Is there a lost US generation?

Results from a Bayesian-based period-cohort model of US male and female adult mortality

Dr David G McCarthy
University of Georgia

The University of Georgia
Terry College of Business

Quick summary

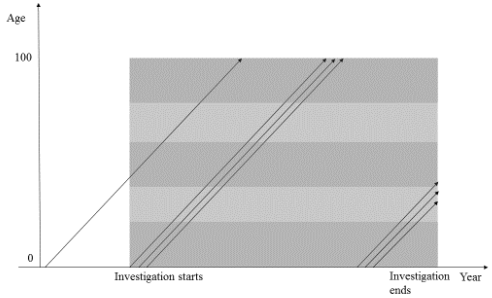
- Advantages of using age and cohort as the primary variables of analysing & forecasting mortality data:
  - Cohort analysis applies to real people, with consequent advantages
  - Age-cohort fits adult mortality data just as well as age period (if not better)
  - Cohort effects appear stable across (adult) life
  - Key: period-related randomness has declined in developed countries, creating challenges for using it alone to forecast future mortality variation
- Disadvantages: estimation difficulties for recent cohorts
- Solution: use Bayesian maximum *a posteriori* estimation to jointly estimate cohort mortality parameters given a prior about how they change over time (that is, slowly)
- Analysis: Apply approach to a variant of the Cairns-Blake-Dowd (2006) model, estimate it on US M & F data
- Findings: Approach seems to produce sensible estimates & reasonable ranges for future cohort adult life expectancy
- Evidence of a significant reduction aggregate mortality improvement for those born in the 1950's, but not as bad as LC

Some history

- First demographers looked at the effect of age only on mortality (e.g. Ulpian, ~150CE, Halley, 1693)
- Next, added period: age-period combinations are the data produced by mortality investigations (Lee & Carter, 1992; Cairns, Blake & Dowd, 2006)
- Next, looked at cohort data in the context of existing age-period models (Willetts, 2003, Haberman & Renshaw, 2006 etc)
- BUT:
  - Age, period & cohort are collinear
  - Including all three effects difficult and contentious (across the social sciences, Yang et al (2008) w/ Luo (2013), Chauvel & Leist (2016) etc)
  - Uncertainty in period-related fluctuations in mortality rates appears to be dying away (in rich countries, at least)
  - Real people have fixed cohorts
- SO:
  - What about using age and cohort as the first two variables, with period as an afterthought?

An ad-hoc investigation on US data (I)

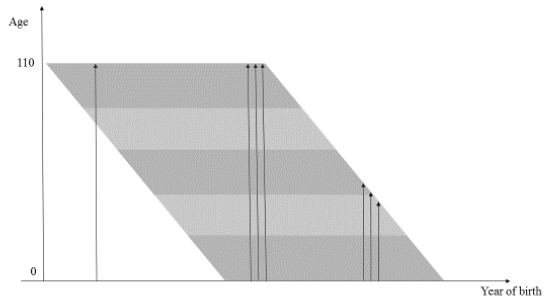
Age-period regressions



$$\log(q_{x,t}) = \alpha + \beta' I_x + \gamma' I_t + \varepsilon_{x,t}$$

Subsample	Males	
	Adj-R squared	Root mean square error
Age 0-9	0.9961	0.08454
Age 10-19	0.9410	0.18702
Age 20-29	0.9660	0.05286
Age 30-39	0.9856	0.04293
Age 40-49	0.9897	0.04083
Age 50-59	0.9932	0.03237
Age 60-69	0.9948	0.02654
Age 70-79	0.9944	0.02574
Age 80-89	0.9853	0.03558
Age 90-99	0.9557	0.03815

Age-cohort regressions

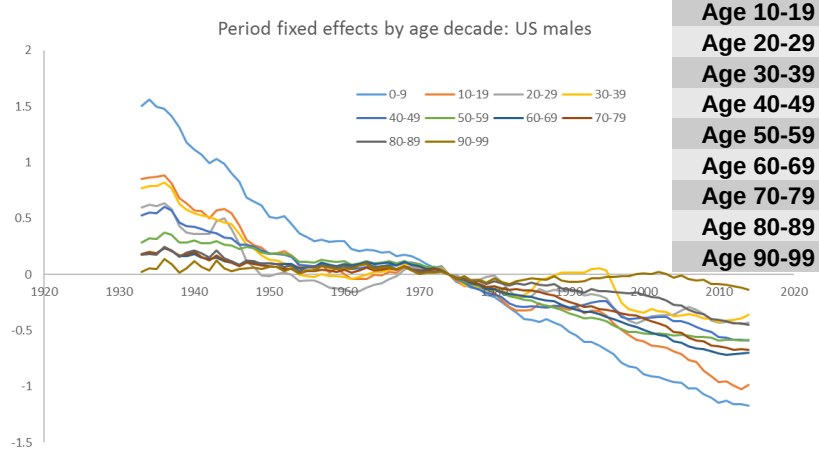


$$\log(q_{x,c}) = \alpha + \beta' I_x + \xi' I_c + \varepsilon_{x,c}$$

Subsample	Males	
	Adj-R squared	Root mean square error
Age 0-9	0.9945	0.10044
Age 10-19	0.9385	0.19097
Age 20-29	0.9176	0.08231
Age 30-39	0.9565	0.07461
Age 40-49	0.9831	0.05227
Age 50-59	0.9952*	0.02728*
Age 60-69	0.9952*	0.02544*
Age 70-79	0.9933	0.02837
Age 80-89	0.9898*	0.02962*
Age 90-99	0.9487	0.04103

An ad-hoc investigation on US data (II)

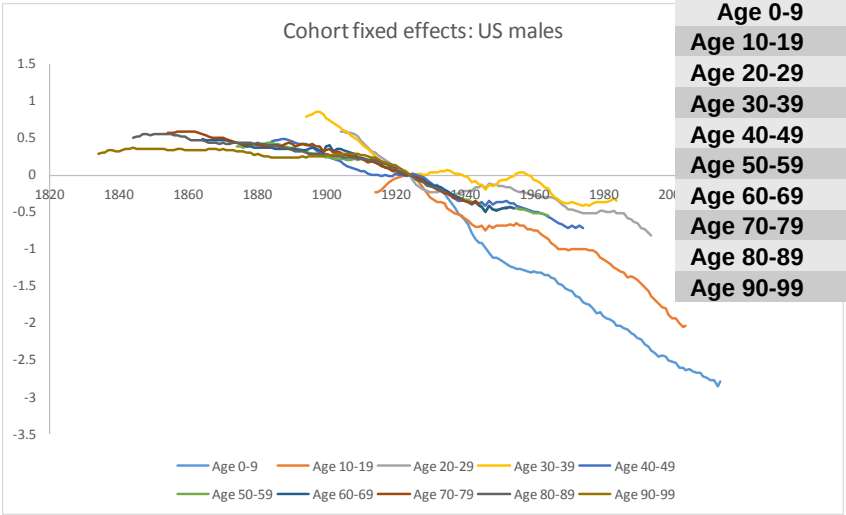
Age-period regressions



	Subsample									
	Age 0-9	Age 10-19	Age 20-29	Age 30-39	Age 40-49	Age 50-59	Age 60-69	Age 70-79	Age 80-89	Age 90-99
Age 0-9	1.000									
Age 10-19	0.656	1.000								
Age 20-29	0.554	0.644	1.000							
Age 30-39	0.537	0.629	0.826	1.000						
Age 40-49	0.529	0.588	0.704	0.770	1.000					
Age 50-59	0.454	0.481	0.646	0.693	0.752	1.000				
Age 60-69	0.348	0.461	0.474	0.539	0.660	0.760	1.000			
Age 70-79	0.288	0.408	0.351	0.440	0.612	0.697	0.906	1.000		
Age 80-89	0.227	0.309	0.267	0.349	0.532	0.606	0.834	0.922	1.000	
Age 90-99	0.242	0.308	0.243	0.310	0.499	0.598	0.787	0.877	0.958	1.000

$$corr(\gamma_t^i - \gamma_{t-1}^i, \gamma_t^j - \gamma_{t-1}^j) \quad [H_0 : LC = 1]$$

Age-cohort regressions



	Subsample									
	Age 0-9	Age 10-19	Age 20-29	Age 30-39	Age 40-49	Age 50-59	Age 60-69	Age 70-79	Age 80-89	Age 90-99
Age 0-9	1.000									
Age 10-19	0.122	1.000								
Age 20-29	-0.174	0.155	1.000							
Age 30-39	0.115	0.147	0.393	1.000						
Age 40-49	0.169	0.530	0.338	0.345	1.000					
Age 50-59	0.181	0.354	0.397	0.357	0.574	1.000				
Age 60-69	0.092	0.391	0.460	0.389	0.508	0.684	1.000			
Age 70-79	-0.269	0.019	0.193	0.083	0.384	0.536	0.703	1.000		
Age 80-89	-0.379	0.164	0.093	-0.226	0.107	0.539	0.498	0.513	1.000	
Age 90-99	-	0.278	0.268	-0.014	-0.293	0.178	0.271	0.281	0.449	1.000

$$corr(\xi_t^i - \xi_{t-1}^i, \xi_t^j - \xi_{t-1}^j) \quad [H_0 : LC^C = 1]$$

The University of Georgia
TERRY COLLEGE OF BUSINESS

Is variance in period innovations declining?

- Evidence that the variance of random innovations in period effects are declining comes from three sources
  - 1. Ad-hoc investigation. Variance ratio tests of the equality of the variance of differenced period coefficients in first and second halves of sample rejected in all subsamples of the data except one (variance of differenced cohort effects, if anything, may be increasing)

	Cohort effects				Period effects			
	Statistic	DF1	DF2	p-value	Statistic	DF1	DF2	p-value
Age 0-9	-	-	-	-	1.8086	39	39	0.034
Age 10-19	0.1290	8	79	0.002	2.2458	39	39	0.007
Age 20-29	0.6875	18	69	0.188	2.8957	39	39	0.001
Age 30-39	0.9978	28	59	0.513	1.7137	39	39	0.048
Age 40-49	0.3621	38	49	0.001	1.3869	39	39	0.156
Age 50-59	0.3962	48	39	0.001	2.8624	39	39	0.001
Age 60-69	0.7900	58	29	0.220	2.4701	39	39	0.003
Age 70-79	0.9873	68	19	0.459	2.7500	39	39	0.001
Age 80-89	0.7194	78	9	0.207	2.1874	39	39	0.008
Age 90-99	-	-	-	-	2.0646	39	39	0.013

  - 2. Estimates of Lee-Carter model clearly show the decline in variation in time parameter (e.g. McCarthy & Miles 2014, *** 2017)
  - 3. McCarthy and Wang (2017), presented in this conference, fit this model to ~30 countries, variance ratio tests within countries and χ^2 tests across them rejected
- Conclusion: cannot rely on variance in period effects to generate *all* mortality uncertainty

So what to do?

- Two main options:
  - Cohort version of Lee-Carter (1992):

$$\log(m_{x,c}) = \alpha_x + \beta_x k_c + \delta' I_{c+x} + \varepsilon_{x,c}$$

    - (we're working on it)
  - Cohort version of Cairns-Blake-Dowd (2006):

$$\log(m_{x,c}) = \alpha' I_c + \beta' I_c \frac{(x-k_1)}{k_2} + \gamma' I_c \left(\frac{x-k_1}{k_2}\right)^2 + \delta' I_{c+x} + \varepsilon_{x,c}$$

    - (this paper)

- Difficulty:
  - How to fit model to recent cohorts where not much data is available?
  - We use Bayesian maximum *a posteriori* estimation (de Groot, 1970)

Model description

- Modified version of CBD (2006)

$$\log(m_{x,c}) = \alpha' I_c + \beta' I_c \frac{(x-k_1)}{k_2} + \gamma' I_c \left(\frac{x-k_1}{k_2}\right)^2 + \delta' I_{c+x} + \varepsilon_{x,c}$$

  - α, β, γ constant for each *cohort* (this is the identifying assumption aka Fienberg and Mason (1978))
  - Model log of central rate of mortality (estimated log hazard rate) to enable easy conversion to survival distribution
  - Add a quadratic term (only needed for US females)
  - Add period random effects to capture non-linear period-related mortality shocks δ: $\delta_{c+x} \sim N(0, \sigma_\delta^2)$ (remember: $(c+x=t)$; linear changes in mortality captured by cohort parameters)
  - Make an appropriate assumption about error term, theory suggests for perfect model:

$$\varepsilon_{x,c} \sim N(0, \omega_{x,c}), \quad \omega_{x,c} = \frac{1}{m^*_{x,c} E_{x,c}} \cong \frac{1}{D_{x,c}}$$

    for our imperfect model we assume:

$$\varepsilon_{x,c} \sim N(0, \sigma_\varepsilon^2 \omega_{x,c})$$

Bayesian maximum *a posteriori* estimation

- Step 1
  - Estimate a prior distribution for the parameters $\alpha, \beta\ [, \gamma]$
- Step 2
  - Calculate the posterior distribution of these parameters, conditional on the data and the chosen prior
- Step 3
  - For point estimates, choose the mode of the posterior distribution
  - For actual posterior distribution, can do Metropolis-Hastings, we use an approximation due to Laplace
- Step 4
  - Point estimates for calculating fitted values, estimating random effects δ, measuring goodness-of-fit
  - Post. dstbn & VAR for forecasting

Step 1: choosing the prior

- Assume that first differences of $\theta_c = (\alpha, \beta, \gamma)_c$ follow a first-order VAR, with exogenous variables (also differenced), x to capture known causes of mortality variation (e.g. smoking)

$$(\Delta\theta)_c - \hat{\mu} - \hat{\Delta}_1(\Delta\theta)_{c-1} - \hat{\kappa}_0(\Delta x)_c = \nu_c, \quad \nu_c \sim N(0, \Sigma) \text{ iid}$$

- Hence, the prior distribution is:

$$p_1(\theta \mid \xi \equiv \hat{\mu}, \hat{\Delta}_1, \hat{\Sigma}, \hat{\kappa}) = \prod_{all\ c} \phi((\Delta\theta)_c - \hat{\mu} - \hat{\Delta}_1(\Delta\theta)_{c-1} - \hat{\kappa}_0(\Delta x)_c, 0, \hat{\Sigma})$$

- Values of fitted parameters estimated off VAR fitted to cohort-by-cohort estimates of $\theta_c = (\alpha, \beta, \gamma)_c$ obtained from regressions on reasonably complete cohorts, fitted using standard ML
- Prior then reflects how we believe the value of $\theta_c = (\alpha, \beta, \gamma)_c$ changes from cohort to cohort

Step 2: obtaining the posterior

- Bayes' Theorem gives the posterior distribution as

$$p_2(\theta,(\sigma_\delta^2,\sigma_\varepsilon^2)\,|\,\xi,\{m_{x,c},\omega_{x,c}\}) = \frac{\ell(\{m_{x,c},\omega_{x,c}\}\,|\,\theta,(\sigma_\delta^2,\sigma_\varepsilon^2))\,p_1(\theta\,|\,\xi)}{\int \ell(\{m_{x,c},\omega_{x,c}\}\,|\,\theta,(\sigma_\delta^2,\sigma_\varepsilon^2))\,p_1(\theta\,|\,\xi)d\theta}$$

or:

$$\ell p_2(\theta,(\sigma_\delta^2,\sigma_\varepsilon^2)\,|\,\xi,\{m_{x,c},\omega_{x,c}\}) = K + \underbrace{\ell\ell(\{m_{x,c},\omega_{x,c}\}\,|\,\theta,(\sigma_\delta^2,\sigma_\varepsilon^2))}_{\text{Fisherian log-likelihood of data } \{m_{x,c},\omega_{x,c}\}} + \underbrace{\ell p_1(\theta\,|\,\xi)}_{\text{logged prior distribution}}$$

(Hence, BMAP estimator is a penalized ML estimator)

Step 3: selecting estimates

- For point estimates choose the mode of the posterior dstbn:

$$(\hat{\theta},(\hat{\sigma}_{\delta}^{2},\hat{\sigma}_{\varepsilon}^{2})) = \underset{\theta,(\sigma_{\delta}^{2},\sigma_{\varepsilon}^{2})}{\arg\max}\, \ell p_{2}(\theta,(\sigma_{\delta}^{2},\sigma_{\varepsilon}^{2}) \mid \xi,\{m_{x,c},\omega_{x,c}\})$$

- To approximate the posterior distribution itself, use a result from Laplace:

$$p_{2}(\theta) \approx K(\hat{\theta}) \exp\left[\frac{1}{2}(\theta-\hat{\theta})' \ell p_{2,\theta\theta}(\hat{\theta})(\theta-\hat{\theta})\right]$$

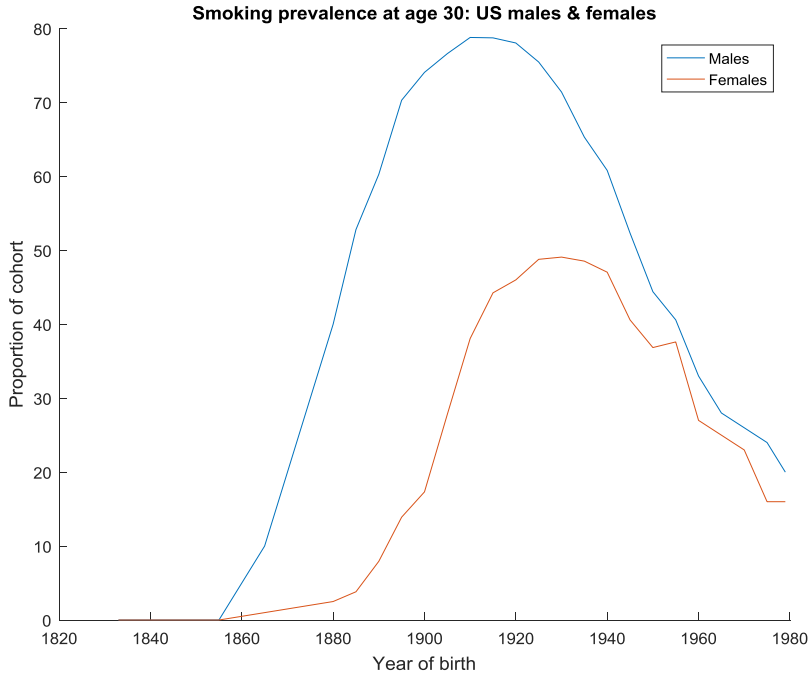
[Note: posterior is approximately MVN (property shared by all modal estimators, including ML)]

(We are also using the Metropolis-Hastings algorithm to test accuracy for marginals)

Application

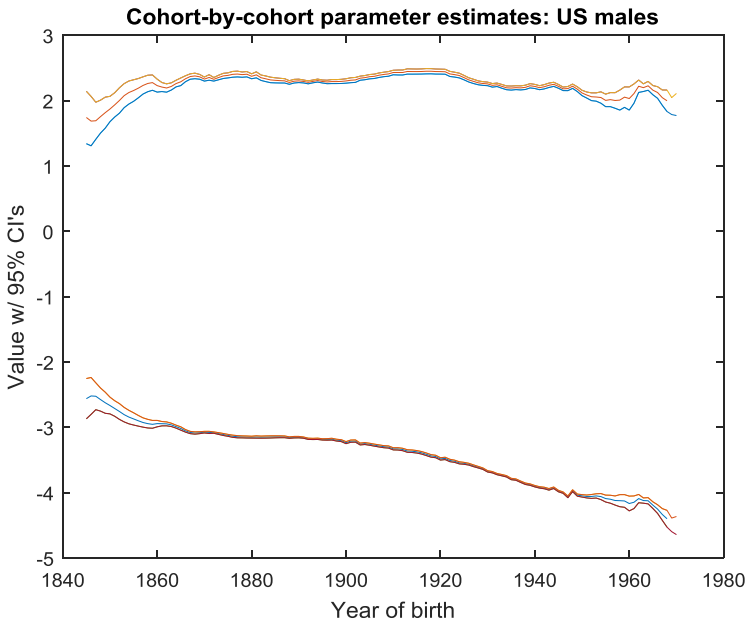
- US male and female mortality data, 1933 – 2014, from www.mortality.org, focus on adult mortality (age 35-100)

- For exogenous variable, use cohort smoking prevalence data collected from Forey *et al* (1997), updated by our own estimates from US survey data (NHIS)
  - Use proportion of the cohort that smoked (cigarettes) at age 30 (usu. ~peak smoking)
  - Pattern similar across cohorts
  - Peaked for men born ~1910
    - (Later for women)
  - Rapid decline
    - (Slower for women)
  - M & F proportion now similar



Step 1: choosing the prior

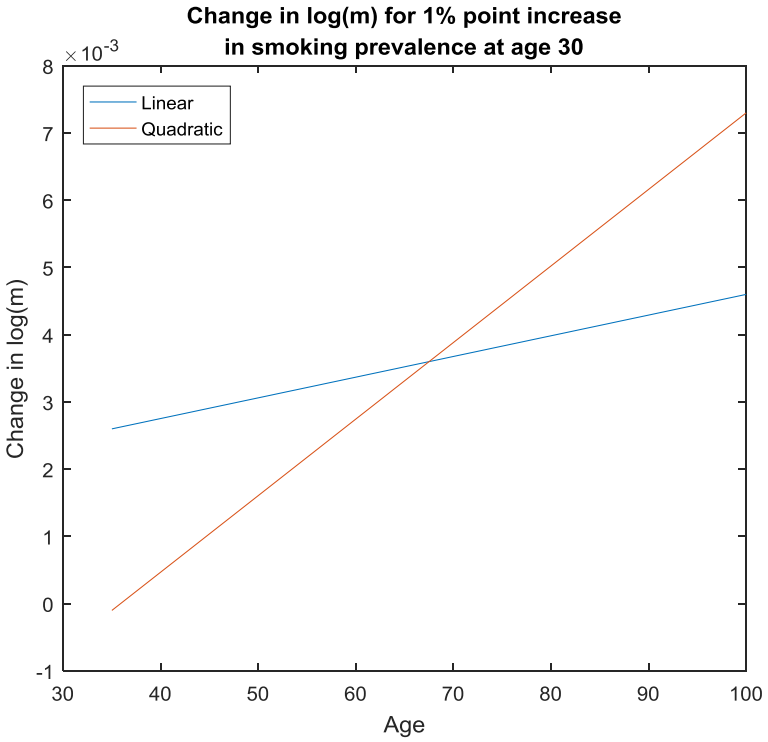
- Use standard ML to fit cohort-by-cohort estimates to $\alpha, \beta\ [, \gamma]$
- Results shown for US males (earliest and latest cohorts excluded)
- Cohorts born after 1870 and before 1942 are reasonably precise
- Fit a VAR to first differences of these estimates only
- Include first differences of exogenous variables, but estimate smoking effect using a SU-VAR on males and female cohorts jointly



Step 1: choosing the prior

VAR estimates			
		Quadratic	Linear
		Females	Males
D.Alpha:	D.Alpha.L1	0.2779	0.3863
		(0.0670)	(0.0763)
	Change in smoking prevalence	0.0036	0.0036
		(0.0007)	(0.0007)
	Constant	-0.0127	-0.0095
		(0.0026)	(0.0023)
	RMSE	0.0196	0.0169
	R-squared	-0.2632	-0.1593
D.Beta:	D.Beta.L1	0.2713	0.3398
		(0.0811)	(0.0870)
	Change in smoking prevalence	0.0037	0.0010
		(0.0011)	(0.0010)
	Constant	-0.0016	-0.0024
		(0.0021)	(0.0020)
	RMSE	0.0169	0.0163
	R-squared	0.2736	0.1161
D.Gamma:	D.Gamma.L1	0.2870	
		(0.0626)	
	Change in smoking prevalence	0.0000	
		(0.0020)	
	Constant	-0.0038	
		(0.0043)	
	RMSE	0.0354	
	R-squared	0.0636	

- Own first-order lags significant
- Steady-state: decline in alpha, beta and gamma
- Increase in smoking prevalence increases cohort mortality at all ages, but more at older ages



Step 2: obtaining the posterior

- Given the estimated parameters of the VAR, ξ, the logged posterior pdf is easily calculated for *any* choice of parameters $\{\theta, (\sigma_\delta^2, \sigma_\varepsilon^2)\}$, from Bayes' Theorem:

$$\ell p_2(\theta, (\sigma_\delta^2, \sigma_\varepsilon^2) \mid \xi, \{m_{x,c}, \omega_{x,c}\}) = K + \underbrace{\ell\ell(\{m_{x,c}, \omega_{x,c}\} \mid \theta, (\sigma_\delta^2, \sigma_\varepsilon^2))}_{\text{Fisherian log-likelihood of data } \{m_{x,c}, \omega_{x,c}\}} + \underbrace{\ell p_1(\theta \mid \xi)}_{\text{logged prior distribution}}$$

- The log likelihood function of the data is penalized by the addition of the prior information

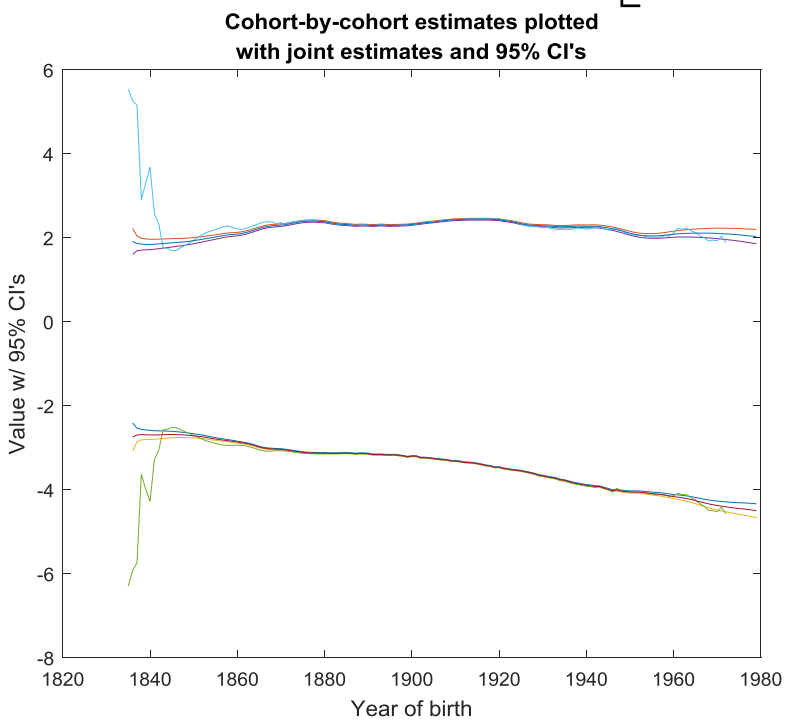
Step 3: obtaining the estimates

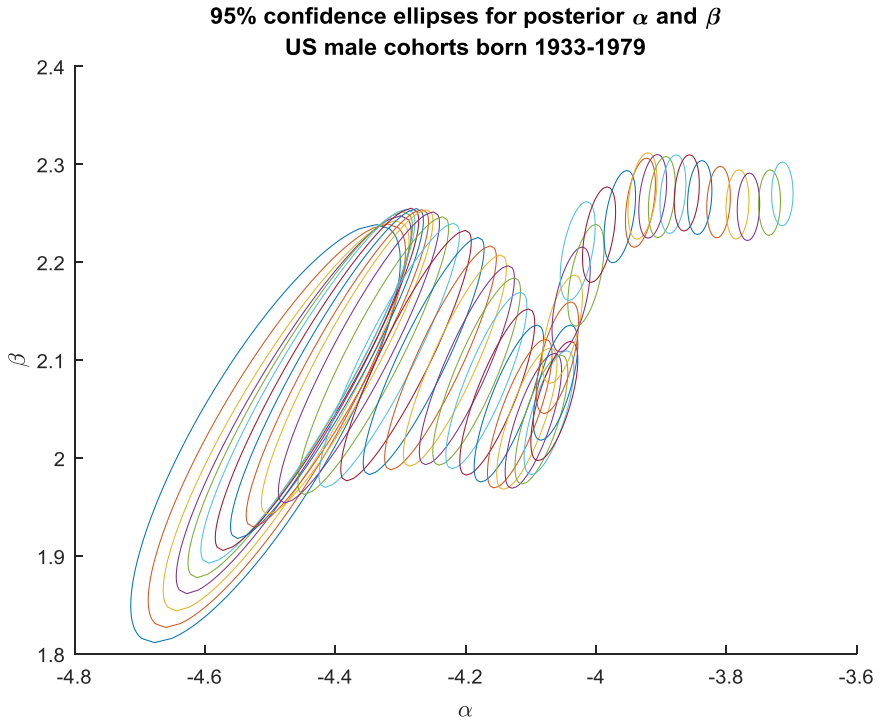
- Use numerical software to obtain point estimates for $\{\theta,(\sigma_\delta^2,\sigma_\varepsilon^2)\}$ by solving:

$$(\hat{\theta},(\hat{\sigma}_\delta^2,\hat{\sigma}_\varepsilon^2)) = \underset{\theta,(\sigma_\delta^2,\sigma_\varepsilon^2)}{\arg\max}\, \ell p_2(\theta,(\sigma_\delta^2,\sigma_\varepsilon^2)\,|\,\xi,\{m_{x,c},\omega_{x,c}\})$$

- Obtain confidence intervals & ellipses using:

$$p_2(\theta) \approx K(\hat{\theta}) \exp\left[\frac{1}{2}(\theta-\hat{\theta})'\,\ell p_{2,\theta\theta}(\hat{\theta})(\theta-\hat{\theta})\right]$$

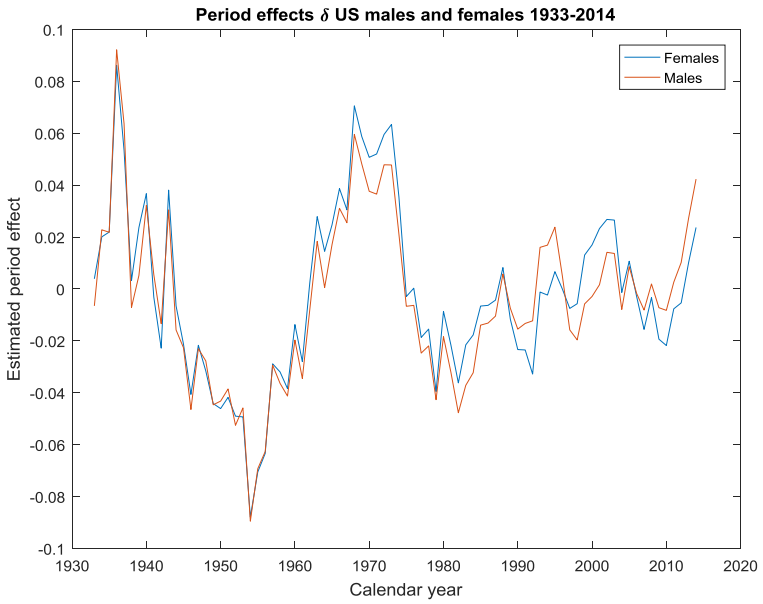




Step 4A: goodness of fit

- Model fits historical data extremely well
- Using adjusted R-squared as a model selection criterion suggests linear model for males & quadratic model for females, as does a check of the residuals (not shown)
- Estimated period effects highly correlated across M & F

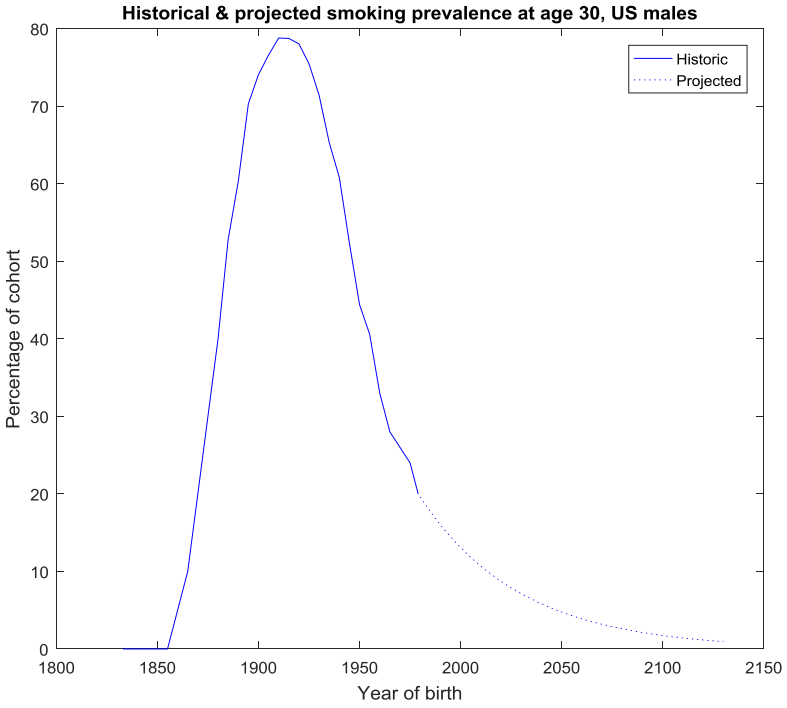
	Males			
Alpha (147 cohorts)	X	X	X	X
Beta (147 cohorts)	X	X	X	X
Gamma (147 cohorts)	X	X	-	-
Random period effects	X	-	X	-
log(sigma²(epsilon))	0.4947	0.5977	0.5219	0.6663
	(0.0199)	(0.0177)	(0.0197)	(0.0188)
log(sigma²(delta))	-3.7114	-	-3.5768	-
	(0.1957)		(0.1906)	
R-squared	0.9989	0.9984	0.9989	0.9985
Adjusted R-squared	0.9988	0.9983	0.9989	0.9984
N	5412	5412	5412	5412
	Females			
Alpha (147 cohorts)	X	X	X	X
Beta (147 cohorts)	X	X	X	X
Gamma (147 cohorts)	X	X	-	-
Random period effects	X	-	X	-
log(sigma²(epsilon))	0.4500	0.5881	0.6137	1.3642
	(0.0197)	(0.0174)	(0.0196)	(0.0378)
log(sigma²(delta))	-3.5289	-	-1.8582	-
	(0.1920)		(0.1759)	
R-squared	0.9993	0.9989	0.9918	0.9932
Adjusted R-squared	0.9992	0.9988	0.9913	0.9928
N	5412	5412	5412	5412

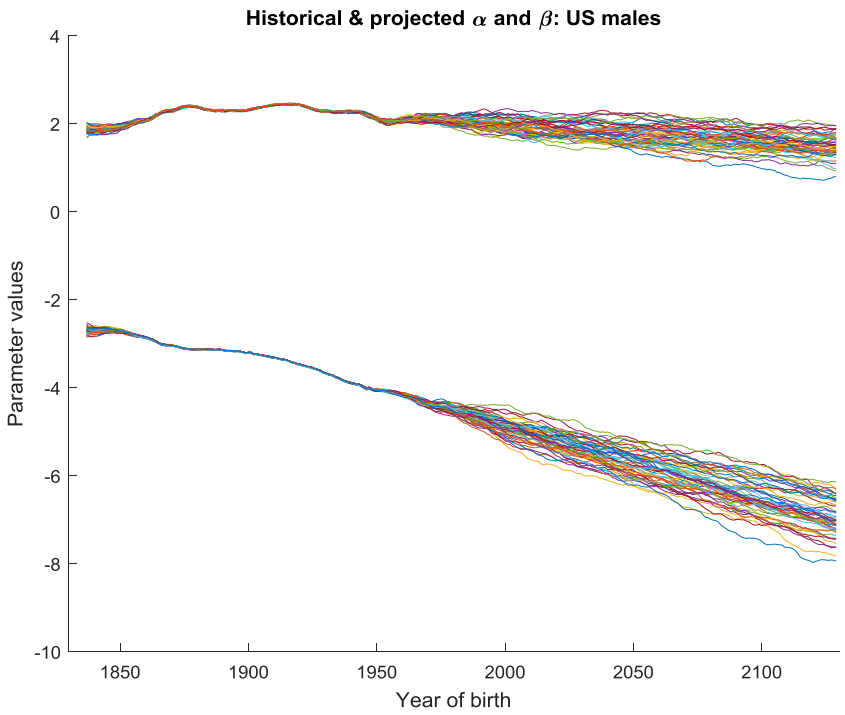


The University of Georgia
TERRY COLLEGE OF BUSINESS

Step 4B: forecasting

- Project what will happen to smoking rates (3% p.b.y. decline seems to fit well)
- Take random draws from posterior distribution of fitted parameters to get starting point
- Use fitted VAR to project these forward (including shocks & projected changes in smoking behaviour)





Step 4B: forecasting

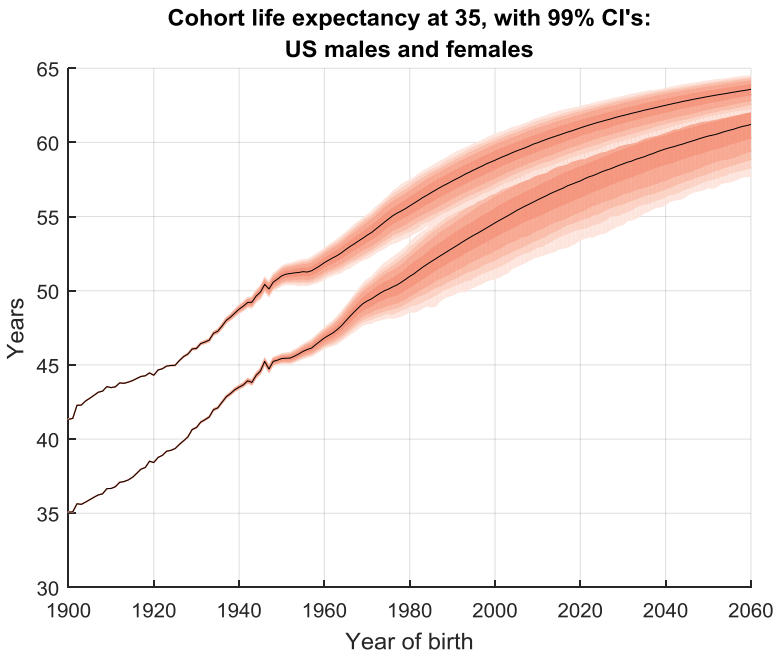
- Use projected values of α, β, γ , fitted values of $\{\log(\sigma_\delta^2), \log(\sigma_\varepsilon^2)\}$ to project future mortality hazard rates using:

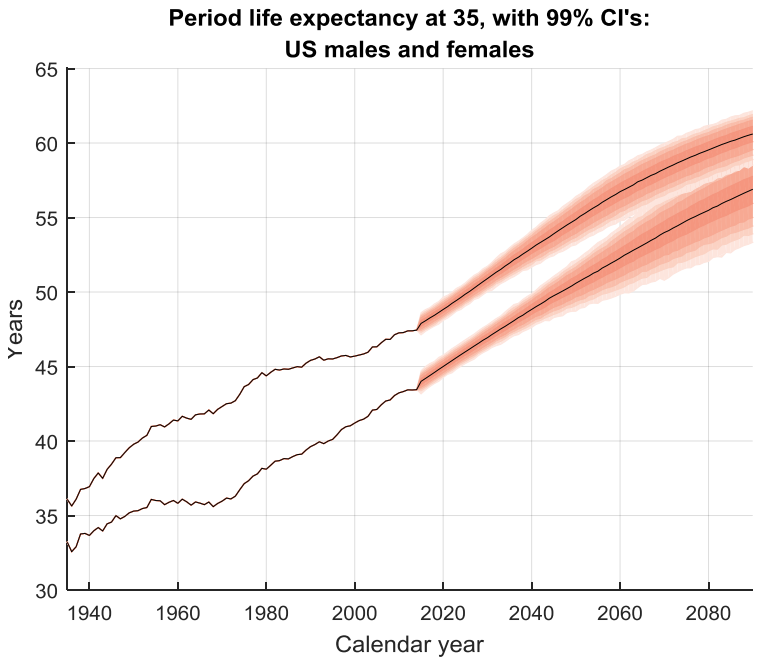
$$\log(m_{x,c}) = \alpha' I_c + \beta' I_c \frac{(x-k_1)}{k_2} + \gamma' I_c \left(\frac{x-k_1}{k_2}\right)^2 + \delta' I_{c+x} + \varepsilon_{x,c}$$

- Use these to estimate future life expectancy (or other quantities of interest, e.g. pension fund, annuity, SS liabilities)
- Two options
  - Do lots of Monte-Carlo runs to obtain CI's for variables of interest
  - Use (known) distribution of α, β, γ & known VAR to obtain computationally quicker but approximate theoretical distributions using survival distribution theory
    - Generate distributions of 'pseudo-parameters', use these & known Bayesian results to generate approximate posterior

Step 4B: forecasting

- Use forecasts of mortality rates to generate
  - Period life expectancy at 35 and 65, by calendar year
  - Cohort life expectancy at 35 and 65, by year of birth

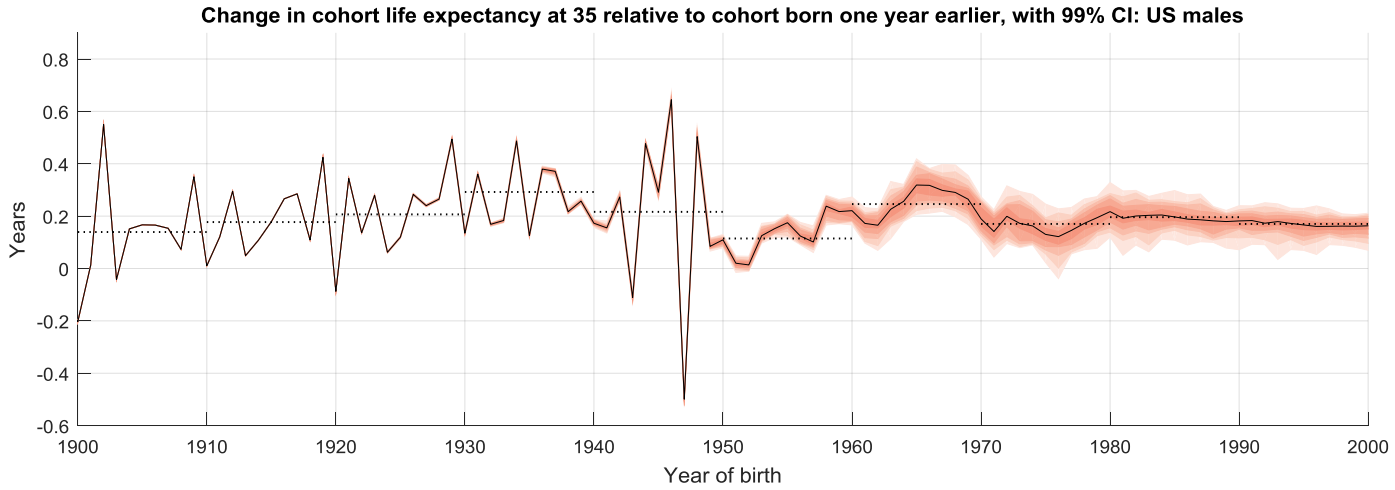


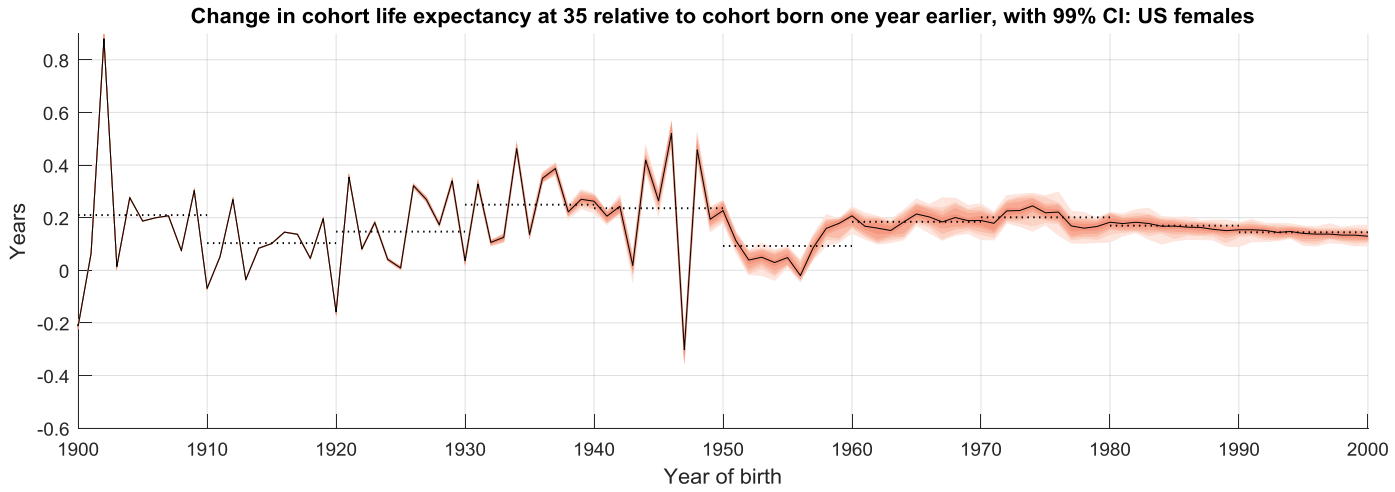


The University of Georgia
TERRY COLLEGE OF BUSINESS

Step 4B: forecasting

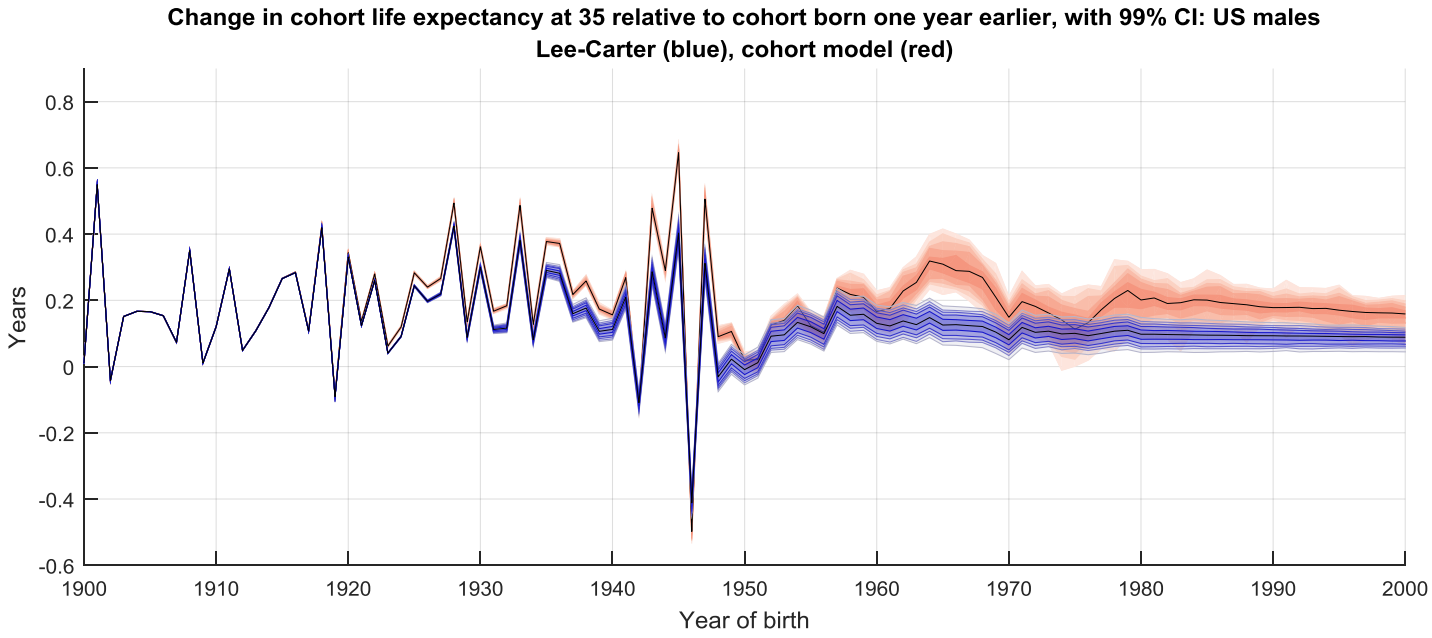
- Decadal cohort born 1950-1960 has lowest average rate of improvement in life expectancy at 35 of any such cohort in our dataset





Step 4B: forecasting

- Fitted Lee-Carter model with parameter uncertainty to same data using ML (shown for males)
- Cohort-based model predicts:
  - Greater increases in life expectancy of cohorts starting in 1925 (largely due to projected decline in smoking rates)
  - Greater improvement in LE of those born in 1960-1970 (cohort effect)
  - Models converge from ~2040 (not shown)



Conclusion

- Proposed use of Bayesian maximum *a posteriori* estimation to jointly estimate cohort mortality parameters given a prior about how they change over time (that is, slowly)
- Applied approach to a variant of the Cairns-Blake-Dowd (2006) model, with period effects, estimated it on US M & F data
- Findings: Approach seems to produce sensible estimates & reasonable ranges for future cohort adult life expectancy
- Evidence of a significant reduction aggregate mortality improvement for those born in the 1950's
  - But projected mortality improvements appear to be better than the LC model would predict
  - Largely due to projected effect of reduced smoking rates on cohort mortality
  - (Models converge in differences after around 2040)
- Future work
  - Use this model to investigate international mortality patterns (w/ Wang)
  - Estimate cohort-based LC using BMAP