were more susceptible to errors, i.e. their biases were positively greater (1975, 1980 and 1985) and tended to be (1995), or were negatively greater (2006), meaning that estimations were higher or lower than Census data, respectively. In the last Census, carried out in 2006, the largest chicken populations were found in the states of São Paulo, Paraná and Santa Catarina. So, once again, caution should be paid for states with larger chicken populations, due to the lower ability of the estimation to be consistent with the censused values.

Table 2 – Accuracy (mean bias) and precision (residual error) for poultry data (x 1,000 chicken heads) collected from Census and PPM (estimations)

	1975	**1980**	**1985**	**1995**	**2006**
n [1]	24	26	25	27	27
Censused chicken	11,732.63	15,891.58	16,928.24	26,612.56	42,350.30
Estimated chicken	12,757.29	16,973.96	18,121.04	27,019.74	37,463.56
Mean bias [2,3]	1,024.67 A**	1,082.38 A**	1,192.80 A	407.19 A	-4,886.74 A
RMSPE [4]	1,491.56	1,633.06	1,955.10	12,630.68	14,448.89
Residual error [5]	1,088.01 B	1,222.84 B	1,549.08 B	12,624.11 A	13,597.43 A
	Regression of bias against censused chicken [6]				
Linear bias [7]	0.0841 A	0.0578 B	0.1052 A	-0.0969 BC	-0.1847 C
R^2 [8]	0.7836	0.6974	0.7040	0.0972	0.7568
Pr > | t | [9]	0.0001	0.0001	0.0001	0.1134	0.0001

[1] Number of states. [2] Means within line, followed by different letters, differ (P<0.05). [3] Mean different from 0: * (P<0.05), ** (P<0.01). [4] Root mean square prediction error. [5] Residual errors within line, followed by different letters, differ (P<0.05). [6] Regression of bias (estimated chicken – censused chicken) against censused chicken. [7] Slope coefficient of regression of bias (estimated chicken – censused chicken) against censused chicken. Values within line, followed by different letters, differ (P<0.05). [8] Coefficient of determination for the linear model of regression of bias against the censused chicken. [9] Statistical probability of considering the linear bias equals to 0.

With regards to pork production (results shown in Table 3), it is possible to state that the pig population declined throughout the years (for both censused and estimated values, in different extents), unlike other species of livestock, like bovine and poultry, which had their populations increased over the years. By observing the mean bias, the 1975, 1980 and 1995 Censuses were not accurate, despite the strong indicative of increasing mean bias in the latest years. The residual errors rose over the years, even without growth of pig population (or even with reduction in its average), showing that the data published in 2006 were the least precise ones. By the graphic evaluation (not shown) and linear biases, it was found that the linear biases declined numerically (especially for 1985, 1995 and 2006 not differing from 0), therefore indicating that the loss in accuracy and precision over the years was no longer dependent on the pig population of the states.

Table 3 – Accuracy (mean bias) and precision (residual error) for pig population data collected from Census and PPM (estimations)

	1975	**1980**	**1985**	**1995**	**2006**
n [1]	24	26	25	27	27
Censused pigs	1,425,782.50	1,254,949.31	1,145,710.88	1,030,046.07	1,155,161.15
Estimated pigs	1,524,871.79	1,320,432.15	1,193,955.80	1,335,633.44	1,302,734.22
Mean bias [2,3]	99,089.29 AB**	65,482.85 B**	48,244.92 B	305,587.37 A**	147,573.07AB
RMSPE [4]	131,597.10	80,879.26	144,732.02	373,596.66	403,543.01
Residual error [5]	86,597.40 CD	47,470.53 D	136,454.33 BC	214,920.51 AB	375,591.43 A
	Regression of bias against Censused pigs [6]				
Linear bias[7]	0.0415 A	0.0369 A	0.0185 A	0.0244 A	-0.0400 A
R^2 [8]	0.3114	0.5550	0.0295	0.0073	0.0306
Pr > | t | [9]	0.0044	0.0001	0.4116	0.6725	0.3826

[1] Number of states. [2] Means within line, followed by different letters, differ (P<0.05). [3] Mean different from 0: * (P<0.05), ** (P<0.01). [4] Root mean square prediction error. [5] Residual errors within line, followed by different letters, differ (P<0.05). [6] Regression of bias (estimated pigs – censused pigs) against censused pigs. [7] Slope coefficient of regression of bias (estimated pigs – censused pigs) against censused pigs. Values within line, followed by different letters, differ (P<0.05). [8] Coefficient of determination for the linear model of regression of bias against the censused pigs. [9] Statistical probability of considering the linear bias equals to 0.

Unfortunately, other publications or studies related to this issue (comparison between the censused and the estimated values) were not found in the Brazilian literature. However, Meyer & Rodrigues (2014), by studying exclusively the Census database concerning cattle and pasture, presented that throughout the country, as cattle grew by 1.5%/year, total pasture dropped (–0.5%/year), and the stocking rate (bovine heads/ha of pasture) grew by 2.1%/year, from 1985 up to 2006. The authors also evaluated the relevance of the Census dataset and pointed out that the Agricultural/Livestock Census is the only assessment that reaches all the farms (the productive units and the ones for subsistence of the families), generating a huge number of variables, and allowing, in some cases, some variable crossings. About the quality, they believed that if there is a bias, this bias is distributed among the variables, once the methodology is the same throughout the data collection in every part of the country. In their opinion, even though there is another official research about livestock (PPM) – also collected and published by the only official Institute (IBGE) – published annually, this only estimates (not counting in loco) the number of animal heads, and therefore not allowing an in-depth study like they accomplished. So, their results were only possible to be obtained because of the range of variables collected and published by the Census.

For the three species of livestock studied in the present study, it is not possible to assure if the estimation values may be overestimating the real ones or if the Census may be underestimating the values, even by under-declaration by the ranchers, even though the confidentiality is ensured by law.

5. Conclusions

As conclusions, it can be stated that the accuracy and the precision varied among the livestock species studied, when comparing the values from the Livestock Census and PPM (estimates) throughout the years. It could be evidenced that, in most of the situations (depending on the years and the livestock species), the values estimated from PPM were greater than the ones released by the Census. Especially for cattle and poultry, caution should be paid for states with larger bovine herds and chicken populations, due to the lower ability of the estimation to be consistent with the censused values. Unfortunately, it is not possible to assure if either the estimation may be overestimating or if the Census may be underestimating the values, even by under-declaration by the ranchers.

References

Bibby J. & Toutenburg, H. (1977). *Prediction and improved estimation in linear models*. London, England: John Wiley & Sons.

FAO (2014a). 'Statistics – FAOSTAT.' (Food and Agriculture Organization: Rome). Available at http://faostat.fao.org/site/567/default.aspx#ancor [Verified 20 February 2014].

FAO (2014b). 'Statistics – FAOSTAT.' (Food and Agriculture Organization: Rome). Available at http://faostat.fao.org/site/339/default.aspx [Verified 20 February 2014]

IBGE (2015). 'Banco de dados agregados Sidra – IBGE (Instituto Brasileiro de Geografia e Estatística: Rio de Janeiro). Available at http://www.sidra.ibge.gov.br [Verified 10 December 2015]

Kohn, R.A., Kalscheur, K.F. & Hanigan, M. (1998). Evaluation of models for balancing the protein requirements of dairy cows. *J. Dairy Sci.*, 81, 3401-3414.

Meyer, P.M., Rodrigues, P.H.M. & Millen, D.D. (2013). Impact of biofuel production in Brazil on the economy, agriculture, and the environment. *Anim. Frontiers*, 3, 28–37.

Meyer, P.M. & Rodrigues, P.H.M. (2014). Progress in the Brazilian cattle industry: an analysis of the Agricultural Censuses database. *Anim. Prod. Sci.*, 54, 1338–1344.

Ott, R.L. (1993). *An introduction to statistical methods and data analysis*. 4th ed. Belmont, CA: Wadsworth.

SAS (2010). *SAS/STAT 9.3 user's guide: statistics*. Cary, NC: SAS Institute.