
Accuracy, precision, and bias in HFD data

Laurent Toulemon


Period and cohort fertility dynamics in the developed world.

The first Human Fertility Database Symposium

3-4 November 2011

Max Planck Institute for Demographic Research, Rostock, Germany

Session V: Methods for producing more detailed fertility data
and data quality questions (Chair: Dmitri Jdanov, MPIDR)



Accuracy, precision, and bias in HFD data



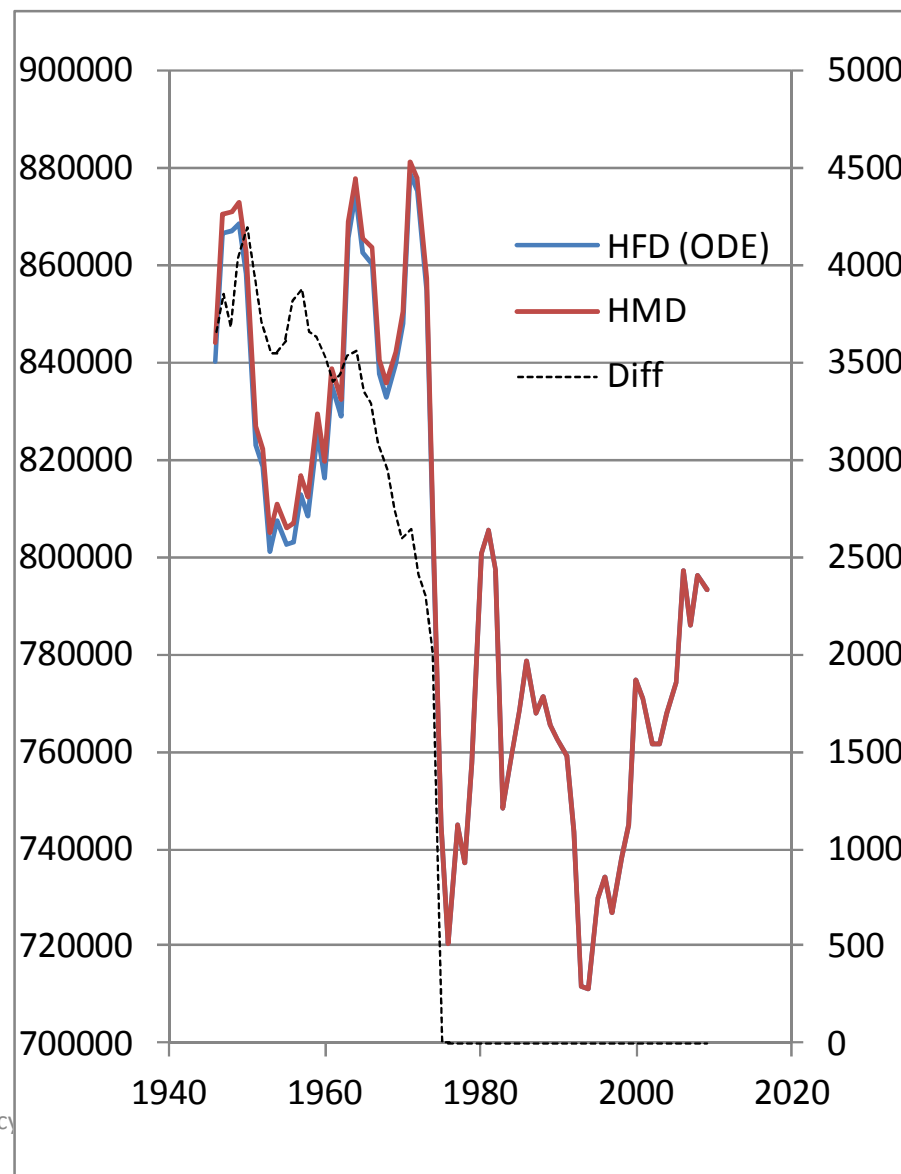
- Accuracy of input data
 - Births total and population size
 - Age, birth order, parity
- Precision
 - Lexis triangles
 - Population by parity
- Bias in HFD
 - Undercount in census
- Conclusion

Accuracy of input data

- Births total
 - Country definition
 - Mainland France; there are 3.9% more births and 3.0% additional population in official French statistics
 - Birth definition
 - Stillbirths are not to be included
 - But “false stillbirths” (a birth followed by a death before registration) should be included
 - 3,500 births per year, 1946-74; 0,4% births missing in France
 - Data are available at Insee

Accuracy of input data

- Births total
 - Yearly number of births in France, including false stillbirths (dotted curve, right scale) in the HMD
- Somebody at Insee (Fabienne Daguet) has collected these data
 - False stillbirths by year and age (and cohort)
- The HFD project is a unique opportunity to publish them
- Other countries?



Accuracy of input data

- Population size
 - Census data, population estimates
 - Interpolation between censuses
 - OK to use HMD population data?
- Age, birth order, parity
 - Missing ages at delivery are imputed (proportionally)
 - Birth order and women's parity are also imputed proportionally (for each age)

Accuracy of input data

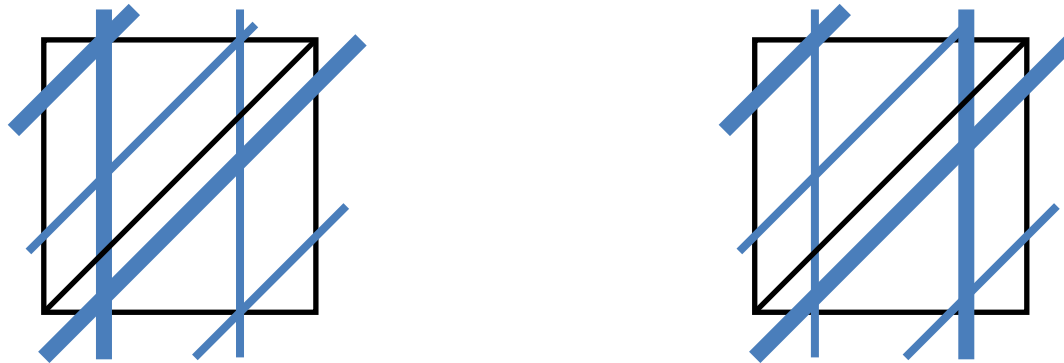
- Birth order
 - How many births have missing birth order?
 - Is a proportional imputation accurate?
Births without order may be more than proportionally out-of-wedlock births, thus more likely of low order
 - Overall quality of birth order data?
 - Too bad in France to be considered. What about other countries?
- Parity
 - Question on children ever born or number of children living in the household?
 - Similar question than for birth order: who are the women for whom no information on parity is not available and has been imputed?

Precision of HFD data

- Lexis triangles

→ A change in seasonality introduces a bias in the share between Lower and Upper triangles

→ On the left, no change; on the right, a change



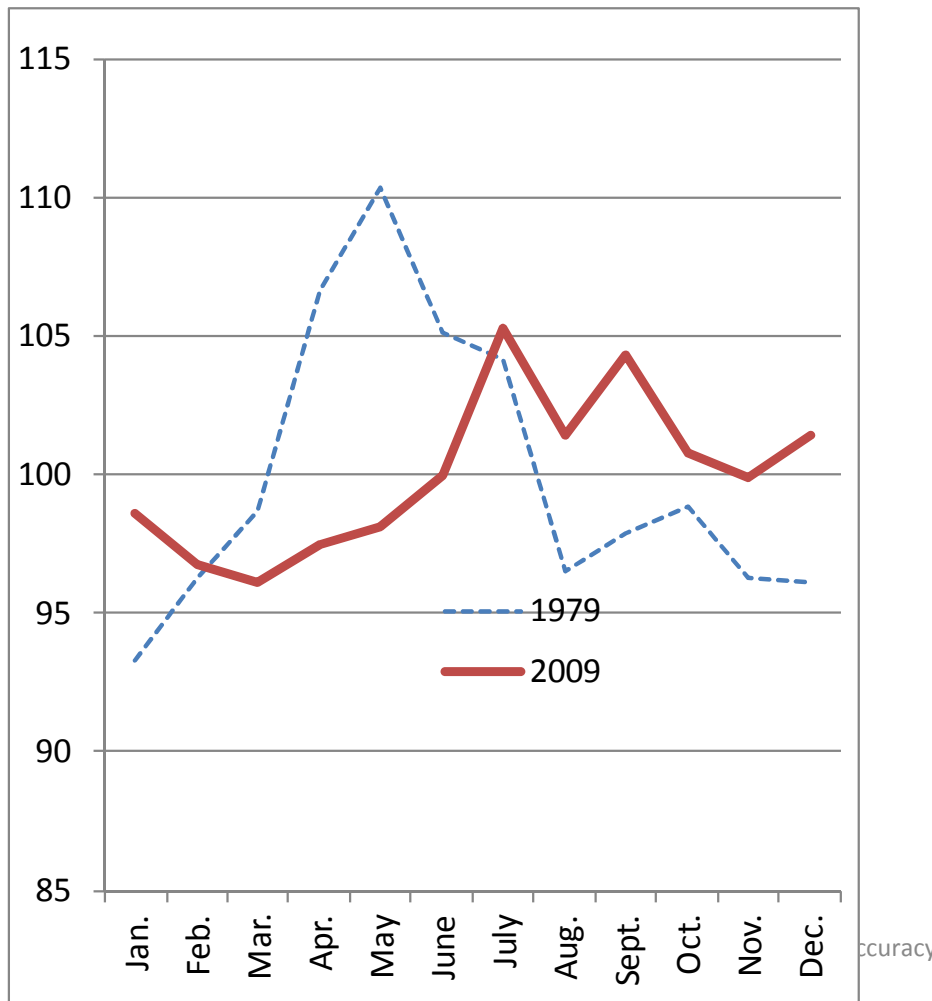
→ On the right, more births in Lower triangles

→ See next slide, current situation in France

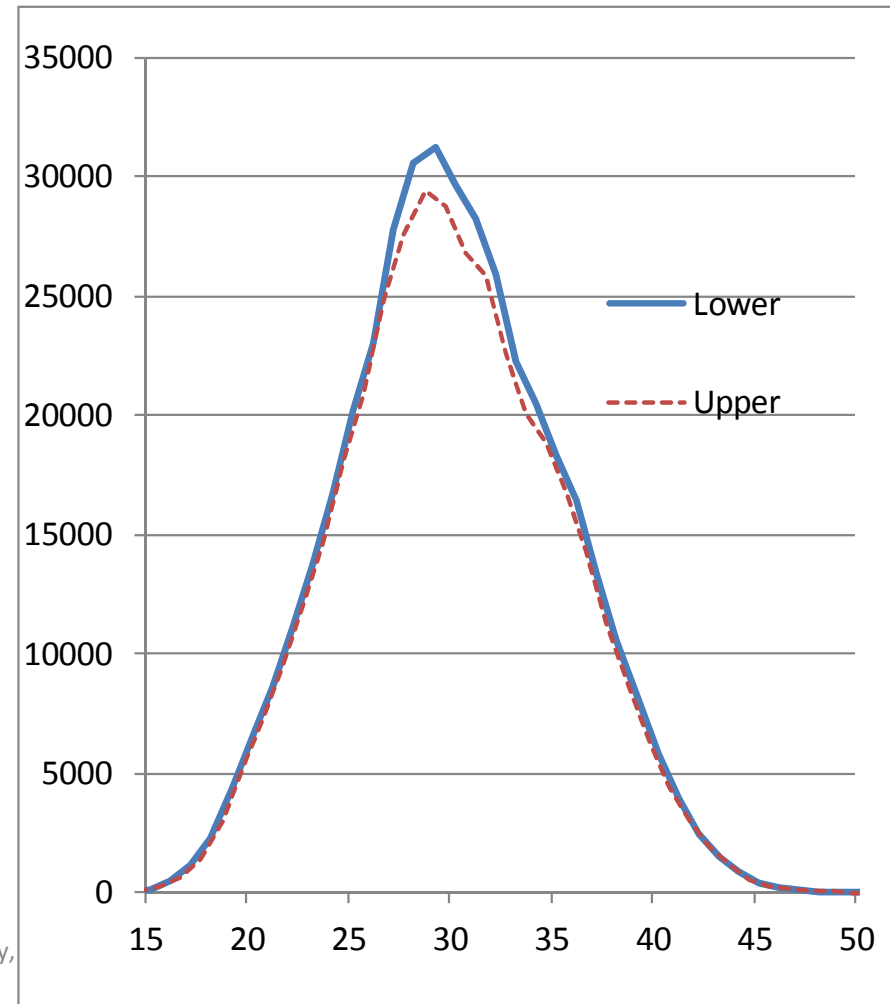
Lexis triangles: births in France, 2009



Seasonality of births (mean=100), France, 1979 and 2009



Age-specific number of births, upper and lower triangles



Lexis triangles

- The exposures are estimated with taking seasonality into account
- It would thus be more consistent to do the same for the births (eq. 4.8 and 4.9)
- The difference may be negligible if the original data, as well as the output results, are “age-period” based
- But not if they are “period-cohort” based (age reached during the year)

Precision of HFD data

- Birth order: categories 1-5+
 - Important to estimate distribution by number of births, including categories 3, 4+
 - Higher order births and higher parities?
- Population distribution by parity
 - The choice has been made to be consistent with cumulated rates by cohort
 - Some bias may be introduced if migrations vary with parity (e.g. childless women emigrate)
 - Can more accurate data be obtained from censuses?
 - Use several “golden censuses” and interpolate?

Bias in Human Fertility Database



- Undercounting of births
 - Very unlikely and minor (false stillbirths)
- Undercounting of women in censuses
 - The HMD census-based data are accurate, updated and documented
 - But the censuses often underestimate the population size (undercount)
 - This of minor importance for mortality, compared to fertility (next slide)

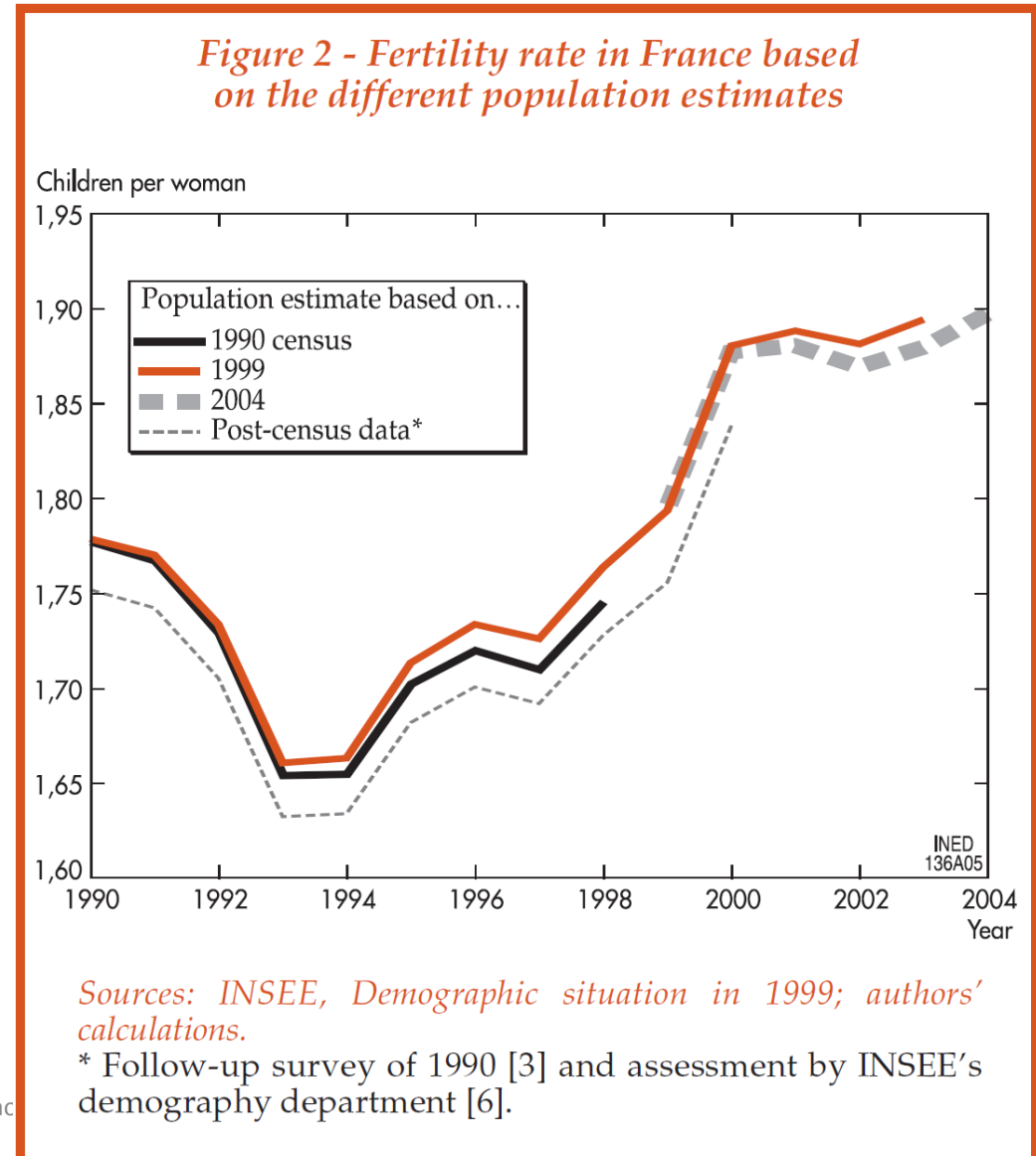
Bias in Human Fertility Database



- Assume a 2% population undercount at census, similar in magnitude for all sex and age groups
 - The **mortality** rates are overestimated by 2%
 - Life expectancy at birth is downward biased of 0,25 year
 - as mortality is increasing by some 8% by year of age
 - A 0.3% relative bias in e_0 ,
 - comparable to the annual increase in life expectancy
 - The **fertility** rates are overestimated by 2%
 - The TFR is upward biased by 0.04 children per woman
 - A 2.0% relative bias in TFR
 - twice the annual mean increase since 15 years in France

Bias in Human Fertility Database

- An estimate based on the French 1990 post-census enumeration survey
 - The bias could be larger because undercounts are larger for adults
 - On the contrary, undercounting may be less severe for women in fertile ages
- Should the census data be corrected?



Conclusion (1/3)

- Many methodological points have been fixed in the new Methods protocol
 - An impressive success!
- Strengths of the Human Mortality Database
 - Availability, quality, documentation of data
 - Population estimates
- A good incentive to make hidden information explicit and available
 - French data on false stillbirths at Insee, probably similar “hidden” data in other countries
 - Not only dissemination, but data building process

Conclusion (2/3)

- Potential improvements on HFD
 - Accurate population data by parity
 - Quality of answers in the censuses
 - Take migrations into account
 - Higher parities to be considered
 - Assessment on census data quality
 - Alternative population estimates?
 - Triangles and seasonality of births
- Other data?
 - Strong added value, compared to survey data
 - Human fertility collection

Conclusion (3/3)

- The Human fertility collection (HFC)
 - Countries where data are not easily available or not accurate (e.g. birth order and parity in France)
 - Alternative data sources: large scale surveys
- Additional variables (for HFC)
 - Duration since previous birth
 - conjugal situation, marital status
 - professional activity
 - Level of education, “race”, social group
 - Urban/rural, sub-national geographical level data
 - Male fertility
 - ...
- How to proceed? (see next roundtable)

The end



Thank you

Laurent Toulemon,
toulemon@ined.fr