

Missing data treatment in administrative fiscal sources in the French structural business statistics production system

Thomas Deroyon - Insee

9 septembre 2013

French Structural Business Statistics main principles 1/2

ESANE : Structural business statistics production system in France since 2008

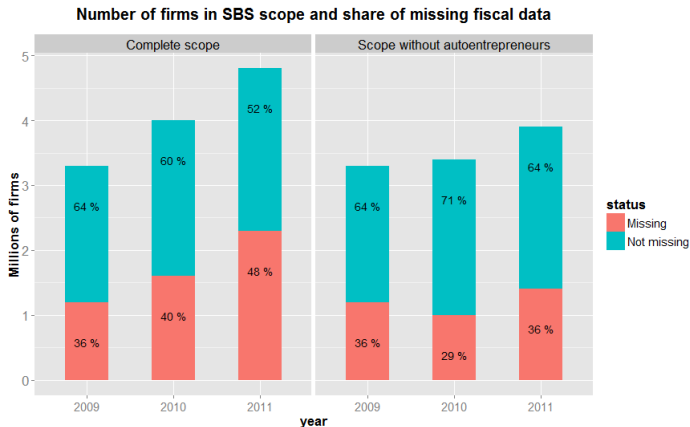
Main principles :

- 1 **Scope defined ex ante** : on to the basis of Insee's statistical firm register
- 2 **Extensive use of exhaustive administrative data** : mainly firms compulsory income returns

French Structural Business Statistic' main principles 2/2

- 3 **Surveys on samples of firms to obtain other information** : especially detailed turnover breakdown and revaluated principal activity
- 4 **Reconciliation of both sources** : on information available in both fiscal files and survey (turnover, its aggregate breakdown and employment and wages)
- 5 **Composite estimators** : to optimally mix the two sources of information, for instance difference estimators of sectorial fiscal aggregates

The Problem of missing fiscal data



Plan of the presentation

Strategie to deal with fiscal missing data

- 1 Characterizing the origins of missing data :**
 - What are the different causes of fiscal missing data ?
 - How each cause can be statistically typified and dealt with : nonresponse / out of scope ?
- 2 Identifying to which cause each missing data belongs**
- 3 Treating each missing data**
- 4 What effect of the treatments on SBS aggregates ?**

Three main causes...

- 1 **Exemptions** : very small (individual) enterprises (named in French *microentreprises* and *autoentrepreneurs*) actually report to fiscal administration on their activity in their household income tax declarations
- 2 **Register overcoverage** : firms have to register when they start, but not when they stop \Rightarrow some missing firms are actually defunct
- 3 **No compliance to tax return obligations** : some firms should send annual income returns to tax authorities but do not

...Corresponding to two different situations

- 1 **Out of scope units** : firms in case 2 should not have been included in Structural business statistics scope
⇒ EXCLUSION FROM SBS'SCOPE
- 2 **Nonrespondent units** : cases 1 and 2 correspond to non response in a census
⇒ NON RESPONSE TREATMENT

- ① **Exemption** : we have a list of firms in case 1
- ② **Other cases** (Out of scope / No compliance) : little external information ⇒ Use of past information
 - VAT declaration during the year
⇒ NO COMPLIANCE
 - Income tax return in one of the two preceding years
⇒ NO COMPLIANCE
 - Other cases
⇒ OUT OF SCOPE

- ① **Out of scope** : units are discarded from the files and all following treatments

- ② **Non response** : Imputations
 - COLD DECK : for units in exemption case \Rightarrow typical tax returns built with aggregate information imputed to each firm

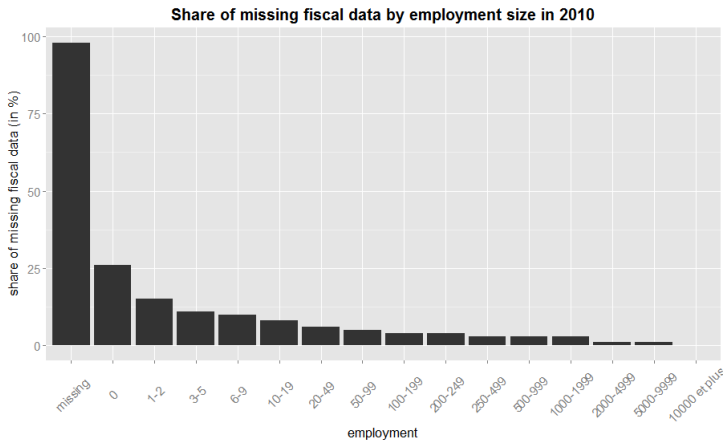
 - MEAN IMPUTATION : for other units (creations / firms entering Esane's scope) \Rightarrow mean imputation in crossings of principal activity (3 digits) and employment size (15 groups)

 - ...

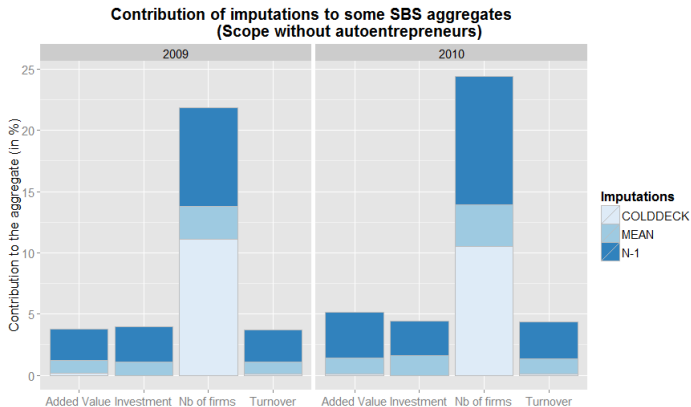
② Non response : Imputations

- N-1 IMPUTATION : for units who were already in Esane's scope the preceeding year
 - ⇒ use of their former declaration
 - ⇒ use of median turnover growth rate in crossings of principal activity (3 digits) and employment size (5 groups)
 - ⇒ imputed fiscal variables are calculated by assuming each fiscal variable had the same growth rate equal to the median turnover growth rate of the crossing

Imputations concern mainly small firms

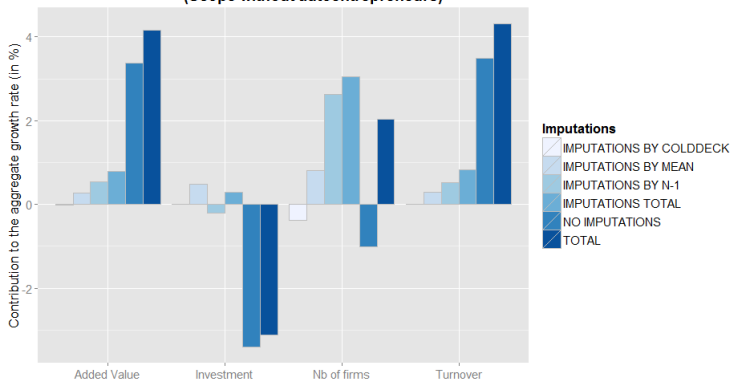


A limited effect on aggregate levels ...



... and on aggregate growth rates

**Contribution of imputations to some SBS aggregates growth rate
between 2009 and 2010
(Scope without autoentrepreneurs)**



Further methodological developments :

- Tests to use household income returns to build better colddeck imputations
- Tests of other methods than mean imputation to treat creations / firms entering Esane's scope
- Studies to understand more deeply and precisely imputations effect on detailed aggregate levels and growth rates