Mining the HFD for robust fertility patterns over age and time

Carl Schmertmann Florida State University

Australia Data Official Statistics







OBJECTIVE

Forecast age- and cohort-specific ASFRs f(a,t) so that

- 1. Age schedules $f(\cdot,t)$ are plausible
- 2. Time trends $f(a, \cdot)$ are plausible
- 3. Uncertainty is estimable



STRATEGY

MINE HFD DATA (Austria ... Ukraine) FOR

- 1. Typical Age Schedules (SVD analysis)
- 2. Typical Time Trends (simple time series models)
- 3. (Co)variance of errors around typical patterns

STRATEGY (cont.)

Bayesian approach

A forecast is <u>less</u> likely, given already-observed rates, when it has

- 1. <u>Atypical Age Schedules</u>
- 2. <u>Atypical Time Trends</u>
- 3. <u>Atypical errors</u>

... as measured by squared error penalties (defined next)

Canada ASFR Surface Women Born 1906-1949



HFD (large set of complete rate surfaces)



Treat as giant A x N matrix of age schedules

Singular Value Decomposition: find principal components of age schedules



COHORT

$f(\cdot,t)$ well approx. by three components





Age



Age



HFD-CALIBRATED SHAPE PENALTIES

- calculate residual vectors ε_t for entire HFD
- calculate $\mathbf{\Omega}$ = average $\varepsilon_t \varepsilon_t'$ in HFD
- forecast is more likely when $\epsilon_t' \ \mathbf{\Omega}^{-1} \ \epsilon_t$ is small
- add penalty $-\frac{1}{2} \epsilon_t' \Omega^{-1} \epsilon_t$ to the (posterior) likelihood of a forecast

TIME SERIES MODELS

Two common *ad hoc* models for time series $f(a, \cdot)$

• Freeze-Rate

last observed rate at each age will continue

• Freeze-Slope

recent trends at each age will continue



HFD (large set of complete rate surfaces)



COHORT

Treat as multiple time series at different ages

DISTRIBUTION OF TIME SERIES RESIDUALS IN HFD



HFD-CALIBRATED TIME SERIES PENALTIES (both freeze-rate and freeze-slope)

- calculate residual vectors ε_a over entire HFD
- calculate $\sigma^2(a)$ = average $\varepsilon_a' \varepsilon_a$ in HFD
- forecast is more likely when $\sigma^{-2}(a) \epsilon_a' \epsilon_a$ is small
- add penalty -^{1/2} σ⁻²(a) ε_a'ε_a
 to the (posterior) likelihood of a forecast

Bayesian Posterior Distribution for {f} | Data

 $\ln P(f \mid \text{Data, HFD}) = c + \ln L(\text{ Data} \mid f) - \text{PENALTIES}$

Std Likelihood
surface shouldPrior
surface shouldmatch obs datahave HFD-like
age schedules
and time series

 \rightarrow Each possible forecast has a different likelihood

→With normal likelihood, there are closed-form solutions for means and covariances of forecast rates

Australia Data Official Statistics



Australia Forecast Posterior Mean





Australia 2010 forecast



All data and R code at *http://cohort-fertility.schmert.net*

Thanks!





Emilio Zagheni





Mikko Myrskylä