UNIVERSITY OF Southampton

Southampton Statistical Sciences Research Institute

Estimating the Uncertainty in the Local Authority Mid-year Population Estimates

Peter W. F. Smith

University of Southampton

CMIST, April 2016

Outline

- Mid-year population estimates (2001–10)
  - Cohort component method
  - Adjustments for special population subgroups
- Overall methodology
- The International Passenger Survey
- International migration methodology
- Internal migration methodology
  - Sources of errors
  - Assessment of errors
  - Modelling the errors
  - Simulating from error distribution
- Results
- Summary

Mid-year Population Estimates (MYE)

- Produced annually by the Office for National Statistics (ONS)
- For the usual resident population of England and Wales (E&W)
- Include:
  - long-term (at least 12 months) migrants
  - HM and US armed forces stationed in E&W
  - students at their term-time address
  - prisoners at institution if they have served six months or more
- As at mid-year (30 June)
- Broken down by age, sex and local authority (LA)
- Census 'population base' updated annually using various data sources
- Use a cohort component method

Cohort Component Method

$$\begin{aligned} P_t &= P_{t-1} + B_t - D_t + M_t \\ &= P_{t-1} + B_t - D_t + Im_t - Em_t + In_t - Ou_t \end{aligned}$$

where

$P_t =$ population at time t

$B_t =$ births in the interval from $t-1$ to t

$D_t =$ deaths ...

$M_t =$ net migration ...

$Im_t =$ immigrants (from outside the UK) ...

$Em_t =$ emigrants (to outside the UK) ...

$In_t =$ internal in-migrants (from within the UK) ...

$Ou_t =$ internal out-migrants (to within the UK) ...

Adjustments for Special Population Subgroups

- For prisoners and school boarders
- The armed forces and asylum seekers and their dependents
- Estimated separately since they are not covered by the data sources used for migration estimates

Overview of the Methodology

- Census estimates assumed to be normally distributed with published standard errors
- Births and deaths – civil registration, assumed to be measured without error
- International migration
  - estimated using the International Passenger Survey (IPS) and the Labour Force Survey (LFS)
  - error distributions estimated using a bootstrap
- Internal migration
  - estimated using the National Health Service Central Register (NHSCR) and the GP Patient Register Data System (PRDS)
  - error distributions estimated by modelling the error using the Census as a 'gold standard'
- Adjustments assumed to be measured without error
- Realisations from estimated error distributions combined using cohort component method

The International Passenger Survey (IPS)

- Voluntary sample survey of passengers arriving at, and departing from, the main UK airports, seaports and the Channel Tunnel
- Stratified to ensure that it is representative by mode of travel, route and time of day
- Identifies migrants and their towns of destination or residence prior to departure
- Approximately 250 000 responding passengers per annum
- In 2008, just over 5 000 migrants were interviewed
- Limitation – it asks about intentions

International Migration Methodology

- Migrant counts are weighted to produce national estimates of migration
- Regional and 'New Migration Geography inflow' level estimates of immigrant counts are calibrated using data from the LFS
- Regional and 'New Migration Geography outflow' level estimates of emigrant counts obtained directly from IPS
- LA estimates of migrant counts are obtained by apportioning higher level estimates down to local authority level
    - distributions estimated by Poisson regression models for the IPS-based estimates of LA migrant counts
- Error distributions estimated by bootstrapping the IPS and LFS, and repeating the above procedures 1000 times

Internal Migration Methodology

- Individual moves captured from GP re-registration data
- Annual (end July) download of patient registers
- Moves identified by changes from previous year's download
- LA moves constrained to information provided by the NHSCR

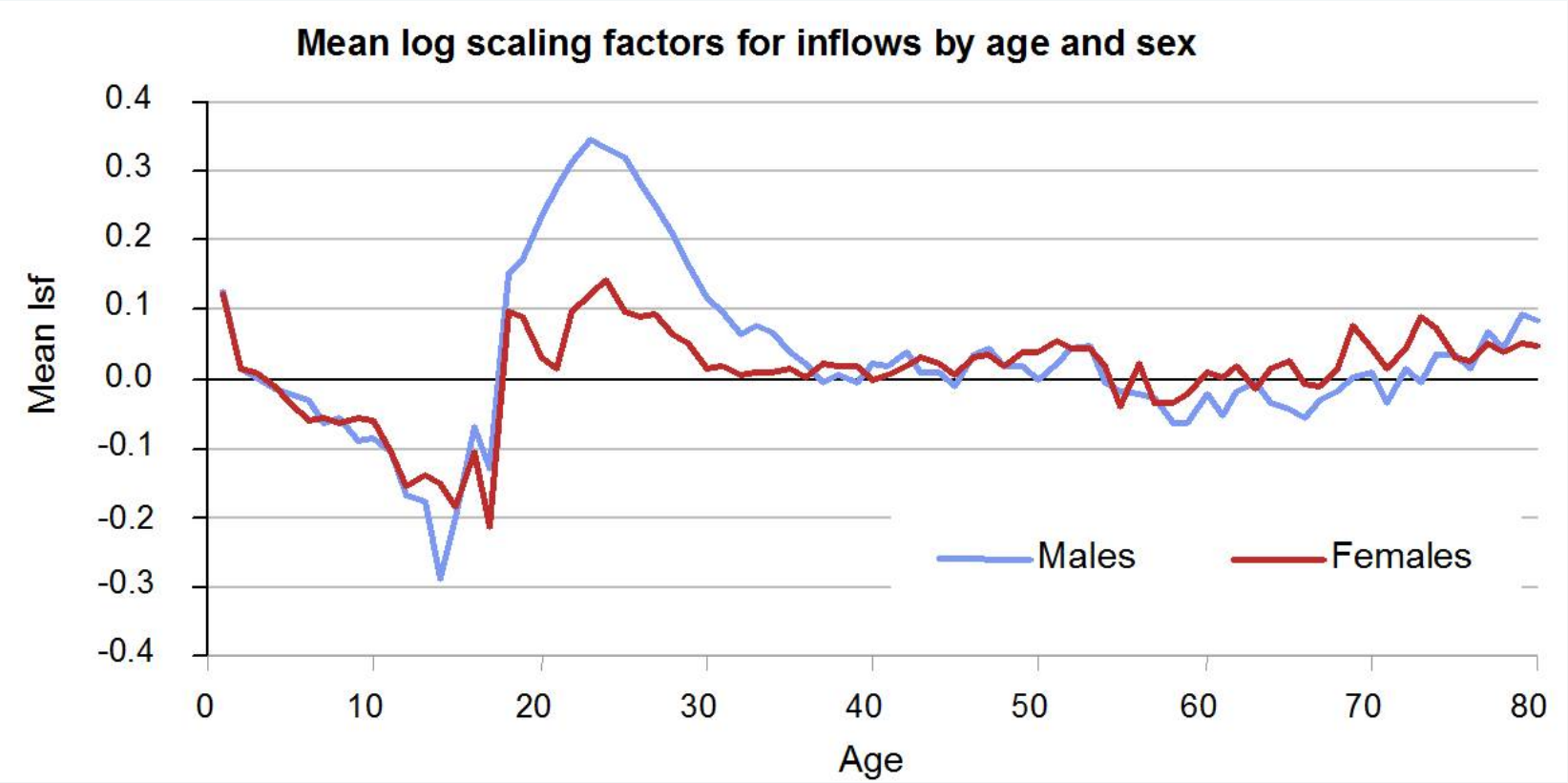
Sources of Errors

- Time lags between moving and re-registering
- Moves not captured by GP registers because patients were not registered at one annual download
- Young people, particularly young men, can be slow to change their registration when they move
- Constraining GP register data to NHSCR
- Potential double counting of school boarders
- Out-flows to Scotland / Northern Ireland and allocating these to LAs

Assessment of Errors

- For each sex, and inflows and outflows separately, compare Census-based counts, c_{ix}, with the GP register-based counts, p_{ix}, for LA i and age x:

$$\text{lsf}_{ix} = \log \frac{c_{ix}}{p_{ix}}$$



Source: ONS

Modelling the Errors

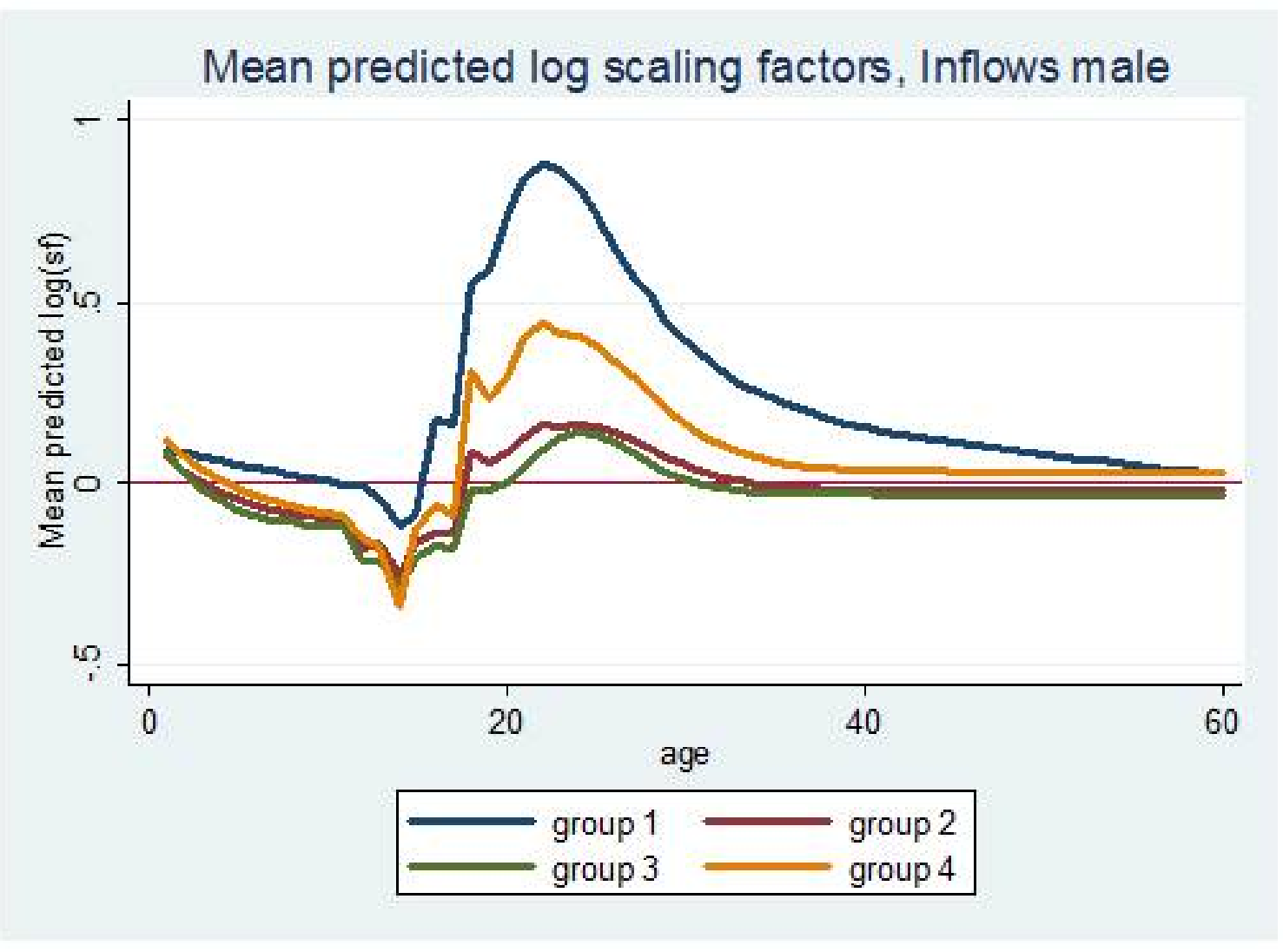
- Cluster the LAs using lsf_{ix} for inflows and outflows separately
- For each sex and cluster fit the model:

$$\begin{aligned}\mathsf{lsf}_{ix} &= a_0 \exp(-a_1 x) + a_2 \exp\left\{-a_3(x-a_4) - e^{-a_5(x-a_4)}\right\} + e_{ix} \\ &\quad -a_7 \sum_{j=0}^{17} I_j(x) + \sum_{j=12}^{17} b_j u_{ij} I_j(x) + \sum_{j=18}^{19} c_j v_{ij} I_j(x) + \boldsymbol{\beta}^T \boldsymbol{z}_i\end{aligned}$$

a modified Roger-Castro curve, where

$$\begin{aligned} u_{ij} &= \text{number of school boarders in LA } i \text{ aged } j \\ v_{ij} &= \text{number of student in LA } i \text{ aged } j \\ I_j(x) &= 1 \text{ if } x = j \text{ and 0 otherwise} \\ \boldsymbol{z}_i &= \text{covariates for LA } i \end{aligned}$$

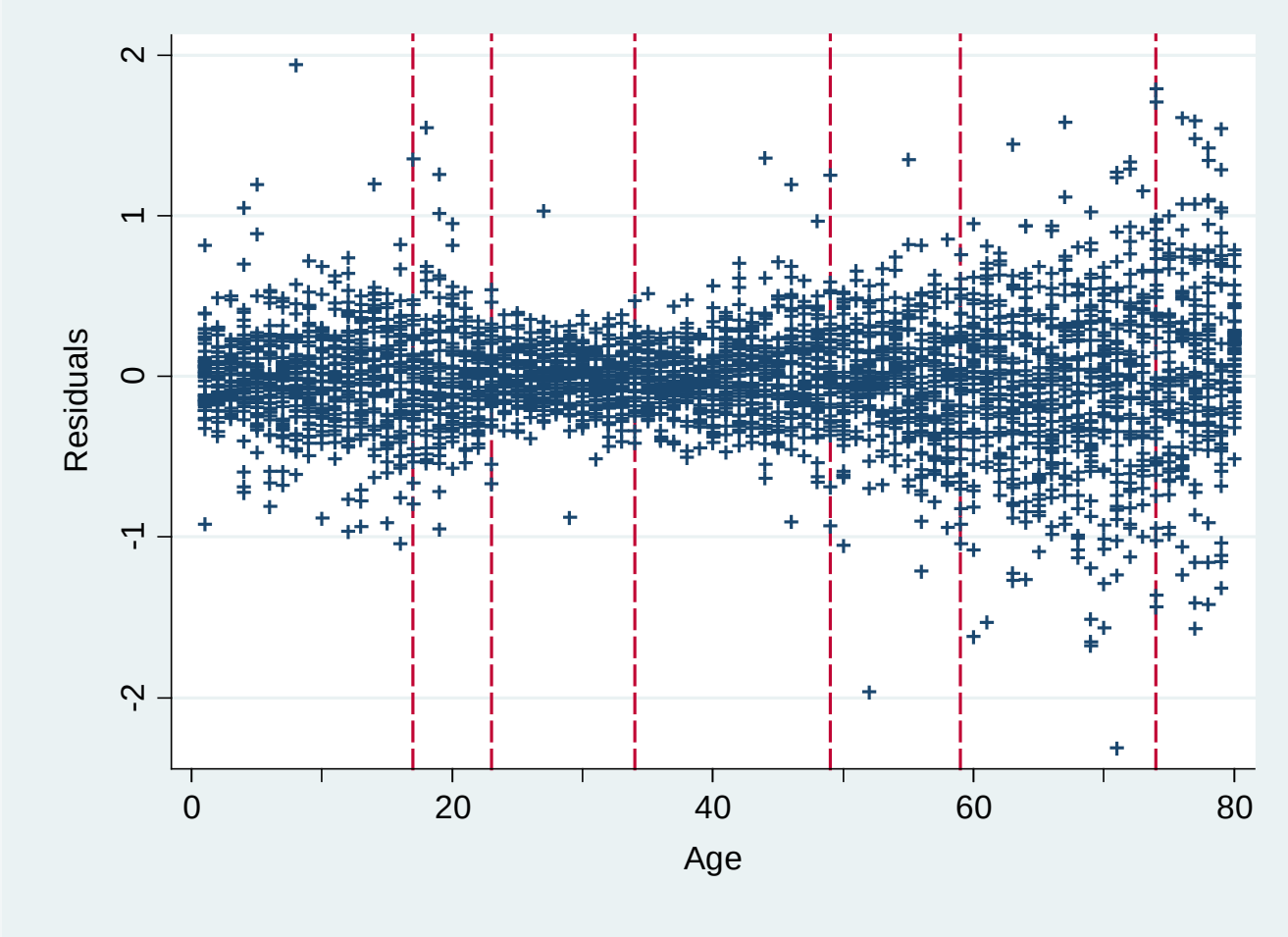
Modelling the Errors



Source: ONS

Modelling the Errors

Figure: Distribution of residuals by age (inflows males) with pots



Source: ONS

Simulating from Error Distribution

- Simulate

$$\mathsf{LSF}^*_{ixt} = \widehat{\mathsf{LSF}}_{ixt} + e^*_{ix}$$

where

$\widehat{\mathsf{LSF}}_{ixt} =$ predicated LSF for year t

$e^*_{ix} =$ residual sampled from the appropriate pot

- Calculate

$$p^*_{ixt} = p_{ixt} \exp(\mathsf{LSF}^*_{ix})$$

where

$p_{ixt} =$ GP register-based count in year t

- Repeat 1000 times for both sexes and all LAs, ages and years, for inflows and outflows

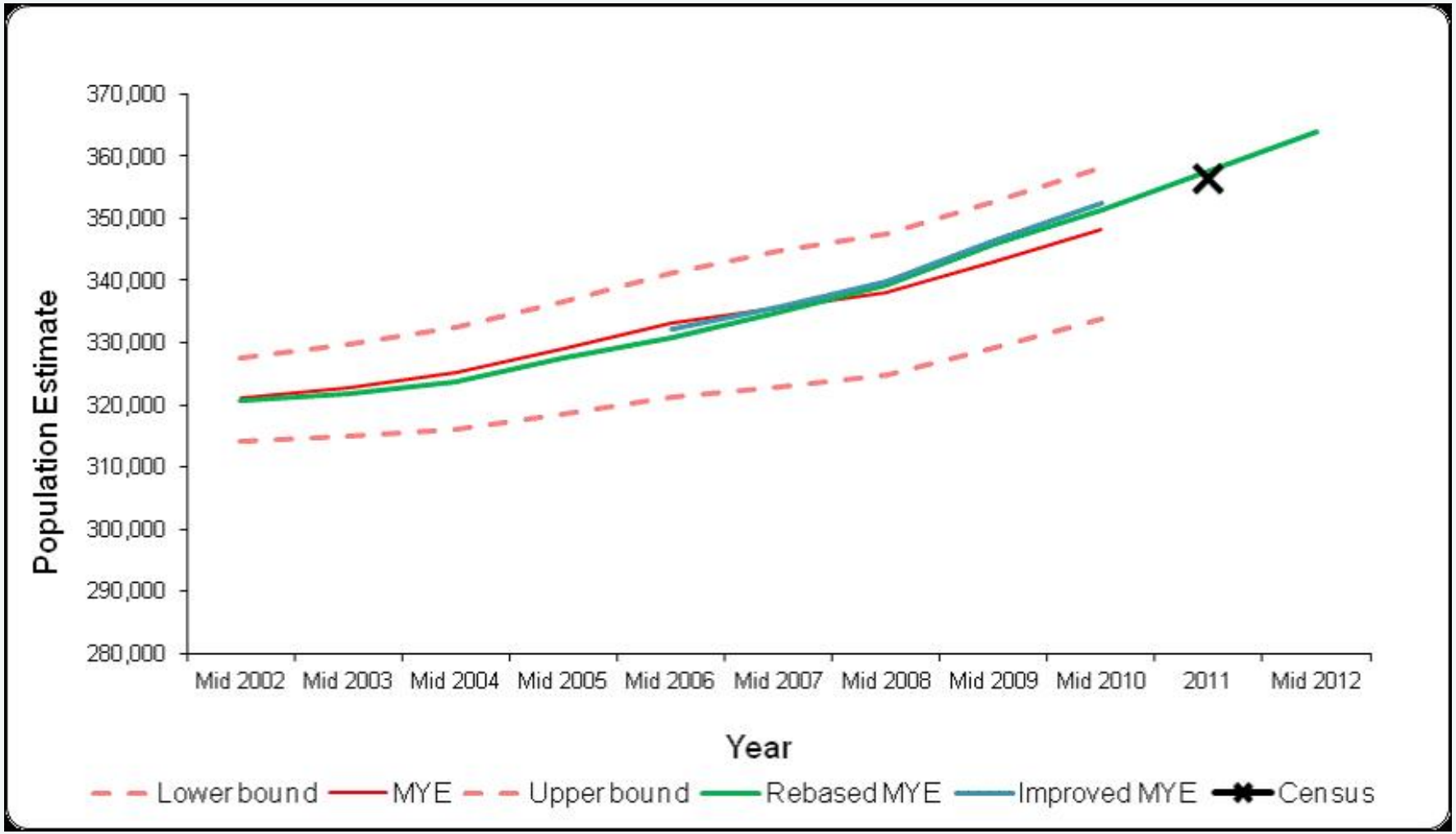
Results

For the 2001–10 series, uncertainty measures for the 376 local authorities in England and Wales included:

- 95% confidence for the MYE
- the uncertainty measure as a percentage of the population
- the percentage contribution that the 2001 census, internal migration and international migration made to the overall measure of uncertainty for each local authority
- an interactive map showing uncertainty levels for local authorities by year

Results

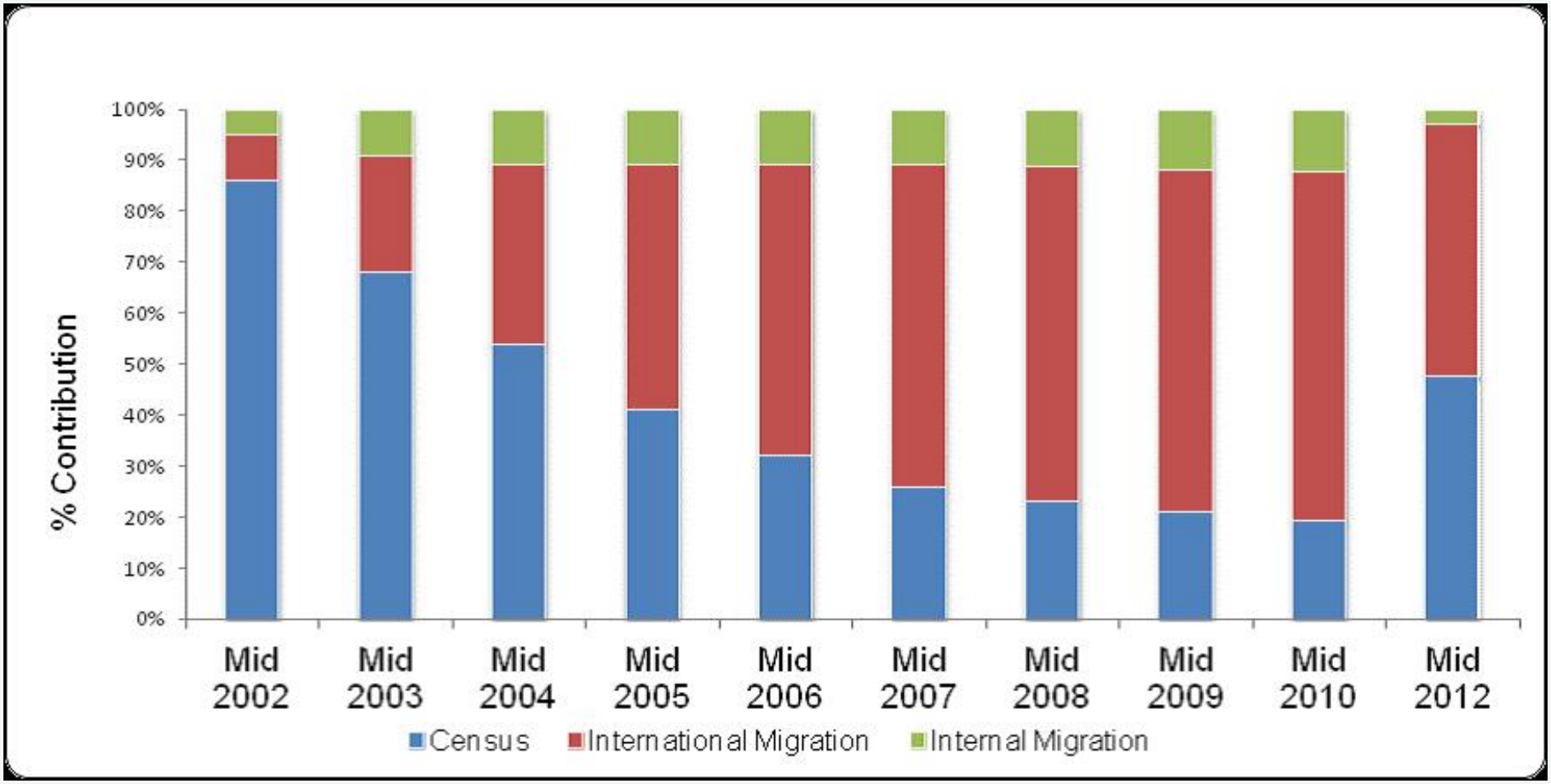
Figure: Uncertainty range around the mid-year population estimate from 2002 for a local authority within England and Wales



Source: ONS

Results

Figure: The proportional contribution that census, international and internal migration make to overall uncertainty for a local authority over the decade



Source: ONS

Summary

- Have developed a bespoke method to estimate the uncertainty in the MYE
- Based on simulation:
  - for internal migration, models are built for the errors in the PR based on discrepancies between the PR and the Census
  - for international migration, the error distributions are estimated by bootstrapping the IPS and LFS
- Work ongoing to allow for changes in the methodology used for the MYE since 2011

Acknowledgements and References

- All at the Office for National Statistics and University of Southampton who have worked on this project

- Search the web for "Measuring uncertainty in ONS population estimates"

- ONS (2010) Mid-year population estimates short methods guide
- ONS (2012) Methodology for deriving a statistical measure of uncertainty for local authority mid-year population estimates 2002–2010
- ONS (n.d.) Measuring uncertainty in ONS population estimates: capturing variability in statistics from combination of census, administrative and survey sources