

# faoswsProduction: A sub-module for the imputation of missing time series data in the Statistical Working System - Livestock items

Francesca Rosa  
Food and Agriculture Organization  
of the United Nations

---

## Abstract

This vignette provides a detailed description of the usage of functions in the **Livestock production** package.

*Keywords:* Imputation, Linear Mixed Model, Ensemble Learning.

---

## 1. Overall Workflow

The key features of the production sub-modules (both the Crop production and the Livestock production sub-modules) are:

- The common theoretical framework based on the *Ensemble approach* developed to produce data-imputations;
- The production imputation process based on the concept of triplet. The triplet contains three elements linked through the relation:

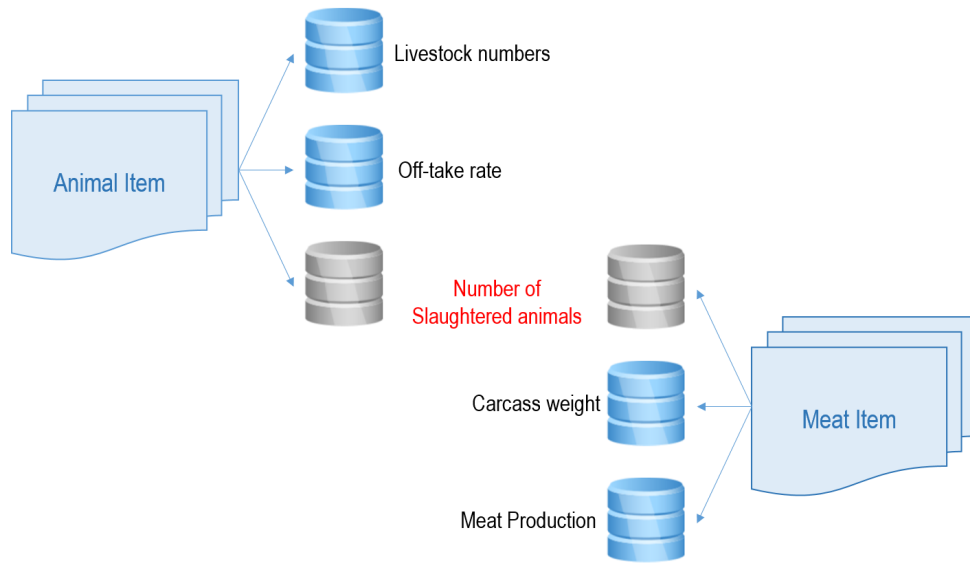
$$output_t = input_t * productivity_t$$

This identity is the basic equation to compute the production quantities (where it is missing) and to validate the output of the module, since we can have an a priori information on feasible *productivities* and/or *inputs* associated to different productive processes.

The specificities of the Livestock sub-module consist in the elements<sup>1</sup> involved, which are not three as usual, but five. In particular we can talk about two different triplets with one element in common. One triplet is associated to the *live animal item*, and its output consists in the *Number of animal slaughtered*. The other triplet is associated to the *derived items* obtained through the slaughtering process, in this latter case the *Number of animal slaughtered* is the input.

---

<sup>1</sup>With the term "Elements" (familiar to the SWS users), we mean the variable/indicator that identifies the measure. It also contains the unit of measurement. For example the element 5510 identifies the "Production [tons]".



The imputation process should be driven by the reliability of available information and it is highly recommended to start with *Livestock numbers* which represent the basis to estimate the number of slaughtered animals which again represent the basis to estimate those items obtained as result of the process of slaughtering animals (meat, hides and skins..) . On the other hand, for many countries we might dispose of official data for production of meat. The current approach has been developed in order to proper exploit this double source of information to produce a coherent and consistent output.

## 2. The animal Triplet

The triplet referring to live animals involves the following three elements:

- **Livestock numbers** - The process starts with the imputation of Livestock numbers. The adopted methodological framework is the *ensemble approach*. It is the first imputed series since we dispose of a broad and reliable set of official/semi-official data. This means that for almost all the countries we can count on a solid basis to produce imputations for filling few gaps in the analyzed time series.

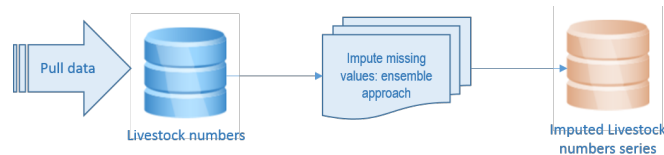


Figure 1: example

- **Off-take rate** - The *off-take rate* is linked to the natural reproduction of the different kind of animals (chickens reach the slaughtering age in 5-6 months while cows need more than one year). This element does not exist in the Statistical Working System and it is artificially computed as identity using the *Livesock numbers* and the *Number of animal slaughtered*. Indeed in this case the triples is represented by the identity:

$$SlaughteredAnimal_t = LivestockNumbers_t * OffTakeRate_t$$

*Off-take rates* can be computed as long as we dispose of Official or Semi-Official *Slaughtered animal* series. The resulting Off-take rates series still contains many gaps that are imputed through the *ensemble approach*.

The output of this step is **stored in a shared folder**. Each time an user launch the Livestock plugin from the SWS, csv tables (by animal) are stored in the `sws_r_share_drive` (be sure to have mapped this folder in your computer in order to be able to browse and analyze the off-take rates.).

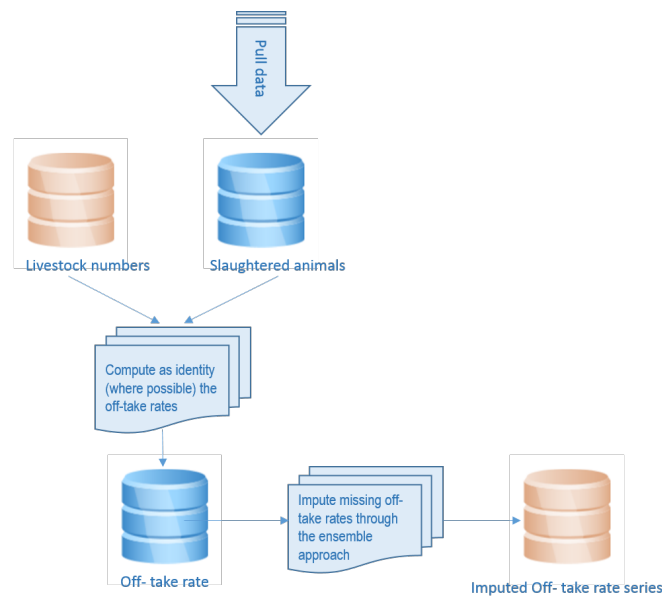


Figure 2: example

- **Number of animal slaughtered** - once both *Livestock numbers* and *Off-take rates* have been properly computed we can proceed computing as identity the *Number of animal Slaughtered* that is the output of this first process and will be the input of the next one.

## 2.1. The role of trade

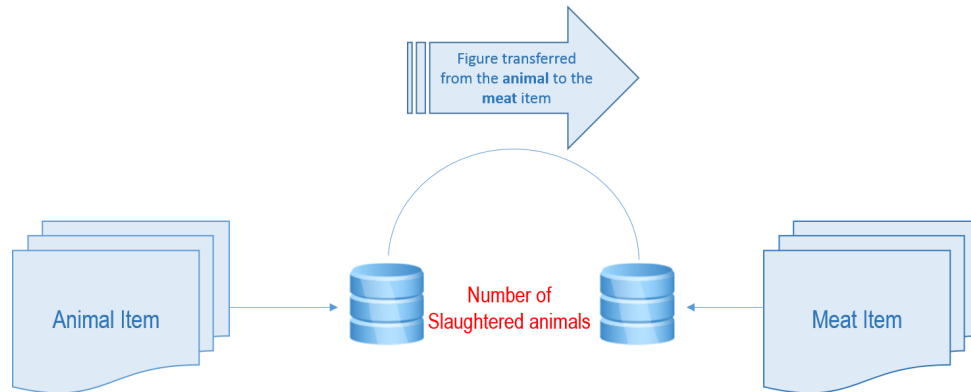
The *Number of slaughtered animals* depends on *Livestock numbers*. Many countries import animals for slaughtering purposes. This means that the total *Livestock numbers* should take into account also the inflow and the outflow of animals from the country.

The *Number of slaughtered animals* depends on the *Livestock numbers*. Many countries import animals for slaughtering purposes. This means that the total *Livestock numbers* should take into account the inflow and the outflow of each country. The current version of the plugin computes the number of slaughtered animals considering only Livestock numbers and considers negligible the animal-trade. The **R** routines have been written in order to take into account an accurate analysis of the output The effect of trade was significant only in a few sets of countries. A possible improvement of the sub-module would consist of re-activate this feature only for those countries where it is really necessary.

## 3. The meat and non-meat item triplet

The triplet referring to the meat items involves the following three elements:

- **Number of animal slaughtered** - This element is transferred in bulk from the element referring to the Animal item. All the transferred figures are flagged with a method flag equal to "c" which means *copied from elsewhere in the SWS*.



There is a specific function that synchronizes the two series (the one referring to the animal with the one referring to the meat / non-meat items). The *shares* stored in the dataset *aupus share* (if any) are used to synchronize this series. When values are transferred from parents to children, shares are applied, this means that we multiply by the shares the series referring to the parent (animal item) in order to obtain the series referring to the child (the meat/ non-meat item). In the reversed direction, the values are divided by the shares. This means that the user has the possibility to decide the percentage of the *number of slaughtered animals* for producing meat, starting from the total number of *slaughtered animals*. The *shares* are almost everywhere equal to 1 and if no shares are available the assumption is that they are equal to 1 by default.

- **Carcass Weight** - It is the element playing the role of *productivity*. It identifies the quantity of meat that can be obtained from slaughtering one animal. We have a priori information at least on the feasible ranges where the imputed figures should lie. This means that it is the element on which ex post checks are performed.
- **Meat Production** - It is the final output of the process. It is imputed through the *ensemble approach* simultaneously with the other two elements of the triplet or balanced as identity if official/semi-official figures of the other two variables <sup>2</sup> exist.

At the end of this process an important check in the *Carcas weight* element is performed. This exigency arose to embed some a priori information into the Livestock sub-module. In particular we check the consistency of the resulting *carcass weight* element.

The a priori information embedded in the module consist in minimum and maximum boundaries for the *carcass weigth* of each animal. This information is contained in a data table stored in the SWS:

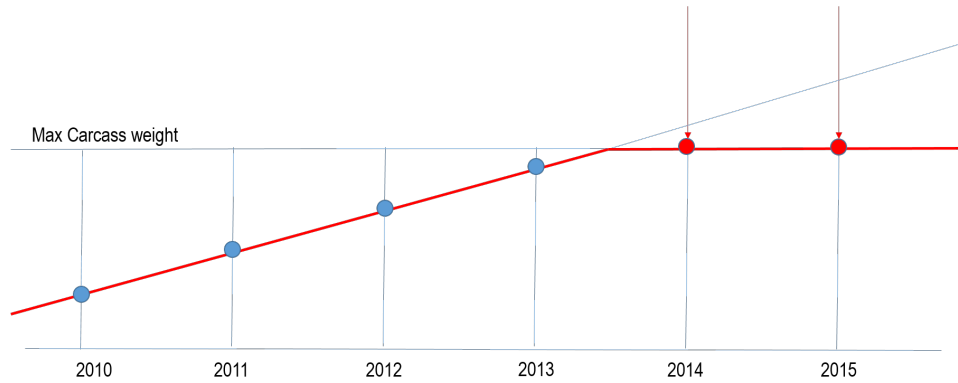
	measuredItemCPC	label	min	max	unit
1:	21111.01	Meat of cattle	60	450	kg
2:	21112	Meat of buffalo	80	450	kg
3:	21113.01	Meat of pig	10	165	kg
4:	21114	Meat of rabbits and hares	600	3000	grams
5:	21115	Meat of sheep	6	50	kg

<sup>2</sup>There is a specific function, *imputeProductionTriple*, which performs the ensemble approach for the elements of the triplet and balances the other two accordingly.

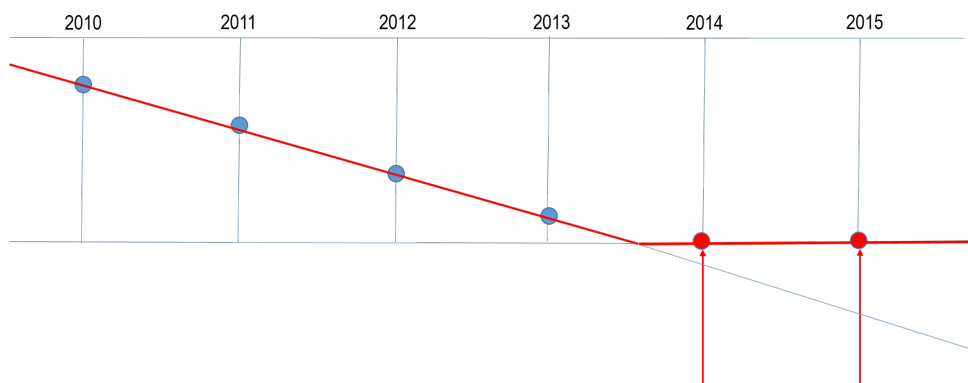
6:	21116	Meat of goat	6	35	kg
7:	21117.01	Meat of camels (fresh)	120	430	kg
8:	21117.02	Meat of other domestic camelids (fresh)	25	70	kg
9:	21118.01	Horse meat (fresh)	80	420	kg
10:	21118.02	Meat of asses (fresh)	60	150	kg
11:	21118.03	Meat of mules (fresh)	100	200	kg
12:	21119.01	Meat of other domestic rodents (fresh)	250	300	grams
13:	21121	Meat of chickens	550	3000	grams
14:	21122	Meat of ducks	700	3800	grams
15:	21123	Meat of geese	1400	8200	grams
16:	21124	Meat of turkeys	1200	18200	grams
17:	21170.01	Meat of pigeons and other birds n.e.	100	50000	grams

If some *Carcass weights* lie outside this ranges, the unfeasible values are replaced by the minimum or the maximum admitted values". In particular:

- if the resulting *Carcass weight* is greater that maximum admitted value it is replaced by the upper boundary of the range;

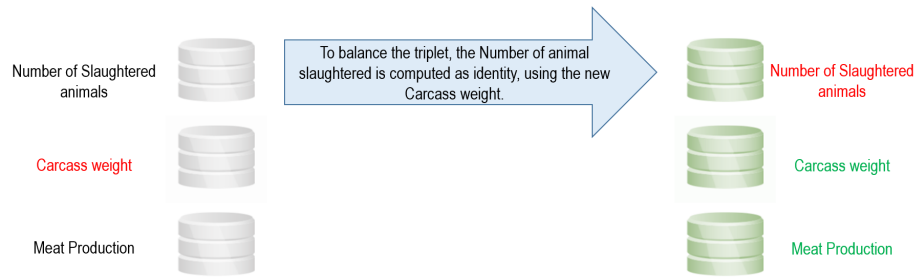


- if the resulting *Carcass weight* is lower that minimum admitted value it is replaced by the lower boundary of the range.

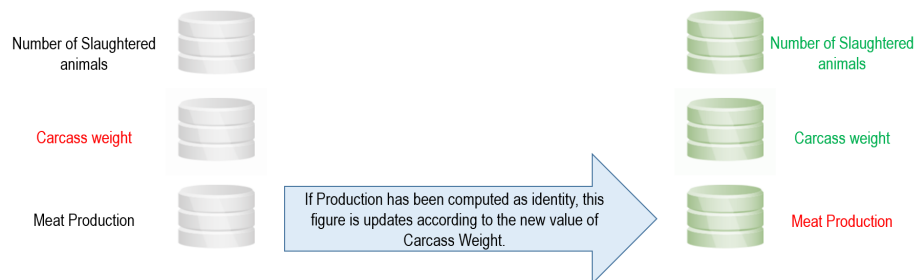


We do not have to forget that the *Carcass weight* is part of a triplet of elements that must always be balanced. To ensure that the *Carcass weight* is within the feasible ranges, we have to balance again the triplet.

In general we decided to use the *Number of animal slaughtered* to re-balance the triplet. This implies that some official/semi-official figures may be overwritten by the module. The user will receive an email all the times this case occurs in order to be able to check those unfeasible figures.



The only scenario where the *Production* element is overwritten is when *Production* had been computed as *identity*, so it was flagged with a "method flag" equal to "i". This choice depends on the exigency of reducing as much as possible the cases where *protected figures* are overwritten.



In this latter scenario there is no risk to overwrite *Official/semi-official* figures, since only *Production* that had been flagged with a "method flag" equal to "i" (and not protected by definition) will be overwritten.

### 3.1. Non-meat item

The so-defined *non-meat* item involves all those secondary items obtained from the slaughtering process. Offals, hides and skins are included in this category. The imputation process is based on the non-meat item triplet and the methodological framework is based once again on the ensemble approach. The number of *Slaughtered animal* is synchronized from the animal-parent item and it is exactly the same associated to the *meat* items. The key difference from the meat and non-meat items is that we do not dispose of a broad set of official/semi-official production data. That's why we decided to *protect* also imputations coming from the *old system*. This means that all the figures flagged as (I, -) are included and protected in the imputation process and are used as basis to produce new imputations. The difference between *protected* and *non-protected* figures is explicitly taken into account in the next section.

## 4. Flag management

### 4.1. Protected flag combinations

As mentioned the imputation method developed to fill the time series that contain missing values, needs a broad set of *official/semi-official* figures. The module exploits as much as possible the precious information already stored in the SWS. This means that the more accurate is the data in SWS the more reliable are the imputed figures obtained through the launch of the module.

It is extremely important to define a rule that identifies which figures have to be used as basis to produce imputations and which figures have to be overwritten by new imputations.

By now we have generically talked about *Official and semi-Official* figure as the backbone information to compute imputations, but for some series we do not dispose of a sufficient set of *official/semi-official* data and so we decided to *protect* (and not overwrite) a larger set of information. In particular, we look at some flag combinations that identify a more reliable set of data.

	flagObservationStatus	flagMethod	Valid	Protected
1:	E	c	TRUE	TRUE
2:	E	f	TRUE	TRUE
3:	E	h	TRUE	TRUE
4:	I	c	TRUE	TRUE
5:	M	-	TRUE	TRUE
6:	T	-	TRUE	TRUE
7:	T	c	TRUE	TRUE
8:	T	h	TRUE	TRUE
9:	T	p	TRUE	TRUE
10:		-	TRUE	TRUE
11:		c	TRUE	TRUE
12:		h	TRUE	TRUE
13:		p	TRUE	TRUE
14:		q	TRUE	TRUE

At each run of the *Livestock Production module* all the figure considered *non-protected* are deleted, and only the *protected* figures are used to produce new imputations.

## 4.2. Open/close a series

The list of *protected flag combinations* includes the (M,-) flag combination. The "M" observation flag indicates a *missing observation*. This means that, despite all those missing value flagged as (M,-) actually contains missing data, they won't be overwritten. In other words those figures have been flagged as missing on purpose, and do not have to be imputed.

It is the case, for example, of items that were produced in the past but are not produced anymore. The series must be *closed* in a way that can be properly interpreted by the R script. The agreed signal to identify this kind of situation is this "special" flag combination.

The situation is more complicated when we take into account the *triplet*. We have to ensure the compliancy between the element of the triplet. If one element of the triplet is flagged as (M,-) all the others are automatically flagged as (M,-) unless they are already flagged as *protected figures*, and *protected figures* are not overwritten.

Once a series has been *closed* it has to be manually re-opened. The only way to *re-open* a closed time-series is to add new protected figures for at least two elements of its triplet in the year where that series starts again. In particular the system requests to add:

- a protected figure for the "Output";
- a protected figure for the "Input".

## 5. Items involved

The current version of the *Livestock* sub-module is working on 17 animal items:

	ItemParentCode	ItemParent
1:	02111	Cattle
2:	02112	Buffaloes
3:	02122	Sheep
4:	02123	Goats
5:	02140	Pigs
6:	02151	Chickens
7:	02154	Ducks
8:	02153	Geese
9:	02152	Turkeys
10:	02131	Horses
11:	02132	Asses
12:	02133	Mules
13:	02121.01	Camels
14:	02191	Rabbits
15:	02192.01	Other Rodents
16:	02121.02	Other Camelids
17:	02194	Other Birds

To each *animal item*, several *meat* and *non-meat* items are associated. For example, we report here the derived items obtained from the pig:

	ItemParentCode	ItemParent	ItemChildCode	ItemChild
1:	02140	Pigs	21113.01	Pigmeat
2:	02140	Pigs	21153	Offals of Pigs, Edible
3:	02140	Pigs	21511.01	Fat of Pigs
4:	02140	Pigs	02959.01	Pigskins, Fresh

	ItemParentCode	ItemParent	ItemChildCode	ItemChild
1:	02111	Cattle	21111.01	Beef and Veal
2:	02111	Cattle	21151	Offals of Cattle, Edible
3:	02111	Cattle	21512	Fat of Cattle
4:	02111	Cattle	02951.01	Cattle Hides, Fresh
5:	02112	Buffaloes	21112	Buffalo Meat
6:	02112	Buffaloes	21152	Offals of Buffalo, Edible
7:	02112	Buffaloes	21513	Fat of Buffalo
8:	02112	Buffaloes	02951.03	Buffalo Hides, Fresh
9:	02122	Sheep	21115	Mutton and Lamb
10:	02122	Sheep	21155	Offals of Sheep, Edible
11:	02122	Sheep	21514	Fat of Sheep
12:	02122	Sheep	02953	Sheepskins, Fresh
13:	02123	Goats	21116	Goat Meat
14:	02123	Goats	21156	Offals of Goats, Edible
15:	02123	Goats	21515	Fat of Goats
16:	02123	Goats	02954	Goatskins, Fresh
17:	02140	Pigs	21113.01	Pigmeat
18:	02140	Pigs	21153	Offals of Pigs, Edible
19:	02140	Pigs	21511.01	Fat of Pigs
20:	02140	Pigs	02959.01	Pigskins, Fresh
21:	02151	Chickens	21121	Chicken Meat
22:	02151	Chickens	21160.01	Offals Liver of Chickens
23:	02154	Ducks	21122	Duck Meat



The relationship between the parent (animal) and child commodities (items obtaining from the slaughtering process) is stored in a data table in the SWS:

This table can be manually modified in case some CPC codes change, or if it is necessary to expand the list of animals and/or derived items to be taken into account.

## 6. Elements involved

There is a triplet associated to each item whatever is its nature: *Animal*, *meat*, *non-meat*... The elements composing each triplet is contained in another data table stored in the SWS:

itemtype	description	areaharvested	yield	production	factor
1: DERA	Derived animal products	5327	5423	5510	1000
2: HIDE	Hides and skins	5320	5417	5510	1000
3: LSNP	Livestock non-primary	5320	5417	5510	1000
4: LSNP	Livestock non-primary	53200	54170	55100	1000
5: LSNP	Livestock non-primary	53201	54171	55101	1000
6: LSPR	Livestock primary	5031	5417	5518	1000
7: OFF	Offals	5320	5424	5510	1000
8: POFF	Poultry offals	5321	5424	5510	1000
9: PONP	Poultry non-primary	5321	5424	5510	1000000
10: PONP	Poultry non-primary	53210	54240	55100	1000000
11: PONP	Poultry non-primary	53211	54241	55101	1000000
12: POPR	Poultry primary	5032	NA	5513	NA
13: PSKN	Poultry skins	5321	5424	5510	1000000

	itemtype	description	areaharvested	yield	production	factor
1:	DERA	Derived animal products	5327	5423	5510	1000
2:	HIDE	Hides and skins	5320	5417	5510	1000
3:	LSNP	Livestock non-primary	5320	5417	5510	1000
4:	LSNP	Livestock non-primary	53200	54170	55100	1000
5:	LSNP	Livestock non-primary	53201	54171	55101	1000
6:	LSPR	Livestock primary	5031	5417	5518	1000
7:	OFF	Offals	5320	5424	5510	1000
8:	POFF	Poultry offals	5321	5424	5510	1000
9:	PONP	Poultry non-primary	5321	5424	5510	1000000
10:	PONP	Poultry non-primary	53210	54240	55100	1000000
11:	PONP	Poultry non-primary	53211	54241	55101	1000000
12:	POPR	Poultry primary	5032	NA	5513	NA
13:	PSKN	Poultry skins	5321	5424	5510	1000000

The column *itemType* identifies a typology of item. All the items belonging to the same category shares the same *Elements* and consequently the same *units of measurement*. The *factor*-column contains the conversion factors associated to each triplet. For example:

- Cows are categorized as *big animals*. The triplet associated to *meat of cattle* is

$$Production_t^{tons} = SlaughteredAnimals_t^{heads} * CarcassWeight_t^{kg} * \frac{1}{1000}$$

- Chicken are categorized as *small animals*. The triplet associated to *meat of Chicken*

$$Production_t^{tons} = SlaughteredAnimals_t^{1000heads} * CarcassWeight_t^{grams} * \frac{1}{1000}$$

## 7. Running the plugin in the SWS

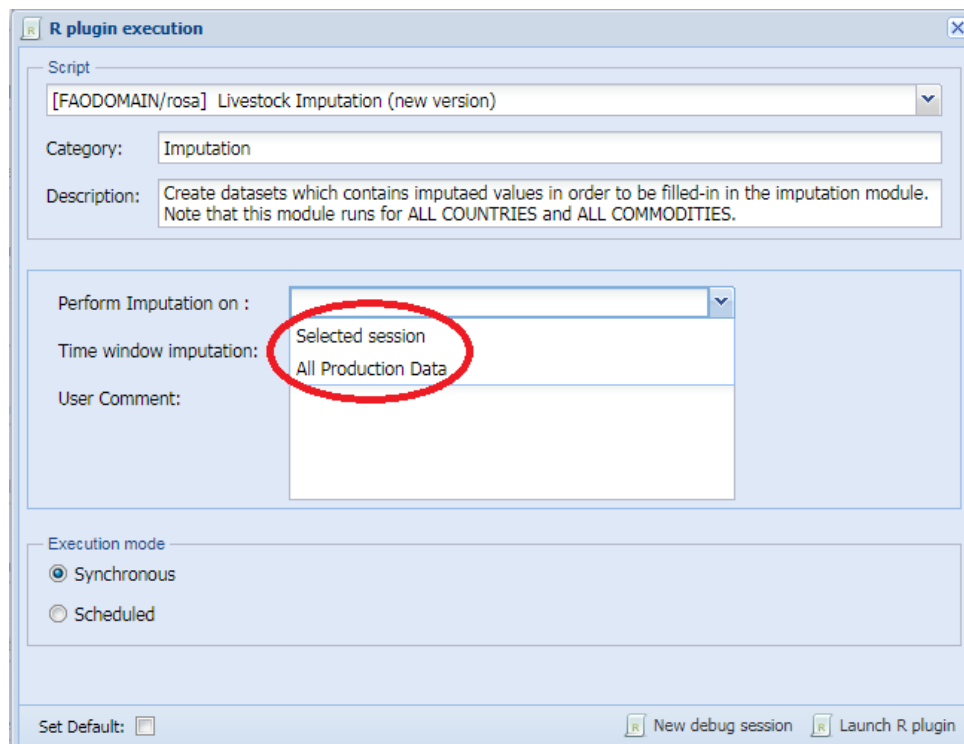
This section contains the list of operations to be performed to successfully run the *Livestock Imputation Module*.

1. Open a new session in the SWS on:

- domain = *Agriculture production*
- dataset = *Agriculture production*

The selection is not particularly important, provided that you select at least one meat item which is, in general, the one (or the those that) you want to impute. Country and time-window are not important at all: all the countries will be automatically selected by the module and the time-window starts from 1990 up to the last available year<sup>3</sup>. The complete *imputation key* is contained in the data table stored in the SWS: "fbs production comm codes". This data table has to be updated at least every year to include the last year for which you want to produce imputations.

2. Click on R plugin and choose: *Livestock Imputation (New version)*
3. Set up the imputation parameters. The imputation process may be performed on the "Selected session" only on the meat (and all its linked items) contained in the session selection. On the other hands the imputation process may be performed on "All Production Data", this means that the original selection will be expanded and the imputation process will be performed for all the 17 meat items and all the linked animal and non-meat items.



4. Set up the imputation parameters. The user has the possibility to choose the *Time window imputation* that can be *since 1990* or *last three years*. In this latter case, the module automatically select the last year contained in the complete imputation key and will produce imputations only for last three years. All previous figures are considered protected and will be used to produce new imputations (see the section about flag management for further clarifications).

<sup>3</sup>The time window is stored in the datatable containing the *complete imputation key*.

5. Click on *Launch R plugin*

## 8. Milk and eggs items

Milk and eggs represent an important foodstuff in terms of nutritive factors. In addition, expecially milk products have a very complex commodity tree which implies that these items contributes, mixed with other commodities, to the production of many derived items.

In terms of data imputation, both *milk* and *eggs* items request the same methodological approach and are characterized by the same structure of elements composing the triplet.

### 8.1. Eggs

The key aspect charachterizing the egg-imputation method consists in the relationship between the *livestock numbers* and the number of *Laying hens* (or *other birds*).

In particular we report the table containing the correspondece in terms of CPC codes between the animal and the egg items.

	Animal code	Animal	Egg code	Egg
1:	02151	Chickens	0231	Hen eggs in shell, fresh
2:	02194	Other birds	0232	Eggs from other birds in shell, fresh, n.e.

While the match between the *Chicken* and *Hen eggs* is more staightfull, the relation between the generic animal item "*Other birds*" and the corresponding typology of *eggs* is not complitely esaustive. In this generic cathegory we may include *eggs of duck - geese - quails ...* The *animal item* is so generic that it is not possible to estimate the number of *laying animals* starting from a specif series of livestock animals. We do not have any other option than using the series *Other birds* <sup>4</sup>.

<sup>4</sup>This choice may not be appropriate for all the countries. It would be extremely helpfull identifying for each country which is the livestock series to be associated to egg production in order to be more precise in imputing laiying birds referring to the item *Eggs from other birds in shell, fresh*.

The key elements involved in this process are:

- *Laying [1000 head]*: in theory under this element should be entered the same data recorded in the element *female in reproducing age* of the chicken (in poultry account) because every female in reproducing age is a potential layer. However it has been decided to register under this element only the females kept primarily for egg production. Data are shown in thousand unit.
- *Eggs production [tons]* data refer to the quantity of eggs, in weight terms, produced during the year by all layers, whether in the traditional or in modern sectors. It includes hatching eggs and eggs wasted at the farm.
- *Yield [g/head]* refers to the average number, in weight terms, of eggs laid by a layer during the year. In this case data are generally not recorded but obtained by dividing the quantity of eggs produced by the number of laying hens/birds. In general we do not dispose of a wide set of official/semi-official figures. We can only use as benchmark the whole time-series already computed in the past. *Yield* can be expressed also in terms of *Number of eggs in shell*. This quantity is obtained dividing the *yield in weight* by the average weight of one egg.

## 8.2. The model

To capture the dependency between the *Laying hens/other birds* and the *Livestock numbers* we implemented a Hierarchical Linear Model. This choice depends on the need to have at least one explanatory variable instead of building a model where the time-series is imputed looking only at its own past.

$$\log(\text{LayingHen}_{i,j,k}) = \beta_0 + \beta_1 \text{Time} + \beta_{2,j,k}(\log(\text{LivestockNumbers}))_i + \epsilon_{i,j,k}$$

$$\beta_{2,j,k} = \gamma_{2,0} + \gamma_{2,1,j,k}(\text{CPC} : \text{Country})_{j,k} + \delta_{j,k}$$

$$\gamma_{2,1,j} = k_{3,1,0} + \gamma_{3,1,1}(\text{CPC})_j + \phi_j$$

where  $\beta, \gamma, k$  are coefficients to be estimated from the data,  $\epsilon, \delta, \phi$  are error estimates, *LivestockNumbers* is the average estimates number of poultry livestock (or other birds livestock) of country  $i$ , and Time is measured in years and is included to capture linear trends over time. The  $i$  indices run over all countries, the  $j$  indices over all CPC item, and the  $k$  indices over all unique country CPC-code combinations.

Thus, the model estimates *Number of laying animals* proportional to the *Livestock Numbers*. The model also accounts for changes over time and differences among countries; the latter are captured by *Livestock numbers* in the country.

If data for a particular country and commodity are sparse, then  $k_{3,1,0}$  and  $k_{3,1,1}$  will likely be estimated as close to 0. Thus,  $\gamma_{2,1,j,k}$  will be close to its mean value, and the model will account only for availability within commodity groups. However, if data are available for a country or commodity, the estimates of  $k_{3,1,0}$  and or  $k_{3,1,1}$  will be far from 0, and thus the model enables adaptation to the individual characteristics of a particular country.

We used the same methodological approach also to formalize the dependency between the *Production of eggs* and *number of laying birds*.

$$\log(\text{EggProd}_{i,j,k}) = \beta_0 + \beta_1 \text{Time} + \beta_{2,j,k}(\log(\text{LayingHen}))_i + \epsilon_{i,j,k}$$

$$\beta_{2,j,k} = \gamma_{2,0} + \gamma_{2,1,j,k}(\text{CPC} : \text{Country})_{j,k} + \delta_{j,k}$$

$$\gamma_{2,1,j} = k_{3,1,0} + \gamma_{3,1,1}(\text{CPC})_j + \phi_j$$

Using the *Hierarchical linear model* as methodological basis to produce imputations for *egg production*, it is sufficient to dispose of a series of *livestock* for the animal item to impute the derived items. This means that, even if the production time-series did not exist at all before the launch of the module, it is possible to obtain imputations, from scratch, for the *egg production* series.

At the moment the module has been built in order to produce imputations only for those time-series for which at least one figure was already available before the imputation process, but in theory we should be able to impute *egg production* also in those countries where no data on *egg production* is available, under the hypothesis that if there are *chickens* or *Other birds* there is no reason why the egg production time series should be empty.

### 8.3. Milk

The key aspect characterizing the milk-imputation method consists in the relationship between the *livestock numbers* and the number of *Milking animals*.

In particular we report the table containing the correspondence in terms of CPC codes between the animal and the milk items.

	Animal code	Animal	milk code	milk
1:	02112	Buffalo	02212	Raw milk of buffalo
2:	02122	Sheep	02291	Raw milk of sheep
3:	02123	Goats	02292	Raw milk of goats
4:	02121.01	Camels	02293	Raw milk of camel
5:	02111	Cattle	02211	Raw milk of cattle

The key elements involved in this process are:

- *Milking animals [heads]* that gives the number of animal which, in the course of the reference period have been milked. It is important to highlight that by *milking animals* it is not meant *animal producing milk*, but *animals milked*. This concept is in relation to the concept of milk production which excludes the milk used for feeding young animals. Note that the number of *milking animals* may be higher than the number of *dairy females*, but it cannot be higher than the number of *female actually reproducing* in the livestock herd account.
- *Milk production [tons]* indicates the quantity of milk milked during the year from the milking animals of specific species.
- *Yield [kg/head]* represents the average quantity of milk produced by a milking animal during the year. It is generally not recorded but obtained by dividing the data stored under the production element by those recorded under milking animal element.

#### 8.4. The model

To capture the dependency between the *Milk Production* and the *Milking animals* we implemented a Hierarchical Linear Model. The reason why we have chosen this model is exactly the same already mentioned for *eggs*: this choice depends on the need to have at least one explanatory variable instead of building a model where the time-series is imputed looking only at the its own past.

$$\log(\text{milkingAnimals}_{i,j,k}) = \beta_0 + \beta_1 \text{Time} + \beta_{2,j,k}(\log(\text{Livestock}))_i + \epsilon_{i,j,k}$$

$$\begin{aligned} \beta_{2,j,k} = & \gamma_{2,0} + \gamma_{2,1,j,k}(\text{CPC} \\ & : \text{Country})_{j,k} + \delta_{j,k} \end{aligned}$$

$$\gamma_{2,1,j} = k_{3,1,0} + \gamma_{3,1,1}(\text{CPC})_j + \phi_j$$

where  $\beta, \gamma, k$  are coefficients to be estimated from the data,  $\epsilon, \delta, \phi$  are error estimates, *Livestock* is the average estimates number of animals of country  $i$ , and Time is measured in years and is included to capture linear trends over time. The  $i$  indices run over all countries, the  $j$  indices over all CPC item, and the  $k$  indices over all unique country CPC-code combinations.

Thus, the model estimates *milking animals* proportional to the *livestock numbers*. The model also accounts for changes over time and differences among countries; the latter are captured by *livestock numbers* in the country.

If data for a particular country and commodity are sparse, then  $k_{3,1,0}$  and  $k_{3,1,1}$  will likely be estimated as close to 0. Thus,  $\gamma_{2,1,j,k}$  will be close to its mean value, and the model will account only for availability within commodity groups. However, if data are available for a country or commodity, the estimates of  $k_{3,1,0}$  and or  $k_{3,1,1}$  will be far from 0, and thus the model enables adaptation to the individual characteristics of a particular country.

As already explained for the *eggproduction*, we used the same methodological approach also to formalize the dependency between the *Production of milk* and *number milking animal*. Details about the methodology are available in the previous section where the methodology to compute *egg production* is presented.

#### 8.5. How to run the milk and eggs plugins in the SWS?

Both milk and egg production depend on *livestock numbers*, this dependency implies that those two sub-modules have to be launched **after** the *livestock imputation sub-module* has been run and validated.

The only parameter to be set is the imputation *time window*: in particular the user can choose if producing new imputations (and overwrite non-protected figures) in the whole time-window or to produce new imputations only in the last three years.

## **9. APPENDIX 1: List of all the data-table/dataset involved in the module**

**9.1. Aupus share**

**9.2. Animal Parent-child mapping**

**9.3. Item Type Yield**

**9.4. range\_carcass\_weight**

**9.5. animal\_eggs\_corrispondence**

**9.6. animal\_milk\_corrispondence**

### **Affiliation:**

Firstname Lastname

Affiliation

Address, Country

E-mail: **name@address**

URL: **http://link/to/webpage/**