9 Supplementary Appendix A. Data Appendix (Cai, Lau and Yuen)

In this appendix, we lay out the details of data construction and definitions in Figure 1. We use the data from the Current Population Survey (CPS) to compute the average retirement age from each cohort. In particular, we obtain the harmonized Basic Monthly CPS data from IPUMS (Flood et al., 2018). We restrict our sample to those US men from the 1920 cohort to the 1950 cohort based on their birth year[41]. All observations are weighted by the Final Basic Weight in the CPS.

We follow the definition of Munnell (2011) to compute the average retirement age. Specifically, for each cohort we calculate the labor force participation rate at each age. The average retirement age is then defined as the age at which the participation rate drops below 50 percent (See Figure A1). In particular, we use the first age if the participation rate rises above 50 percent at a later age. Moreover, we use interpolation to obtain a continuous measure of average retirement age. In Figure 1(a) we further compute the average retirement age for each cohort and education group.

[Insert Figure A1 here.]

Finally, the average schooling years is from the schooling years data set (for US men) used in Goldin and Katz (2008).

References

[1] Flood S., King M., Rodgers R., Ruggles S. and Warren J. R. (2018), Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V6.0.

[2] Munnell, A. H. (2011), What is the average retirement age? Centre for Retirement Research, Boston College, 11(11), 1-7.

[41] The range of cohorts depends solely on the availability of the relevant data in the CPS.

10 Supplementary Appendix B. Derivations for the model with direct utility benefit of schooling (Cai, Lau and Yuen)

In this model, the individual maximizes (32) subject to (1), (3) and (4). Again using standard techniques of dynamic optimization, we can obtain (5). Following the derivation in the baseline model, we denote

$$U_b(S,R) = \int_0^T \exp(-\rho x)\, l(x) \frac{c(x,S,R;\phi_b)^{1-\frac{1}{\sigma}} - 1}{1-\frac{1}{\sigma}} dx$$
$$+\zeta \int_0^S \exp(-\rho x)\, l(x)\, dx - \int_{\underline{W}}^R \exp(-\rho x)\, l(x)\, \nu(x)\, dx,$$

where health index θ_b is absent because mortality decline is not considered in this model. Differentiating $U_b(S,R)$ with respect to S and using (A3) to simplify, we obtain

$$\frac{\partial U_b(S,R)}{\partial S} = \phi_b^{1-\frac{1}{\sigma}} c^n(0,S,R)^{-\frac{1}{\sigma}} \left[h'(S) \int_S^R \exp(-rx)\, l(x)\, dx - \exp(-rS)\, l(S)\, h(S) \right]$$
$$+ \exp(-\rho S)\, l(S)\, \zeta. \tag{B1}$$

Setting $\frac{\partial U_b(S,R)}{\partial S} = 0$ in (B1) and rearranging, we obtain the first-order condition for schooling (33).

Similarly, differentiating with respect to R and using (A4) leads to

$$\frac{\partial U_b(S,R)}{\partial R} = l(R) \left[\phi_b^{1-\frac{1}{\sigma}} c^n(0,S,R)^{-\frac{1}{\sigma}} \exp(-rR)\, h(S) - \exp(-\rho R)\, \nu(R) \right]. \tag{B2}$$

Setting $\frac{\partial U_b(S,R)}{\partial R} = 0$ in (B2), we obtain the same first-order condition for retirement age (8) as in the baseline model in Section 2.

10.1 Comparative statics

The following relationships, obtained from using (5), are useful for subsequent analysis:

$$\frac{1}{c^n(0,S^*,R^*)} \frac{\partial c^n(0,S^*,R^*)}{\partial S} = \frac{-\eta(S^*,R^*;\phi_b)}{\phi_b h(S^*) \int_{S^*}^{R^*} \exp(-rx)\, l(x;\theta_b)\, dx}, \tag{B3}$$

and

$$\frac{1}{c^n(0,S^*,R^*)} \frac{\partial c^n(0,S^*,R^*)}{\partial R} = \frac{\exp(-rR^*)\, l(R^*)}{\int_{S^*}^{R^*} \exp(-rx)\, l(x)\, dx}. \tag{B4}$$

Multiplying both sides of (33) by $\phi_b^{-\frac{1}{\sigma}} c^n \left(0, \tilde{S}(R), R\right)^{-\frac{1}{\sigma}}$ and totally differentiating, we obtain

$$\left\{ \begin{array}{c} \phi_b \exp(-rS^*) l(S^*) h(S^*) \left[\frac{h'(S^*)}{h(S^*)} - \frac{h''(S^*)}{h(S^*)} - \mu(S^*) - r + \frac{\exp(-rS^*) l(S^*)}{\int_{S^*}^{R^*} \exp(-rx) l(x) dx} \right] \\ -\eta(S^*, R^*; \phi_b) \left[\frac{1}{\sigma} \frac{1}{c^n(0,S^*,R^*)} \frac{\partial c^n(0,S^*,R^*)}{\partial S} - \frac{h''(S^*)}{h(S^*)} - \rho - \mu(S^*) + \frac{\exp(-rS^*) l(S^*)}{\int_{S^*}^{R^*} \exp(-rx) l(x) dx} \right] \end{array} \right\} dS^*$$
$$= \left\{ \phi_b h'(S^*) \exp(-rR^*) l(R^*) + \frac{1}{\sigma} \eta(S^*, R^*; \phi_b) \left[\frac{1}{c^n(0, S^*, R^*)} \frac{\partial c^n(0, S^*, R^*)}{\partial R} \right] \right\} dR^*$$
$$- \left[\eta(S^*, R^*; \phi_b) \left(1 - \frac{1}{\sigma}\right) \frac{1}{\phi_b} \right] d\phi_b.$$

Substituting (33), (B3) and (B4) into this equation, we obtain

$$\left\{ \begin{array}{c} \phi_b \exp(-rS^*) l(S^*) h(S^*) \left[-r - \mu(S^*) + \frac{2h'(S^*)}{h(S^*)} - \frac{h''(S^*)}{h(S^*)} \right] \\ +\eta(S^*, R^*; \phi_b) \left[\frac{h''(S^*)}{h(S^*)} + \frac{1}{\sigma} \frac{\eta(S^*, R^*; \phi_b)}{\phi_b h(S^*) \int_{S^*}^{R^*} \exp(-rx) l(x) dx} + \mu(S^*) + \rho \right] \end{array} \right\} dS^*$$
$$= \left\{ \phi_b h'(S^*) \exp(-rR^*) l(R^*) + \frac{\eta(S^*, R^*; \phi_b)}{\sigma} \left[\frac{\exp(-rR^*) l(R^*)}{\int_{S^*}^{R^*} \exp(-rx) l(x) dx} \right] \right\} dR^*$$
$$- \left(1 - \frac{1}{\sigma}\right) \frac{1}{\phi_b} \eta(S^*, R^*; \phi_b) d\phi_b.$$

Hence, we have (35) and (38).

Similarly, totalling differentiating (8) leads to (A11). Using (B3) and (B4), we have (36) and (A16). Note that if $\zeta \geq 0$, then $\eta(S^*, R^*; \phi_b) \geq 0$ and

$$\frac{\partial \tilde{S}(R^*)}{\partial R} > 0$$
$$\frac{\partial \tilde{R}(S^*)}{\partial S} > 0.$$

Hence, Proposition 1 holds for this model as well.

10.2 Proof of Proposition 6

First, note that the total effects of a productivity increase are given by:

$$\frac{\partial S^*}{\partial \phi_b} = \left(\frac{1}{1 - \frac{\partial S^*}{\partial R} \frac{\partial R^*}{\partial S}} \right) \left[\frac{\partial \tilde{S}(R^*; \phi_b)}{\partial \phi_b} + \frac{\partial S^*}{\partial R} \frac{\partial \tilde{R}(S^*; \phi_b)}{\partial \phi_b} \right], \tag{B5}$$

and

$$\frac{\partial R^*}{\partial \phi_b} = \left(\frac{1}{1 - \frac{\partial S^*}{\partial R} \frac{\partial R^*}{\partial S}} \right) \left[\frac{\partial \tilde{R}(S^*; \phi_b)}{\partial \phi_b} + \frac{\partial R^*}{\partial S} \frac{\partial \tilde{S}(R^*; \phi_b)}{\partial \phi_b} \right]. \tag{B6}$$

When $0 < \sigma < 1$ and $\zeta \geq 0$, we have $\frac{\partial \tilde{S}(R^*;\phi_b)}{\partial \phi_b} > 0$ and $\frac{\partial \tilde{R}(S^*;\phi_b)}{\partial \phi_b} < 0$. Then $\frac{\partial S^*}{\partial \phi_b} > 0$ if and only if the direct effect dominates the indirect effect in (B5). That is,

$$-\frac{\partial S^*}{\partial R} \frac{\partial \tilde{R}(S^*; \phi_b)}{\partial \phi_b} < \frac{\partial \tilde{S}(R^*; \phi_b)}{\partial \phi_b}.$$

By substituting (35), (38), (A16), and rearranging, the above condition is equivalent to

$$\zeta > \frac{\exp\left[-\rho\left(R^* - S^*\right)\right] \frac{l(R^*)}{l(S^*)} \frac{h'(S^*)}{h(S^*)}}{r - \rho + \frac{1}{\nu(R^*)} \frac{\partial \nu(R^*)}{\partial x}} \nu\left(R^*\right).$$

Similarly, we have $\frac{\partial R^*}{\partial \phi_b} < 0$ if and only if the direct effect (in absolute value) dominates the indirect effect in (B6). That is,

$$\frac{\partial R^*}{\partial S} \frac{\partial \tilde{S}(R^*; \phi_b)}{\partial \phi_b} < -\frac{\partial \tilde{R}(S^*; \phi_b)}{\partial \phi_b}.$$

By substituting (36), (38), (A16), and rearranging, the above condition is equivalent to

$$\zeta < \frac{\exp\left[(r - \rho)\left(R^* - S^*\right)\right] \left[\frac{2h'(S^*)}{h(S^*)} - \frac{h''(S^*)}{h'(S^*)} - \mu\left(S^*\right) - r\right]}{\frac{h'(S^*)}{h(S^*)} - \frac{h''(S^*)}{h'(S^*)} - \mu\left(S^*\right) - \rho} \nu\left(R^*\right).$$

Hence, Proposition 6 follows by considering different cases of the signs of $\frac{\partial S^*}{\partial \phi_b}$ and $\frac{\partial R^*}{\partial \phi_b}$.

11 Supplementary Appendix C. Sensitivity analysis: The details (Cai, Lau and Yuen)

We report the results of the sensitivity analysis regarding the parameters listed in Table 1.

First, we are interested to know how sensitive our quantitative results are with respect to the values of interest rate and subjective discount rate. In the baseline case, we have followed Bloom et al. (2014) to assume $r = \rho = 3\%$. The first sensitivity analysis we perform is to assume that $r = \rho = 4\%$

(Case 1a), as in Restuccia and Vandenbroucke (2013).[42] When the values of these two parameters are the same, (5) implies that the consumption path is constant. It is also interesting to consider other parameter values such that consumption is changing over time. We consider the case that $r = 5\%$ and $\rho = 2\%$ (Case 1b), as in Barro et al. (1995). The results for these two cases, together with the baseline case, are presented in panel (a) of Figure C1. It can be observed that there are only minor differences in retirement age and years of schooling. Graphically, the profiles remain mostly unaltered, with a mildly inverted U-shaped graph for the retirement age and a concavely increasing one for schooling years. These analyses show that the baseline results are not sensitive to the choice of r and ρ, at least when they vary within a reasonable range.

[Insert Figure C1 here.]

Second, we vary the value of the growth rate of productivity. In the baseline case, we assume that the growth rate of productivity is 1.27%.[43] We would like to know how sensitive our main results are with respect to this parameter. We consider $g = 1\%$ (Case 2a), lower than the baseline value, and $g = 2\%$ (Case 2b), higher than the baseline value. The results are given in panel (b) of Figure C1. When productivity is growing at a slower rate ($g = 1\%$), the retirement age path is above that of the baseline case, with the retirement age for the 2000 cohort at 69.8, which is almost 2 years higher than 67.9 of the baseline case. Intuitively, when the growth rate of productivity is lower, the effect of productivity increase (and thus, wealth) is weaker. Thus, the effect of mortality decline on retirement age becomes relatively more important, resulting in a higher retirement age path. Interestingly, the schooling years profile is not too much different from the baseline case. When productivity is growing faster ($g = 2\%$), the stronger productivity effect drags down the retirement age path substantially, to an extent that the level of retirement age for the 2000 cohort is even lower than the level for the 1900 cohort. The schooling years profile now has an inverted U-shape. For Case 2b, the decrease in retirement age is substantial (4.6 years) from 1945 to 2000, and similarly, the corresponding decrease in schooling years is quite large (1.71 years), offering support regarding the interaction of the effects on these two variables.

[42] Note that in each of the following cases, we re-calibrate the three parameters (γ, λ and δ) according to the procedures described in Section 4.2.

[43] Note that this is the value assumed in Bloom et al. (2014, p. 852), corresponding to the long-run growth rate of real wage in the USA. On the other hand, Restuccia and Vandenbroucke (2013) assume $g = 2\%$.

We next perform analysis with different values of the intertemporal elasticity of substitution (σ). In the baseline case, we focus on the region $\sigma < 1$ (with the income effect dominating the substitution effect), and assume $\sigma = 0.6$. We consider two cases in this region: $\sigma = 0.5$ (Case 3a), which is assumed in Bloom et al. (2014), and $\sigma = 0.7$ (Case 3b). We also consider two other cases that the income effect does not dominate the substitution effect. In Case 3c, we consider $\sigma = 1$ such that the two effects cancel out. Finally, we consider $\sigma = 1.5$ (Case 3d), in which the substitution effect dominates. The results vary significantly when σ changes. This pattern is particularly clear when we switch off the mortality effect and focus only on the effect of productivity increase. The variation of the relative magnitude of the income and substitution effects with respect to σ is confirmed: retirement age and schooling years decrease over time (with increasing wealth) when $\sigma < 1$, but increase when $\sigma > 1$. (See Table C1.) Quantitatively, the magnitude of the changes in retirement age and schooling years with respect to σ is large. In panel (c) of Figure C1, we focus on the comparison of the baseline case with Cases 3a and 3b, regarding the combined effect of changes in mortality decline and productivity increase. When σ decreases slightly from the baseline value to 0.5, the productivity effect becomes stronger, leading to decreases in the retirement age and schooling years in the second half of the twentieth century. The effect is substantial, and the retirement age for the 2000 cohort is even lower than that for the 1900 cohort. Similarly, the productivity effect becomes weaker when σ increases to 0.7, leading to increases in both retirement age and schooling years over almost the entire period. Again the effect is large, with the retirement age for the 2000 cohort increases to 71.2 years, compared with 67.9 years in the baseline case.

[Insert Table C1 here.]

It is seen from the above analysis that a decrease in g shifts up the retirement age path (and the schooling years in recent years), but a decrease in σ produces opposite effects. In panel (d) of Figure C1, we present the results of $g = 0.85\%$ and $\sigma = 0.5$ (Case 4). In this case, decreases in g and σ (from the baseline values) produce almost completely offsetting effects, resulting in retirement age and schooling years paths similar to the baseline case. We also consider results with different assumptions of the age that individuals begin making economic decisions (N) and maximum age in the model (T). The results of $N = 6$ (Case 5) and $T = 105$ (Case 6) are given in panel (d) of Figure C1. We see no major differences in the retirement age and schooling years, relative to the baseline case. Our results are robust to the choice of N and T.

To summarize, the computational results are robust with respect to r, ρ, N and T, at least when they are within some relevant ranges. They are less robust with respect to g and σ individually, but we also find that simultaneous changes in g and σ can lead to retirement age and schooling years paths close to the baseline case.

For each of the above cases, we also perform the decomposition exercises by focusing on one exogenous change (mortality decline or productivity increase) only. As seen in Table C1, we find that in each case, optimal retirement age and schooling years always change monotonically, and these two variables always move in the same direction.[44]

References

[1] Barro, R. J., Mankiw, N. G. and Sala-i-Martin, X. (1995), Capital mobility in neoclassical models of growth. American Economic Review 85, 103-115.

[44]This includes the special case that these two variables remain unchanged in response to productivity increase when $\sigma = 1$.

Figure A1: Labor Force Participation Rate By Cohort

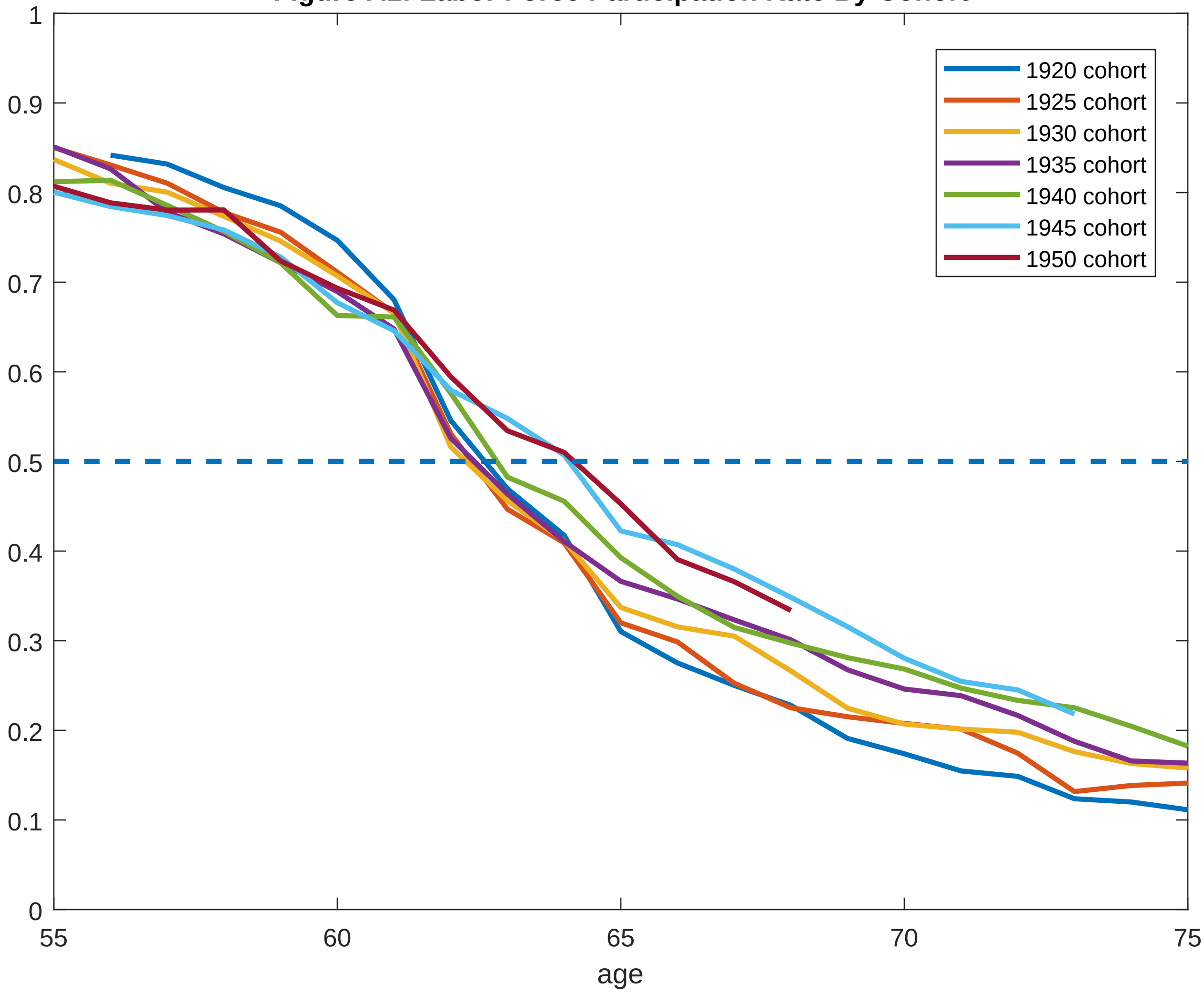
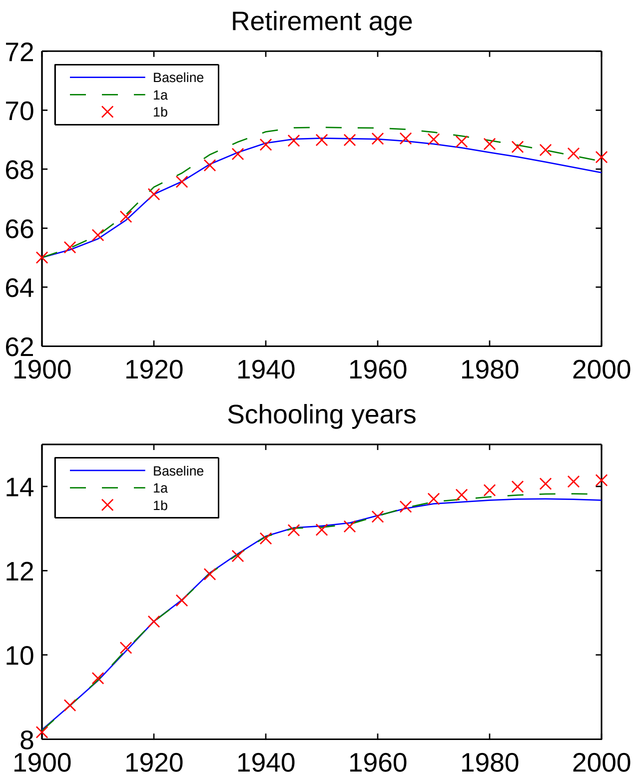
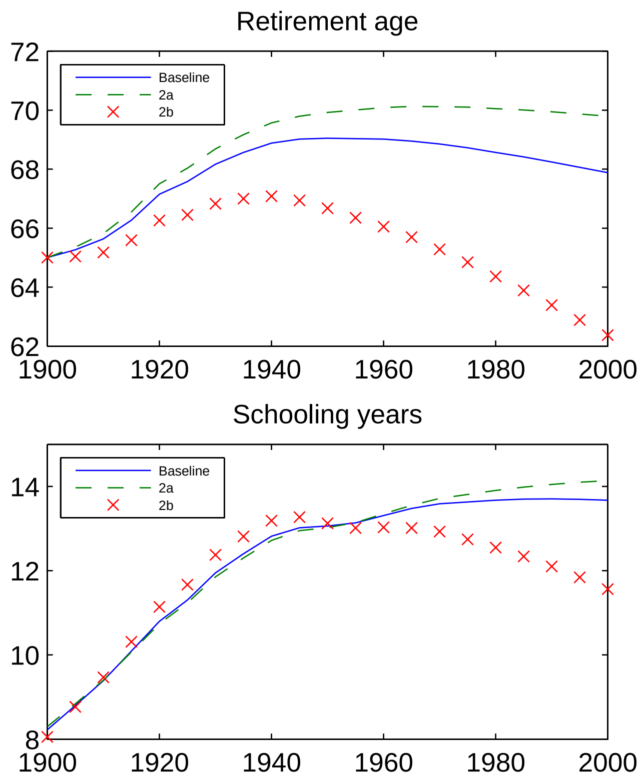


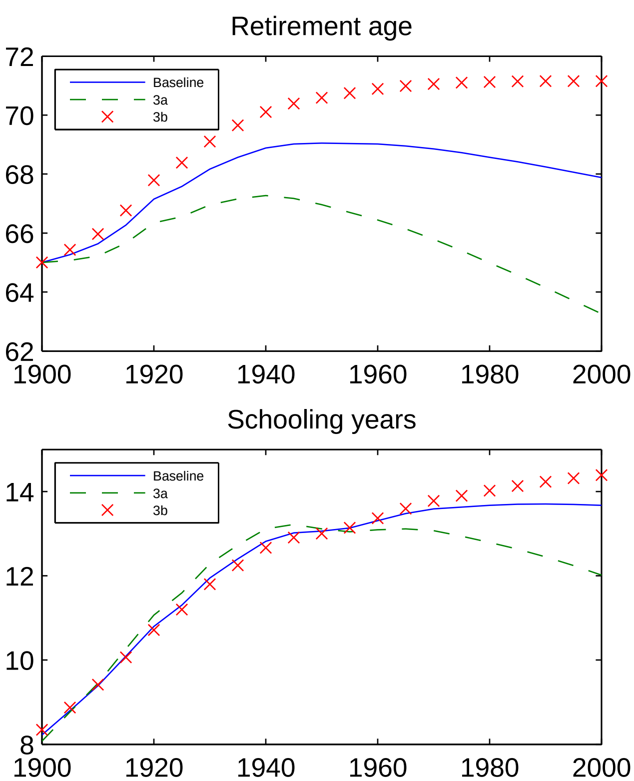
Figure C1. Sensitivity Analysis (With $\sigma < 1$)



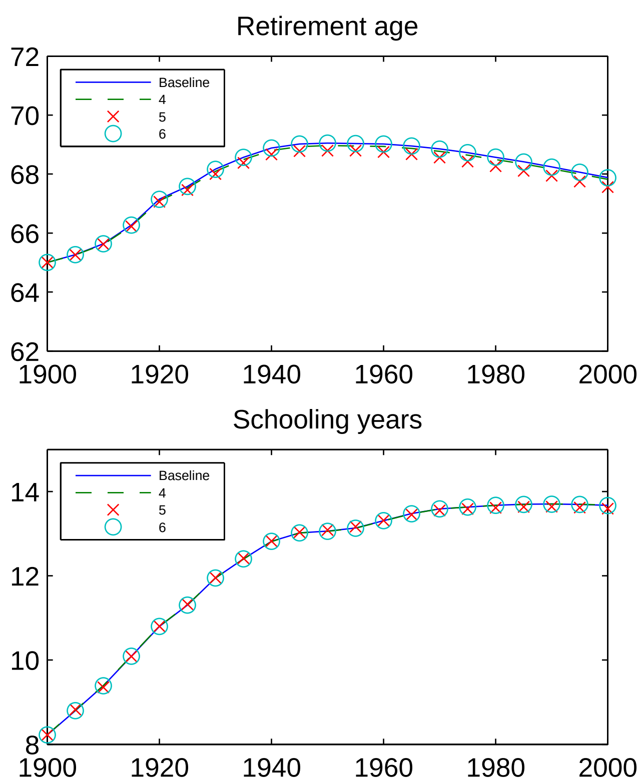
(a) r & ρ



(b) g



(c) σ



(d) Others

Cases	Values	RMSE		Mortality decline & productivity increase		Mortality decline only		Productivity increase only	
				R*+N	S*+N-6	R*+N	S*+N-6	R*+N	S*+N-6
Baseline	/	0.302	1900	65.0	8.22	65.0	8.22	65.0	8.22
			Max. value	69.0	13.71	78.1	18.45	65.0	8.22
			(Year)	(1950)	(1990)	(2000)	(2000)	(1900)	(1900)
			2000	67.9	13.67	78.1	18.45	53.0	5.29
1a	$r = 4\%$; $\rho = 4\%$	0.303	1900	65.0	8.21	65.0	8.21	65.0	8.21
			Max. value	69.4	13.82	78.9	18.03	65.0	8.21
			(Year)	(1950)	(1995)	(2000)	(2000)	(1900)	(1900)
			2000	68.3	13.81	78.9	18.03	52.3	5.47
1b	$r = 5\%$; $\rho = 2\%$	0.305	1900	65.0	8.17	65.0	8.17	65.0	8.17
			Max. value	69.0	14.15	76.4	17.14	65.0	8.17
			(Year)	(1965)	(2000)	(2000)	(2000)	(1900)	(1900)
			2000	68.4	14.15	76.4	17.14	56.0	6.38
2a	$g = 1\%$	0.263	1900	65.0	8.30	65.0	8.30	65.0	8.30
			Max. value	70.1	14.14	77.7	17.46	65.0	8.30
			(Year)	(1965)	(2000)	(2000)	(2000)	(1900)	(1900)
			2000	69.8	14.14	77.7	17.46	53.1	5.56
2b	$g = 2\%$	0.615	1900	65.0	8.05	65.0	8.05	65.0	8.05
			Max. value	67.1	13.27	79.5	21.88	65.0	8.05
			(Year)	(1940)	(1945)	(2000)	(2000)	(1900)	(1900)
			2000	62.4	11.56	79.5	21.88	52.6	4.43
3a	$\sigma = 0.5$	0.541	1900	65.0	8.08	65.0	8.08	65.0	8.08
			Max. value	67.3	13.22	78.9	21.23	65.0	8.08
			(Year)	(1940)	(1945)	(2000)	(2000)	(1900)	(1900)
			2000	63.3	12.02	78.9	21.23	47.6	4.10
3b	$\sigma = 0.7$	0.250	1900	65.0	8.34	65.0	8.34	65.0	8.34
			Max. value	71.2	14.39	77.6	16.90	65.0	8.34
			(Year)	(1995)	(2000)	(2000)	(2000)	(1900)	(1900)
			2000	71.1	14.39	77.6	16.90	57.2	6.58
3c	$\sigma = 1$	0.252	1900	65.0	8.48	65.0	8.48	65.0	8.48
			Max. value	77.1	15.06	77.1	15.06	/	/
			(Year)	(2000)	(2000)	(2000)	(2000)	/	/
			2000	77.1	15.06	77.1	15.06	65.0	8.48
3d	$\sigma = 1.5$	0.263	1900	65.0	8.52	65.0	8.52	65.0	8.52
			Max. value	81.8	15.22	77.0	14.22	71.0	9.37
			(Year)	(2000)	(2000)	(2000)	(2000)	(2000)	(2000)
			2000	81.8	15.22	77.0	14.22	71.0	9.37
4	$g = 0.85\%$; $\sigma = 0.5$	0.302	1900	65.0	8.22	65.0	8.22	65.0	8.22
			Max. value	69.0	13.71	77.9	18.42	65.0	8.22
			(Year)	(1950)	(1990)	(2000)	(2000)	(1900)	(1900)
			2000	67.8	13.68	77.9	18.42	47.8	4.50
5	$N = 6$	0.304	1900	65.0	8.22	65.0	8.22	65.0	8.22
			Max. value	68.8	13.64	77.7	18.04	65.0	8.22
			(Year)	(1950)	(1990)	(2000)	(2000)	(1900)	(1900)
			2000	67.6	13.59	77.7	18.04	52.4	4.25
6	$T = 105$	0.302	1900	65.0	8.22	65.0	8.22	65.0	8.22
			Max. value	69.0	13.70	78.1	18.45	65.0	8.22
			(Year)	(1950)	(1990)	(2000)	(2000)	(1900)	(1900)
			2000	67.9	13.67	78.1	18.45	53.0	5.29

Table C1: Sensitivity Analysis