Online Appendix to "Sharing High Growth Across Generations: Pensions and Demographic Transition in China"

Zheng Song (University of Chicago Booth), Kjetil Storesletten (University of Oslo), Yikai Wang (University of Oslo), and Fabrizio Zilibotti (University of Zurich)

May 2014

A Technical analysis and extensions related to section 2

We now restate and prove Proposition 1, which characterizes the optimal allocation and the associated pension policy. For simplicity, and without loss of generality, we abstract from a hump-shaped age-profile of wages (so the age profile is flat and $\eta_j = 1$), human capital deepening over time (so $\varpi_j = 1$), and mortality before J (so $s_j = 1$ and all agents survive until age J, at which point they die for sure).

Proposition 1 (restated) *Consider an economy where wages grow at the constant rate $\tilde{g}$ during the transition and $g < \tilde{g}$ in steady state, i.e., $g_t = \tilde{g}$ for $t \in \{0, 1, .., T\}$, and $g_t = g$ for $t > T$. The size of the cohort born in period t is denoted μ_t and s_j denotes the unconditional probability of surviving until age j. Agents live for $J \geq 2$ periods and retire after $J_W < J$ periods. The optimal allocation (*first best*) solves the following planning program:*[36]

$$\sum_{t=0}^{\infty} \mu_t \phi^t \sum_{j=0}^{J} \beta^j \left(\log \left(c_{t,j} \right) - \frac{h_{t,j}^{1+\frac{1}{\theta}}}{1+\frac{1}{\theta}} \right), \tag{14}$$

subject to

$$\begin{aligned} \sum_{t=0}^{\infty} \frac{\mu_t}{R^t} \sum_{j=0}^{J} \frac{c_{t,j}}{R^j} &= A_0 + \sum_{t=0}^{\infty} \frac{\mu_t}{R^t} \sum_{j=0}^{J_w} \frac{w_{t+j} h_{t,j}}{R^j} \\ h_{t,j} &= 0 \text{ for all } j > J_W, \end{aligned}$$

where $c_{t,j}$ and $h_{t,j}$ are consumption and labor supply of an individual of age j born at date t. Then, the first-best allocation is given by:

$$\begin{aligned} c_{t,0} &= \lambda^{-1} \left(\phi R \right)^t, \\ c_{t,j} &= c_{t,0} \left(\beta R \right)^j, \text{ for } j \in \{1, 2, ..., J\}, \\ h_{t,j} &= \begin{cases} \left(\frac{w_{t+j}}{c_{t,j}} \right)^{\theta} & \text{for} \quad j \in \{0, 1, ..., J_w\} \\ 0 & \text{for} \quad j \in \{J_w + 1, J_w + 2, ..., J\} \end{cases}. \end{aligned}$$

[36] We ignore for simplicity the generations born before $t = 0$.

where λ is a decreasing function of A_0.
Consider a cohort born at k, and let $W_k = \sum_{j=0}^{J_w} (1-\tau_{t,j}) w_{k+j} \bar{h}_{k,j} R^{-j}$ denote the present value of expected (after-tax) labor income for a representative household, where $\bar{h}_{k,j}$ is the average labor supply of workers of cohort k with experience j. Denote by $b_{k,j}$ the pension paid to a retiree of cohort k and age j. Define cohort $\acute{k}$'s pension replacement rate ζ_k as the present value of pensions as a share of W_k, i.e., $\zeta_k = \left(\sum_{j=J_w+1}^{J} b_{k,j} R^{-j}\right)/W_k$. The first-best allocation can be implemented by a Ramsey sequence of cohort-specific taxes and pension replacement rates. These sequences are characterized as follows:
(1) Taxes are zero in all periods, $\tau_{t,j} = 0$ for all t and j;
(2) The pension replacement sequence satisfies

$$\frac{1+\zeta_{t+1}}{1+\zeta_t} = \left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta} \times F(t), \tag{15}$$

where

$$F(t) = \begin{cases} 1 & \text{if} \quad t \le T - J_w \\ \dfrac{\sum_{j=0}^{T-t} \beta^j \left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot j} + \left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot (T-t)} \sum_{j=T-t+1}^{J_w} \beta^j \left(\frac{1+g}{\beta R}\right)^{(1+\theta)\cdot(j-(T-t))}}{\sum_{j=0}^{T-(t+1)} \beta^j \left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot j} + \left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot (T-(t+1))} \sum_{j=T-t}^{J_w} \beta^j \left(\frac{1+g}{\beta R}\right)^{(1+\theta)\cdot(j-(T-(t+1)))}} & \text{if} \quad t \in \{T-J_w+1, ..., T\} \\ \left(\frac{1+\tilde{g}}{1+g}\right)^{1+\theta} & \text{if} \quad t > T \end{cases} \tag{16}$$

is a non-decreasing function of the birth date t. Finally ζ_0 is given by

$$1+\zeta_0 = \frac{\sum_{j=0}^{J} \beta^j}{\sum_{j=0}^{J_w} \beta^j \left(\frac{w_j}{(\beta R)^j}\right)^{1+\theta}} \times \frac{1}{\lambda^{1+\theta}}. \tag{17}$$

Proof. The characterization of the first-best allocation, (5)–(7) follows from the problem (14)-(4) using standard methods. Consider, next, the Ramsey policy. Since $\tau_{t,j} = 0$, the intratemporal first-order condition implies equation (7). The Euler equation implies that $c_{t,j} = (\beta R)^j c_{t,0}$ as in (6). Next, plugging in (6) and (7) into the budget constraint, and recalling that ζ_t is proportional to the present value of earnings, yields

$$\sum_{j=0}^{J} \frac{(\beta R)^j}{R^j} c_{t,0} = (1+\zeta_t) \sum_{j=0}^{J_w} \frac{w_{t+j}}{R^j} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{\theta} (c_{t,0})^{-\theta}.$$

Solving for $c_{t,0}$ yields

$$(c_{t,0})^{1+\theta} = (1+\zeta_t) \frac{\sum_{j=0}^{J_w} w_{t+j} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{\theta} R^{-j}}{\sum_{j=0}^{J} \beta^j}.$$

Lagging the expression, taking the ratio of $c_{t+1,0}/c_t$, and using (8)-(16), yields

$$\left(\frac{c_{t+1,0}}{c_{t,0}}\right)^{1+\theta} = \left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta} \times F(t) \times \frac{\sum_{j=0}^{J_w} \beta^j \left(\frac{w_{t+1+j}}{(\beta R)^j}\right)^{1+\theta}}{\sum_{j=0}^{J_w} \beta^j \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta}}.$$

We now show that replacing $F(t)$ by its expression in (16) yields $c_{t+1,0}/c_{t,0} = \phi R$, which is consistent with the optimality condition (5).

Suppose, first, that $t > T$. Then, replacing $F(t)$ by its expression in (16) and simplifying terms yields

$$\left(\frac{c_{t+1,0}}{c_{t,0}}\right)^{1+\theta} = \left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta} \times \left(\frac{1+\tilde{g}}{1+g}\right)^{1+\theta} \times (1+g)^{1+\theta} = (\phi R)^{1+\theta},$$

which is consistent with (5).

Suppose, next, that $t \in \{T - J_w + 1, ..., T\}$. Then, proceeding as above,

$$\left(\frac{c_{t+1,0}}{c_{t,0}}\right)^{1+\theta} = \left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta} \times$$

$$\frac{\sum_{j=0}^{T-t}\beta^j\left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot j} + \left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot(T-t)}\sum_{j=T-t+1}^{J_w}\beta^j\left(\frac{1+g}{\beta R}\right)^{(1+\theta)\cdot(j-(T-t))}}{\sum_{j=0}^{T-(t+1)}\beta^j\left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot j} + \left(\frac{1+\tilde{g}}{\beta R}\right)^{(1+\theta)\cdot(T-(t+1))}\sum_{j=T-t}^{J_w}\beta^j\left(\frac{1+g}{\beta R}\right)^{(1+\theta)\cdot(j-(T-(t+1)))}} \times \frac{\sum_{j=0}^{J_w}\beta^j\left(\frac{w_{t+1+j}}{(\beta R)^j}\right)^{1+\theta}}{\sum_{j=0}^{J_w}\beta^j\left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta}}.$$

Then, simplifying terms yields

$$\left(\frac{c_{t+1,0}}{c_{t,0}}\right)^{1+\theta} = \left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta} \times \left(\frac{w_{t+1}}{w_t}\right)^{1+\theta} = (\phi R)^{1+\theta},$$

which is again consistent with (5).

Suppose, finally, that $t \leq T - J_w$. Then, proceeding as above,

$$\left(\frac{c_{t+1,0}}{c_{t,0}}\right)^{1+\theta} = \left(\frac{\phi R}{1+g}\frac{1+g}{1+\tilde{g}}\right)^{1+\theta} \times 1 \times (1+\tilde{g})^{1+\theta} = (\phi R)^{1+\theta},$$

which is again consistent with (5).

Finally, we show that the individual optimization yields $c_{0,0} = \lambda^{-1}$ proving that the entire Ramsey sequence satisfies the first-best condition (5). To this aim, note that

$$c_{0,0}\sum_{j=0}^{J}\beta^j = (1+\zeta_0) \times \sum_{j=0}^{J_w}\beta^j\left(\frac{w_j}{(\beta R)^j}\right)^{1+\theta} c_{0,0}^{-\theta}.$$

Collecting terms and replacing ζ_0 by (17) yields $c_{00} = \lambda^{-1}$. ■

Corollary 1 *Suppose $\phi = (1+g)/R$. Then, the optimal pension benefit sequence is strictly decreasing for all transition generations, $t \leq T$, and constant for all generations born after the end of the transition, $\zeta_t = \zeta_L$ for all $t > T$.*

$$\frac{1+\tilde{\zeta}_{t+1}}{1+\tilde{\zeta}_t} = \left(\frac{1+g}{1+\tilde{g}}\right)^{1+\theta}.$$

Proof. The proof follows from (8)-(16), recalling that $\tilde{g} > g$. ■

Corollary 2 *Consider the environment of Proposition 1. Suppose $\phi = (1+g)/R$, $A_0 \geq 0$, and that the Ramsey implementation is subject to the additional constraint that pensions are non-negative, i.e., $\zeta_t \geq 0$ for all t. The second-best Ramsey allocation has the following characterization: Either the*

constraint $\zeta_t \geq 0$ is never binding (A_0 is very large), and the first best can be implemented by the policy described in Proposition 1, or there exists $\hat{T} < \infty$ such that:
(1) If $t < \hat{T}$, then, up to an increase in λ (implying a lower $c_{0,0}$), the Ramsey policy sequence is identical to the unconstrained policy sequence that implements first best, i.e., taxes are zero in all periods, $\tau_{t,j} = 0$ for all t and j, and pensions are given by (8)—(17);
(2) If $t \geq \hat{T}$, then, $\zeta_t = 0$ and taxes are constant and positive for the cohort, $\tau_{t,j} = \tau_t > 0$.

Proof. The second-best Ramsey problem can be formulated as follows

$$\max_{\left\{\{\tau_{t,j},c_{t,j},h_{t,j}\}_{j=1}^{J_W},\zeta_t\right\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \mu_t \phi^t \sum_{j=0}^{J} \beta^j \left(\log\left(c_{t,j}\right) - \frac{h_{t,j}^{1+\frac{1}{\theta}}}{1+\frac{1}{\theta}} \right), \tag{18}$$

subject to the non-negative-pension constraint $\zeta_t \geq 0$, to the resource constraint

$$\sum_{t=0}^{\infty} \frac{\mu_t}{R^t} \sum_{j=0}^{J} \frac{c_{t,j}}{R^j} = A_0 + \sum_{t=0}^{\infty} \frac{\mu_t}{R^t} \sum_{j=0}^{J_w} \frac{w_{t+j} h_{t,j}}{R^j},$$

and to the constraint that households optimize given the fiscal policy sequence $\left\{\{\tau_{t,j}\}_{j=1}^{J_W}, \zeta_t\right\}_{t=0}^{\infty}$. Household optimization implies

$$\begin{aligned}
c_{t,j} &= c_{t,0}\left(\beta R\right)^j, \\
h_{t,j} &= \begin{cases} \left(1-\tau_{t,j}\right)^{\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{\theta} \left(c_{t,0}\right)^{-\theta} & \text{for} \quad j \in \{0,1,...,J_w\} \\ 0 & \text{for} \quad j \in \{J_w+1, J_w+2, ..., J\} \end{cases}, \\
\sum_{j=0}^{J} \beta^j c_{t,0} &= \left(1+\zeta_t\right) \sum_{j=0}^{J_W} \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-\theta}.
\end{aligned}$$

We use the household's optimal decisions substitute out the labor supply from the planner constraints. Moreover, the Euler equation of consumers allows us to express the problem as a function of $c_{t,0}$ rather that of the entire consumption sequence of each cohort. This leaves only the resource constraint and the non-negative pension constraint, expressed in terms of tax rates and the sequence $\{c_{t,0}\}_{t=0}^{\infty}$. Using these constraints, we can express the second-best problem in terms of the following Lagrangian:

$$\begin{aligned}
L = \sum_{t=0}^{\infty} \phi^t &\left(\begin{array}{c} \sum_{j=0}^{J} \beta^j \log\left(c_{t,0}\left(\beta R\right)^j\right) - \sum_{j=0}^{J_w} \beta^j \frac{(1-\tau_{t,j})^{1+\theta}\left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} (c_{t,0})^{-(1+\theta)}}{1+\frac{1}{\theta}} \\ +\xi_t \left(\sum_{j=0}^{J} \beta^j c_{t,0} - \sum_{j=0}^{J_w} \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-\theta} \right) \end{array} \right) + \\
\lambda &\left(\sum_{t=0}^{\infty} \frac{\mu_t}{R^t} \sum_{j=0}^{J_w} \beta^j \left(1-\tau_{t,j}\right)^{\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-\theta} - \sum_{t=0}^{\infty} \frac{\mu_t}{R^t} \sum_{j=0}^{J} \beta^j c_{t,0} \right)
\end{aligned}$$

where $\xi_t \geq 0$ is the Lagrangian multiplier associated with the constraint $\zeta_t \geq 0$, and $\lambda > 0$ is the Lagrange multiplier associated with the resource constraint.

The FOCs with respect to $c_{t,0}$ and $\tau_{t,j}$ yield, respectively:

$$\frac{\partial L}{\partial c_{t,0}} = \phi^t \left(\begin{array}{c} \sum_j \beta^j \frac{1}{c_{t,0}} + \sum^{J_w} \beta^j \theta \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} c_{t,0}^{-(2+\theta)} + \\ \xi_t \left(\sum \beta^j + \theta \sum^{J_w} \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-(1+\theta)} \right) \end{array} \right) -$$

$$\lambda \frac{\mu_t}{R^t} \left(\sum_j^{J_w} \theta \beta^j \left(1-\tau_{t,j}\right)^{\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-(1+\theta)} + \sum_j \beta^j \right) = 0, \tag{19}$$

$$\frac{\partial L}{\partial \tau_{t,j}} = \phi^t \left(\beta^j \theta \left(1-\tau_{t,j}\right)^{\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-(1+\theta)} + \xi_t \left((1+\theta) \beta^j \left(1-\tau_{t,j}\right)^{\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-\theta} \right) \right) -$$

$$\lambda \left(\frac{1}{R^t} \theta \beta^j \left(1-\tau_{t,j}\right)^{\theta-1} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-\theta} \right) = 0. \tag{20}$$

Consider, next, two separate cases:

1. $\xi_t = 0$, i.e., the constraint $\zeta_t \geq 0$ is slack. In this case, the problem is identical to the implementation of the first best in Proposition 1, up to an increase in the value of λ. In particular, letting $\tau_{t,j} = \tau_t = 0$ implies that $c_{t,0} = \lambda^{-1} (\phi R)^t$ (see equation (5)) and $h_{t,j} = \left(\frac{w_{t+j}}{c_{t,j}}\right)^{\theta}$, for $j \in \{0, 1, ..., J_w\}$ (see equation (7)). Since λ is larger, consumption is lower and labor supply is higher. Moreover, if the constraint is slack at $t > 0$, it must also be slack for all $k \leq t$. To see why, note that the pension sequence ζ_t given by (8)-(17) is non-increasing, so $\zeta_t > 0$ (and, thus, $\xi_t = 0$) implies $\zeta_k > 0$ (thus, again, $\xi_k = 0$) for all $k < t$.

2. $\xi_t > 0$, i.e., the constraint that pensions cannot be negative is binding. Thus, $\zeta_t = 0$ and the individual budget constraint yields:

$$\sum \beta^j c_{t,0} = \sum^{J_w} \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} \left(c_{t,0}\right)^{-\theta} \tag{21}$$

   Combining (19)-(20) yields:

$$\phi^t \left(\sum_j \beta^j \frac{1}{c_{t,0}} + \xi_t \left(\sum \beta^j - \sum^{J_w} \beta^j \left(1-\tau_{t,j}\right)^{1+\theta} \left(\frac{w_{t+j}}{(\beta R)^j}\right)^{1+\theta} c_{t,0}^{-\theta-1} \right) \right) - \lambda \left(\frac{1}{R^t} \sum_j \beta^j \right) = 0.$$

   Substituting into this expression the budget constraint, (21), implies:

$$\mu_t \phi^t \sum_j \beta^j \frac{1}{c_{t,0}} - \lambda \frac{\mu_t}{R^t} \sum_j \beta^j = 0 \Rightarrow$$
$$c_{t,0} = \lambda^{-1} (\phi R)^t .$$

   Finally, substituting this condition into (20), and solving for τ_t, after rearranging terms, yields:

$$\tau_{t,j} = \tau_t = \frac{\xi_t (1+\theta) c_{t,0}}{\theta + \xi_t (1+\theta) c_{t,0}} > 0,$$

where the inequality follows from the assumption that $\xi_t > 0$. Finally, we can prove by *reductio ad absurdum* that if $\xi_t > 0$, then $\xi_k > 0$ for all $k > t$. Suppose not, and $\exists k > t$ such that $\zeta_k > 0$. Then, for the argument provided in the proof of part 1 of this proposition, the non-negativity constraint should be slack for all $k' < k$, including $k' = t$, raising a contradiction.

Finally, note that either the constraint $\zeta_t \geq 0$ is slack for all T, and then the first best can be implemented, or there exist a T such that the constraint is slack for all $t < T$ and is binding for all $t \geq T$.

■

B Estimation method of the rural-urban migration

In this appendix, we present the estimation method of the rural-urban migration. $n_{2000}^{h,i,j}$ and $n_{2005}^{h,i,j}$ represent the population of group (h,i,j) in the 2000 census and 2005 survey, respectively, where $h \in \{u, r\}$, $i \in \{f, m\}$, and $j \in \{0, 1, \cdots, 100\}$ stand for residential status (u for urban and r for rural residents), gender (f for females and m for males), and age, respectively. $\hat{n}_{2005}^{h,i,j}$ represents the projected "natural" population in 2005. Denote $m^{i,j}$ the net flow of the rural-urban migration from 2000 to 2005. We observe $n_{2000}^{h,i,j}$ and $n_{2005}^{h,i,j}$ from the 2000 census and 2005 survey. Moreover, we can use $n_{2000}^{h,i,j}$, together with the observed birth and mortality rates, to project $\hat{n}_{2005}^{h,i,j}$; i.e., the "natural" population in 2005. In other words, both $n_{2005}^{h,i,j}$ and $\hat{n}_{2005}^{h,i,j}$ in (22) and (23) are observable. The 2005 urban and rural population gender-age structure can thus be composed into three parts:

$$n_{2005}^{u,i,j} = \hat{n}_{2005}^{u,i,j} + m^{i,j} + \varepsilon^{u,i,j}, \tag{22}$$

$$n_{2005}^{r,i,j} = \hat{n}_{2005}^{r,i,j} - m^{i,j} + \varepsilon^{r,i,j}, \tag{23}$$

where $\varepsilon^{h,i,j}$ captures measurement errors in the census and survey.

In the ideal case with no measurement errors, either (22) or (23) can back out $m^{i,j}$. The measurement error on the total population, $\sum_{h,i,j} \varepsilon^{h,i,j}$, is small. When $\sum_{h,i,j} \varepsilon^{h,i,j} = 0$, (22) and (23) imply that the projected total population, $\sum_{h,i,j} \hat{n}_{2005}^{h,i,j}$, would be equal to the total population in the 2005 survey, $\sum_{h,i,j} n_{2005}^{h,i,j}$. The difference between $\sum_{h,i,j} \hat{n}_{2005}^{h,i,j}$ and $\sum_{h,i,j} n_{2005}^{h,i,j}$ is less than 1%. [37] However, the match of the sum of the rural and urban population in each gender-age group is less perfect. Figure A-1 plots the projected 2005 "natural" population gender-age structure (solid line) and the 2005 survey data (dotted line). The discrepancy between the two lines reveals the measurement error on the population of each gender-age group, $\varepsilon^{i,j}$, where

$$\varepsilon^{i,j} \equiv \sum_h \varepsilon^{h,i,j} = \sum_h \left(n_{2005}^{h,i,j} - \hat{n}_{2005}^{h,i,j} \right). \tag{24}$$

Figure I suggests $\varepsilon^{i,j}$ to be quantitatively important.[38] To understand how $\varepsilon^{i,j}$ affects the estimated migration gender-age structure, let us assume the measurement error on urban population, $\varepsilon^{u,i,j}$, is proportional to $\varepsilon^{i,j}$:

$$\varepsilon^{u,i,j} = \pi \cdot \varepsilon^{i,j}, \tag{25}$$

[37] Despite the small discrepancy, to avoid biased estimates, we adjust $n_{2000}^{h,i,j}$ by a scale of κ, where κ is calibrated to 1.0073 by matching the projected 2005 total population with the 2005 survey data. $\kappa = 1.0073$ suggests the discrepancy of the total population to be less than 1%.

[38] If all the discrepancies are due to sampling errors in the 2005 survey, the comparison between the two lines in figure I indicates that a major drawback of the 2005 survey is the undercounted young labor force (age 16 to 40). Our calculation suggests 66 million young labor force (11% of total young labor force) missing from the 2005 survey.

where $\pi \in [0,1]$. It follows that the measurement error for the rural population is

$$\varepsilon^{r,i,j} = (1-\pi)\cdot\varepsilon^{i,j}. \tag{26}$$

Rearranging (22) gives the net flow of migration:

$$\begin{aligned}\sum_i\sum_j m^{i,j} &= \sum_i\sum_j\left(n_{2005}^{u,i,j} - \hat{n}_{2005}^{u,i,j}\right) - \pi\sum_i\sum_j\varepsilon^{i,j} \\ &= \sum_i\sum_j\left(n_{2005}^{u,i,j} - \hat{n}_{2005}^{u,i,j}\right) - \pi\sum_h\sum_i\sum_j\left(n_{2005}^{h,i,j} - \hat{n}_{2005}^{h,i,j}\right).\end{aligned} \tag{27}$$

The second equality comes from (24). Let us consider two extreme cases of π. When $\pi = 1$, (27) can be written as

$$\sum_i\sum_j m^{i,j} = \underbrace{\sum_i\sum_j \hat{n}_{2005}^{r,i,j}}_{\text{projected ``natural'' rural population}} - \underbrace{\sum_i\sum_j n_{2005}^{r,i,j}}_{\text{rural population in the survey data}}.$$

When $\pi = 0$, (27) reduces to

$$\sum_i\sum_j m^{i,j} = \underbrace{\sum_i\sum_j n_{2005}^{u,i,j}}_{\text{urban population in the survey data}} - \underbrace{\sum_i\sum_j \hat{n}_{2005}^{u,i,j}}_{\text{projected ``natural'' urban population}}.$$

Therefore, the choice of π boils down to the choice of using rural or urban population to back out migration. It has been widely acknowledged that the urban population survey tends to underestimate the "floating population," that is, rural migrants without *hukou* - the local household registration status (e.g., Liang and Ma 2004). So, we set $\pi = 1$. We will discuss the results using $\pi = 0.5$.

It is instructive to compare the actual migration structure with our estimates. The migration *flow* structure is hard to obtain. However, the migration *stock* structure may shed some light on the *flow* structure. The age structure of migrants in the 2000 census is presented in the second row of Table A-1, which has a high concentration in the 15-29 age group. The same pattern also appears in our estimates under $\pi = 1$ (the third row). $\pi = 0.5$ results in a much more dispersed age structure (the fourth row). This provides a justification for using $\pi = 1$.[39]

Table A-1 Age distribution of migration (percent)

age	<15	15-29	30-44	45-59	60+
migration stock in the 2000 census	9.0	60.5	22.2	5.8	2.5
estimated flow from 2000 to 2005 with $\pi = 1$	25.8	64.8	26.5	-8.6	-8.6
estimated flow from 2000 to 2005 with $\pi = 0.5$	17.8	39.5	27.7	8.9	6.1

Note: The age structure in the 2000 census is from Liang and Ma (2004).

[39] One caveat is that the data from the 2000 census are the age structure of narrowly defined migrants, whereas our estimate is on broadly defined migrants including urbanized population.

Finally, we compute $mr^{i,j}$, the age–gender specific migration rate defined as the average annual net flow of migration per hundred rural population with gender i and age j. We assume that $mr^{i,j}$ is time-invariant and the mortality rates for migrants are the same as those for rural residents. Then, $m^{i,j}$ can be written as follows:

$$\begin{aligned}
m^{i,j} = & \underbrace{mr^{i,j-5} n_{2000}^{r,i,j-5}}_{\text{migration of 2000}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2000 to 2005}} \\
& + \underbrace{mr^{i,j-4}\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2001}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}} \\
& + \underbrace{mr^{i,j-3}\left(1-mr^{r,j-4}\right)\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2002}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}} \\
& + \underbrace{mr^{i,j-2}\left(1-mr^{r,j-3}\right)\left(1-mr^{r,j-4}\right)\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2003}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}} \\
& + \underbrace{mr^{i,j-1}\left(1-mr^{r,j-2}\right)\cdots\left(1-mr^{r,j-5}\right) n_{2000}^{r,i,j-5}}_{\text{migration of 2004}} \underbrace{\left(1-d_{2000}^{r,i,j-1}\right)\cdots\left(1-d_{2000}^{r,i,j-5}\right)}_{\text{survival rate from 2001 to 2005}}.
\end{aligned}$$

Here, $n_{2000}^{r,i,j-5}$ is the mortality rate of rural residents in the 2000 census. In other words, $m^{i,j}$ measures an accumulated migration stock from 2000 to 2005. The above equation allows us to back out the age-gender specific migration rates. Specifically, for $j = J + 5$:

$$m^{i,J+5} = \underbrace{mr^{i,J}\hat{n}_{2000}^{r,i,J}}_{\text{migration of 2000}} \underbrace{\left(1-d_{2000}^{r,i,J+4}\right)\cdots\left(1-d_{2000}^{r,i,J+4}\right)}_{\text{survival rate from 2000 to 2005}}$$

$$\Rightarrow mr^{i,J} = \frac{m^{i,J+5}}{n_{2000}^{r,i,J}\left(1-d_{2000}^{r,i,J+4}\right)\cdots\left(1-d_{2000}^{r,i,J}\right)}.$$

For $j = J + 4$:

$$\begin{aligned}
m^{i,J+4} = & \underbrace{mr^{i,J-1}\hat{n}_{2000}^{r,i,J-1}}_{\text{migration of 2000}} \underbrace{\left(1-d_{2000}^{r,i,J+3}\right)\cdots\left(1-d_{2000}^{r,i,J-1}\right)}_{\text{survival rate from 2000 to 2005}} \\
& + \underbrace{mr^{i,J}\left(1-mr^{r,J-1}\right) n_{2000}^{r,i,J-1}}_{\text{migration of 2001}} \underbrace{\left(1-d_{2000}^{r,i,J+3}\right)\cdots\left(1-d_{2000}^{r,i,J-1}\right)}_{\text{survival rate from 2000 to 2005}}
\end{aligned}$$

$$\Rightarrow mr^{i,J-1} = \frac{m^{i,J+4} - mr^{i,J} n_{2000}^{r,i,J-1}\left(1-d_{2000}^{r,i,J+3}\right)\cdots\left(1-d_{2000}^{r,i,J-1}\right)}{\left(1-mr^{i,J}\right) n_{2000}^{r,i,J-1}\left(1-d_{2000}^{r,i,J+3}\right)\cdots\left(1-d_{2000}^{r,i,J-1}\right)}.$$

All the migration rates can thus be solved in a recursive way.

C Details on the Chinese pension system

This appendix provides a description of the basic features of the Chinese pension system. We start with the urban pension system, and then provide a brief description of the rural pension system, which has been introduced experimentally in 2009.

C.1 The urban pension system

The pre-1997 urban pension system was primarily based on state and urban collective enterprises in a centrally planned economy. Retirees received pensions from their employers, with replacement rates that could be as high as 80 percent (see, e.g., Sin, 2005; OECD, 2007). The coverage was low in the work-unit-based system, though. Many non-state-owned enterprises had no pension scheme for their employees. The coverage rate, measured by the ratio of the number of workers covered by the system to the urban employment, was merely 44% in 1992 according to *China Statistical Yearbook* 2009. The rapid expansion of the private sector caused a growing disproportion between the number of contributors and beneficiaries and, therefore, a severe financial distress for the old system (Zhao and Xu, 2002). To deal with the issue, the government initiated a transition from the traditional system to a public pension system in the early 1990s. The new system was implemented nationwide after the State Council issued "A Decision on Establishing a Unified Basic Pension System for Enterprise Workers (Document 26)" in 1997.

The reformed system mainly consists of two pillars. The first pillar, funded by 17% wage taxes paid by enterprises, guarantees a minimum replacement rate of 20% of local average wage for retirees with a minimum of 15 years of contribution. It is worth emphasizing that the pension fund is managed by local governments (previously at the city level and now at the provincial level). The second pillar provides pensions from individual accounts financed by a contribution of 3% and 8% social security tax paid by enterprises and workers, respectively. There is a third pillar adding to individual accounts through voluntary contribution. The return of individual accounts is adjusted according to bank deposit rates. The system also defines monthly pension benefits from individual accounts equaling the account balance at retirement divided by 120.

More recently, a new reform was implemented after the State Council issued "A Decision on Improving the Basic Pension System for Enterprise Workers (Document 38)" in 2005. The reform adjusted the proportion of taxes paid by enterprises and individuals and the proportion of contribution for individual accounts. Individual accounts are now funded by the social security 8% tax paid by workers only.[40] The first and second pillars deliver target replacement rates of 35% and 24.2%, respectively (Hu, Stewart and Yermo, 2007).

Two features of the current urban pension system is particularly important for our modeling. First, the pension reform was cohort-specific. There were three types of cohorts when the pension reform took place: cohorts entering the labor market after 1997 (*Xinren*), cohorts retiring before 1997 (*Laoren*) and cohorts in-between (*Zhongren*). Pension contributions and benefits of *Xinren* are entirely determined by the new rule. According to Item 5 in Document 26, the government commits to pay *Laoren* the same pension benefits as those in the old system subject to an annual adjustment by wage growth and inflation. For *Zhongren*, their contributions follow the new rule, while their benefits consist of two components: (1) pensions from the new system identical to those for *Xinren,* and (2) a transitional pension that smooths the pension gap between *Laoren* and *Xinren.* For simplicity, we ignore *Zhongren* and take pensioners retiring before and after 1997 as *Laoren* and *Xinren*, respectively. Following Sin (2005), we set the replacement rate for *Laoren* and *Xinren* to 78% and 60%, respectively.

Second, like private savings, pension funds are allowed to invest in domestic stock markets. The

[40] The reform also adjusted the pension benefits. The replacement rate of an individual is now determined by years of contribution: A one year contribution increases the replacement rate of a wage index averaged from local and individual wages by one percentage point. However, the article did not state explicitly how to compute the wage index.

In practice, the index appears to differ across provinces. For instance, the increase in the average pension benefits per retiree in 2011 was almost the same across Beijing and GanSu (the monthly increase was RMB210 in Beijing and RMB196 in GanSu), though the average wage in Beijing is more than two times as high as that in GanSu and the gap has been rather stable over time.

baseline model assumes the annual rate of returns to pension funds to be 2.5%, which is identical to the rate of returns to private savings. According to the latest information released by the National Council for Social Security Fund, the average share of pension funds invested in stock markets was 19.22% in 2003-2011.[41] If 20% of pension funds have access to the market with an annual return of 6% and the rest of the funds gain an annual return of 1.75% as the one-year bank deposits, the average annual rate of returns would be equal to 2.6%, almost equal to 2.5% set in the baseline model.

It is also worth emphasizing that the actual urban pension system deviates from statutory regulations in a number of ways and our model has been adapted to capture some major discrepancies. First, the individual accounts are basically empty. Despite the recent efforts made by the central government to fund these empty individual accounts, there are only 270 billion RMB in all individual accounts of around 200 million workers participating in the urban pension system.[42] Therefore, we take the individual accounts as notional and ignore any distinction between the different pension pillars throughout the paper. In addition, we assume that 40% of pension benefits are indexed to wage growth. The level of indexation is set on the conservative side since the actual level is between 40% and 60% (see Sin, 2005).

Second, the statutory contribution rate including both basic pensions and individual accounts is 28%, of which 20% should be paid by firms and 8% should be paid by workers (see the above discussion on Document 26 and 38). However, there is evidence that a significant share of the contributions is evaded. For instance, in the annual National Industrial Survey – which includes all state-owned manufacturing enterprises and all private manufacturing enterprises with revenue above 5 million RMB – the average pension contributions paid by firms in 2004-2007 amounts to 11% of the average wages, 9 percentage points below the statutory rate.[43] Most evasion comes from privately owned firms, whose contribution rate is a merely 7%.

The actual contribution rate is substantially lower than the statutory rate even for workers participating in the system. A simple way of estimating the actual contribution rate conditional on participation is to look at the following ratio:

$$\begin{aligned} BR &\equiv \frac{\text{per retiree pension benefits}}{\text{per worker pension contributions}} \\ &\equiv \frac{\frac{\text{total pension fund expenditure}}{\text{total retirees covered by the system}}}{\frac{\text{total pension fund revenue - government subsidy}}{\text{total workers covered by the system}}}. \end{aligned}$$

If the replacement rate is indeed 60%, a contribution rate of 28% would imply BR to be 2.1. However, we find that the average BR in the data from 1997 to 2009 is 3.1, much higher than 2.1 by the statutory contribution rate. With a targeted replacement rate of 60%, the ratio of 3.1 would imply an actual contribution rate of 19.4%.[44] So, we set the actual contributkion rate to 20% in the paper.

Finally, although the coverage rate of the urban pension system is still relatively low, it has grown from about 40% in 1998 to 57% in 2009, where we measure the coverage rate by the number of

[41] Source: http://www.ssf.gov.cn/xw/xw_gl/201205/t20120509_4619.html.

[42] The number of 270 billion RMB comes from the information released by the Ministry of Human Resources and Social Security in the 2012 National People's Congress. Source: http://lianghui.people.com.cn/2012npc/GB/239293/17320248.html

[43] In addition, with a labor income share less than 20%, wages appear to be severely underreported.

[44] All the data are available from *China Statistical Yearbook*, except for the government subsidies. Fortunately, since 2010, the Ministry of Finance has started to publicize detailed expenditure items. The government subsidy to the pension fund amounted to 191 billion RMB in 2010, accounting for 21% of the total government social security and employment expenditure. We then use 21% to back out annual government subsidy to pension funds from annual total government social security and employment expenditure, which is available from *China Statistical Yearbook*.

employees participating in the pension system as a share of the number of urban employees.[45] There is a concern that the rapidly growing size of migrant workers might lead to downward-biased urban employment. Our estimation suggests that the urban population (including migrants) between age 22 and 60 increases by 130 million from 2000 to 2009. A labor participation rate of 80% would imply an increase of 104 million in the urban employment, whereas the increase by the official statistics is 79 million. Restoring the 25 million "missing" urban employment would lower the pension coverage rate from 57% to 53% in 2009. Our baseline model assumes a constant coverage rate of 60%, reflecting a trade-off between the low coverage of the current pension system and the potentially higher one in the future.

C.2 The rural pension system

The pre-2009 rural pension program had two features. First, it was "fully-funded" in the sense that pension benefits were essentially determined by contributions to individual accounts. Second, the coverage rate was low since farmers did not have incentives to participate. A pilot pension program was launched for rural residents in 2009. Like those in the urban pension system, the new rural program entails two benefit components. The first one is referred to as basic pension, mainly financed by the Ministry of Finance, and the second one is referred to as pension from individual account. If a migrant worker who joined the urban pension system returns to her home town, the money accumulated in her account will be transferred to her new account in the rural pension program. The program was first implemented in 10% of cities and counties on a trial basis. The government targeted to extend the program to 60% of cities and counties in 2011. Many of the cities and counties report high participation rates (above 80%). This is not surprising since the program is heavily subsidized (see below for more details).

We then lay out some basic features of the new program upon which the model is based. According to "Instructions on New Rural Pension Experiments" issued by the State Council in 2009, the new program pays a basic pension of RMB55 ($8.7) per month. Suppose that the rural wage equals the rural per capita annual net income, which was RMB5153 in 2009 (*China Statistical Yearbook* 2010). Then, the basic pension would correspond to a replacement rate of 12.8%. Notice that provinces are allowed to choose more generous rural pensions. So, the replacement rate of 9% should be viewed as a lower bound.[46] In practice, some places set a much higher basic pension standard. Beijing, for instance, increased the level to RMB280. The monthly basic pension in Shanghai has a range from RMB150 to RMB300, dependent of age, years of contribution and status in the old pension program.[47] Since the rural per capita net income in Beijing and Shanghai is about 1.4 times higher than the average level in China, a monthly pension of RMB280 would imply a replacement rate of 27.2%. In the quantitative exercise, we then set the replacement rate to 20% to match the average of the basic level of 12.8% and the high level of 27.2%.[48] On the contribution side, rural residents in

[45] Both numbers are obtained from *China Statistical Yearbook* 2010.

[46] The Ministry of Human Resources and Social Security has made it clear that there is no upper bound for basic pension and local governments may increase basic pension according to their public financing capacity.

[47] See "Detailed Rules for the Implementation of Beijing Urban-Rural Household Pension Plans," Beijing Municipal Labor and Social Security Bureau, 2009 and "Implementation Guidelines of State Council's Instructions on New Rural Pension Experiments," Shanghai Municipal Government, 2010.

[48] All rural residents above age 60 are entitled to a basic pension. The only condition is that children of a basic pension recipient, if any, should participate in the program. In practice, basic pension might be contingent on years of contribution and status in the old pension program (see the above example from Shanghai).

In addition, an official policy report from the Ministry of Human Resources and Social Security (http://news.qq.com/a/20090806/000974.htm) states that by the rule of the new system, a rural worker paying an annual contribution rate of 4% for 15 years should be entitled to pension benefits with a replacement rate of 25%.

Planner's discount rate	Delayed until 2050 high	Delayed until 2050 low	Delayed until 2100 high	Delayed until 2100 low	Fully Funded high	Fully Funded low	PAYGO high	PAYGO low
Baseline (ret. age at 60)	6.4%	0.9%	8.9%	0.6%	-3.3%	0.2%	12.4%	1.6%
Retirement age at 57	9.9%	1.3%	13.4%	0.7%	-3.2%	0.3%	11.8%	1.8%

Table 2: The table summarizes the welfare effects (measured in terms of compensated variation in consumption for the high- and low-discount rate planners, respectively) under the alternative assumption about retirement age compared to the results under the baseline calibration.

principle should contribute 4% to 8% of the local average income per capita in the previous year. We take the mean and set a contribution rate of 6%.[49]

The current pension program heavily relies on government subsidy. *China Statistical Yearbook* 2010 reports a rural population of 712.88 million. According to the 2005 one-percent population survey, 13.7% of rural population is above age 60. These two numbers give a rural population of 97.66 million who are entitled to a basic pension. This, in turn, implies an annual government subsidy of 64.46 billion RMB, if monthly basic pension is set to RMB55. The central government revenue is 3592 billion RMB in 2009. So, a full-coverage rural pension program in 2009 would require subsidy as a share of the central government revenue of 1.8% and a share of GDP of 0.19%.

D A retirement age of 57

In this section we report the results under an alternative calibration which assumes that the retirement age is 57 instead of 60, as in the benchmark calibration. 57 is an average of the current statutory retirement age for men (60) and women (55). We have opted for using a retirement age of 60 as a benchmark because we expect that the pension age is likely to increase as the health of the population improves with economic progress.

The fiscal imbalance of the system is now larger than under the baseline calibration. Consequently, a larger reduction in replacement rate is required to balance the system. Under the draconian reform the replacement rate now is 32.5%, compared to 39.1% in the baseline calibration. When the reform is delayed until 2050 (2100), the required replacement rate fall to 28.0% (18.3%). The welfare results are reported in Table 2. As is evident from the table, the main conclusions hold up, being even stronger in the sense that delaying the reform would be even more beneficial than under the baseline calibration.

E A dynamic general equilibrium model

In this section, we construct a dynamic general equilibrium model that delivers the wage and interest rate sequence assumed in the baseline model of section 2 as an equilibrium outcome. These prices are sufficient to compute the optimal decisions of workers and retirees (consumption and labor supply) as well as the sequence of budget constraints faced by the government. The model is builds on SSZ, augmented with the demographic model of section 3.1 and the pension system of section 2.

The production sector: The urban production sector consists of two types of firms: (i) *financially integrated* (F) firms, modeled as standard neoclassical firms; and (ii) *entrepreneurial* (E) firms,

[49] Rural residents are allowed to contribute more. But the contribution rate cannot exceed 15% for each person. Moreover, to be eligible for pension from individual account, a rural resident must contribute to the program for at least 15 years. The monthly pension benefit is set equal to the accumulated money in individual account divided by 139 (the same rule applied to the urban pension program).

owned by (old) entrepreneurs, who are residual claimants on the profits. Entrepreneurs delegate the management of their firms to specialized agents called *managers*. E firms can run more productive technologies than F firms (see Song *et al.*, 2011 for the microfoundation of this assumption). However, they are subject to credit constraints that limit their growth. In contrast, the less productive F firms are unconstrained. Motivated by the empirical evidence (see Song *et al.*, 2011) that private firms are more productive and more heavily financially constrained than state-owned enterprises (SOE) in China, we think of F firms as SOE and E firms as privately owned firms.

The technology of F and E firms are described, respectively, by the following production functions:

$$Y_F = K_F^{\alpha} (AN_F)^{1-\alpha}, \qquad Y_E = K_E^{\alpha} (\chi AN_E)^{1-\alpha},$$

where Y is output and K and N denote capital and labor, respectively. The parameter $\chi > 1$ captures the assumption that E firms are more productive. A labor market-clearing condition requires that $N_{E,t} + N_{F,t} = N_t$, where N_t denotes the total urban labor supply at t, whose dynamics are consistent with the demographic model. The technology parameter A grows at the exogenous rate z_t; $A_{t+1} = (1 + z_t) A_t$.

The capital stock of F firms, $K_{F,t}$, is *not* a state variable, since F firms have access to frictionless credit markets, and the capital stock adjusts so that the rate of return on capital equals the lending rate. Note that we assume no irreversibility in investments, so F firms can adjust the desired level of capital in every period. Let r_t^l denote the net interest rate at which F firms can raise external funds. Let w denote the market wage. Profit maximization implies that $K_F = AN_F \left(\alpha / \left(r_t^l + \delta\right)\right)^{-\frac{1}{1-\alpha}}$, where δ is the depreciation rate. The capital-labor ratio and the equilibrium are determined by r^l. Thus,

$$w_t \geq (1-\alpha) \left(\frac{\alpha}{r_t^l + \delta}\right)^{\frac{\alpha}{1-\alpha}} A_t. \tag{28}$$

As long as there are active F firms in equilibrium ($N_F > 0$), equation (28) holds with strict equality.

Let $K_{E,t}$ denote the capital stock of E firms. E firms are subject to an agency problem in the delegation of control to managers. The optimal contract between managers and entrepreneurs requires revenue sharing. We denote by ψ the share of the revenue accruing to managers.[50] Profit maximization yields, then, the following optimal labor hiring decision:

$$\begin{aligned} N_{Et} &= \arg\max_{\tilde{N}_t} \left\{ (1-\psi) (K_{Et})^{\alpha} \left(\chi A_t \tilde{N}_t\right)^{1-\alpha} - w_t \tilde{N}_t \right\} \qquad (29) \\ &= ((1-\psi)\chi)^{\frac{1}{\alpha}} \left(\frac{r_t^l + \delta}{\alpha}\right)^{\frac{1}{1-\alpha}} \frac{K_{Et}}{\chi A_t}. \end{aligned}$$

The gross rate of return to capital in E firms is given by

$$\rho_{E,t} = \left((1-\psi) K_{Et}^{\alpha} (\chi A_t N_{Et})^{1-\alpha} - w_t N_{Et} + (1-\delta) K_{Et}\right) / K_{E,t}. \tag{30}$$

We assume that E firms are also subject to a credit constraint, modeled as in Song *et al.* (2011, p. 216). According to such a model, E firms can borrow funds at the same interest rate as F firms, but the incentive-compatibility constraint of entrepreneurs implies that the share of investments financed externally must satisfy the following constraint:

[50] Managers have special skills that are in scarce supply. If a manager were paid less than a share ψ of production, she could "steal" it. No punishment is credible, since the deviating manager could leave the firm and be hired by another entrepreneur. See Song et al. (2011) for a more detailed discussion.

$$K_E - \Omega_{E,t} \leq \frac{\sigma \rho_E}{1+r^l} K_E, \tag{31}$$

where $\Omega_{E,t}$ denotes the stock of entrepreneurial wealth invested in E firms at t, and, hence, $K_E - \Omega_{E,t}$ denotes the external capital of E firms. Thus, the constraint implies that the entrepreneurs can only pledge to repay a share σ of next-period net profits.

Three regimes are possible: (i) during the first stage of the transition, the credit constraint (31) is binding and F firms are active (hence, the wage is pinned down by (28) holding with equality); (ii) during the mature stage of the transition, the credit constraint (31) is binding and F firms are inactive; (iii) eventually, the credit constraint (31) ceases to bind (F firms remain inactive). In regimes (ii) and (iii), (28) holds with strict inequality.

Consider, first, regime (i). Substituting N_{Et} and w_t into (30) by their equilibrium expressions, (28) and (29), yields the gross rate of return to E firms: $\rho_{E,t} = (1-\psi)\left((1-\psi)\chi\right)^{\frac{1-\alpha}{\alpha}}\left(r_t^l + \delta\right) + (1-\delta)$. The corresponding gross rate of return to entrepreneurial investment is given by $R_{E,t} = \left(\rho_{E,t} K_{E,t} - \left(1+r_t^l\right)\left(K_{E,t} - \Omega_{E,t}\right)\right)/\Omega_{E,t}$. We assume that $(1-\psi)^{\frac{1}{\alpha}}\chi^{\frac{1-\alpha}{\alpha}} > 1$, ensuring that the return to capital is higher in E firms than in F firms (i.e., that $R_{E,t} > r_t^l + 1$). Note that the rate of return to capital is a linear function of r_t^l in both E and F firms. The equilibrium in regime (i) is closed by the condition that employment in the F sector is determined residually, namely,

$$N_{F,t} = N_t - \left((1-\psi)\chi\right)^{\frac{1}{\alpha}} \left(\frac{r_t^l + \delta}{\alpha}\right)^{\frac{1}{1-\alpha}} \frac{K_{Et}}{\chi A_t} \geq 0.$$

Consider, next, regime (ii), where only E firms are active ($N_{E,t} = N_t$) and the borrowing constraint is binding, so (31) holds with equality. In this case, the rates of return to capital and labor equal their respective marginal products. More formally, $w_t = (1-\alpha)(1-\psi)(\chi A_t)^{1-\alpha}(K_{E,t}/N_t)^{\alpha}$, and the gross rate of return on entrepreneurial wealth is given by

$$\rho_{E,t} = \left(\alpha(1-\psi)\chi^{1-\alpha}\left(\frac{K_{Et}}{A_t N_t}\right)^{\alpha-1} + (1-\delta)\right),$$

whereas the borrowing constraint implies that $K_{E,t} = \left(1 + \frac{\sigma \rho_{E,t}}{R^l - \sigma \rho_{E,t}}\right)\Omega_{E,t}$. Given the stock of entrepreneurial wealth, $\Omega_{E,t}$, the two last equations pin down $\rho_{E,t}$ and $K_{E,t}$. The rate of return to entrepreneurial investment is then determined by the expression used for regime (i).

Finally, in regime (iii) the rate of return to capital in E firms is identical to the rate of return offered by alternative investment opportunities (e.g., bonds). Namely,

$$R_{E,t} = 1 + r_t^l.$$

Thus, $K_{E,t}$ ceases to be a state variable, and the wage is given by $w_t = (1-\alpha)\left(\alpha/\left(r_t^l + \delta\right)\right)^{\alpha/(1-\alpha)}\chi A_t$.

In all regimes, the law of motion of entrepreneurial wealth is determined by the optimal saving decisions of managers and entrepreneurs, described below.

The rural production sector consists of rural firms whose technology is assumed to be similar to that of urban F firms, $Y_{Rt} = K_{Rt}^{\alpha_R}(\chi_R A_t N_{Rt})^{1-\alpha_R}$, where $\chi_R < 1$. Like urban F firms, rural firms can raise external funds at the interest rate r_t^l in each period, and adjust their capital accordingly. So, r_t^l pins down capital-labor ratio and wage in the rural economy. This description is aimed to capture, in a simple way, the notion that there are constant returns to labor in rural areas, due to, e.g., rural overpopulation.

Banks: Competitive financial intermediaries (*banks*) with access to perfect international financial markets collect savings from workers and hold assets in the form of loans to domestic firms and foreign bonds. Foreign bonds yield an exogenous net rate of return denoted by r, constant over time. Arbitrage implies that the rate of return on domestic loans, r_t^l, equals the rate of return on foreign bonds, which in turn must equal the deposit rate. However, lending to domestic firms is subject to an *iceberg cost*, ξ, which captures the operational costs, red tape, and so on, associated with granting loans. Thus, ξ is an inverse measure of the efficiency of intermediation. In equilibrium, $r^d = r$ and $r_t^l = (r + \xi_t) / (1 - \xi_t)$, where r_t^l is the lending rate to domestic firms.

Households' saving decisions: Workers and retirees face the problem discussed in section 2, given the equilibrium wage sequence, and having defined $R \equiv 1 + r$. As in the previous section, we hold fixed the share of workers participating in the pension system.

The young managers of E firms earn a managerial compensation m. Throughout their experience as managers, they acquire skills enabling them to become entrepreneurs at a later stage of their lives. The total managerial compensation in period t equals $M_t = \psi Y_{E,t}$. Managers work for J_E years, and during this time can only invest their savings in bank deposits (as can workers) which yields an annual gross return R. As they reach age $J_E + 1$, they retire as managers, and have the option (which they always exercise) to become entrepreneurs. In this case, they invest their wealth in their own business yielding the annual return $R_{E,t}$ and hire managers and workers. Thereafter, they are the residual claimants of the firm's profits. We assume that entrepreneurs are not in the pension system. Their lifetime budget constraint is then given by

$$\sum_{j=0}^{J_E} \frac{s_j}{R^j} c_{t+j} + \sum_{j=J_E+1}^{J} \frac{1}{R^{J_E}} \frac{s_j}{\Pi_{v=t+J_E+1}^{t+j} R_{E,\nu}} c_{t+j} = \sum_{j=0}^{J_E} \frac{s_j}{R^j} m_{t+j}.$$

The right hand-side is the PDV income from the managerial compensation. The left hand-side yields the PDV of consumption. This is broken down in two parts: the first term is the PDV of consumption when young, when the manager faces a constant rate of return, R; the second part is the PDV of consumption when being an entrepreneur, and is discounted at the rate R until J_E, and at the entrepreneurial rate of return thereafter.

Mechanics of the model: The dynamic model is defined up to a set of initial conditions including the wealth distribution of entrepreneurs and managers, the wealth of the pension system, the aggregate productivity (A_0), and the population distribution. The engine of growth is the savings of managers and entrepreneurs. If the economy starts in regime (i), then all managerial savings are invested in the entrepreneurial business as soon as each manager becomes an entrepreneur. As long as managerial investments are sufficiently large, the employment share of E firms grows and that of F firms declines over time.

The comparative dynamics of the main parameters is as follows: (i) a high β implies a high propensity to save for managers and entrepreneurs and a high speed of transition; (ii) a high world interest rate (r) and/or a high iceberg intermediation cost (ξ) increases the lending rate, implying a low wage, a high rate of return in E firms, a high managerial compensation, and, hence, a high speed of transition; (iii) a high productivity differential (χ) implies a high rate of return in E firms, a high managerial compensation, and, hence, a high speed of transition; (iv) a high σ implies that entrepreneurs can leverage up their wealth and earn a higher return on their savings, which speeds up the transition; and (v) a high managerial rent (ψ) implies a low rate of return in E firms, a high managerial compensation, and, hence, has ambiguous (and generally non-monotonic) effects on the speed of transition.

Note that the savings of the worker do not matter for the speed of transition, because the lending rate offered by banks depends only on the world market interest rate and on the iceberg cost.

E.1 Calibration

In SSZ, we show that a calibrated version of the model outlined in the previous section matches well a number of salient macroeconomic trends for the recent period. In particular, the model reproduces realistic trends for output growth, wage growth, return to capital, transition from state-owned to private firms, and foreign surplus accumulation. The current model - which incorporates additional features including demographics and the pension system - the model is calibrated to match the same macroeconomic trends after 2000.

We must calibrate two parameters related to the financial system, ξ and σ, and four technology parameters, α, δ, χ and ψ. The parameters α and δ are set exogenously: $\alpha = 0.5$ so that the capital share of output is 0.5 in year 2000 (Bai *et al.*, 2006), and $\delta = 0.1$ so that the annual depreciation rate of capital is 10%.

The remaining parameters are calibrated internally, so as to match a set of empirical moments. We set the parameters ψ and χ so that the model is consistent with two key observations: (i) the capital-output ratio in E firms is 50% of the corresponding ratio in F firms (as documented by SSZ for manufacturing industries, after controlling for three-digit industry type), (ii) the rate of return on capital is 9% larger in E firms than in F firms.[51] The implied parameter values are $\psi = 0.27$ and $\chi = 2.73$. This implies that the TFP of an E firm is 1.65 times larger than the TFP of an F firm.[52]

We set ξ so as to target an average gross return on capital of 20% in year 2000 (Bai *et al.*, 2006). With $\delta = 10\%$, this implies an average net rate of return on capital of 10%. This average comprises both F firms and E firms. Since the DPE employment share in the period 1998-2000 was on average 10%, this implies $\rho_F = 9.3\%$, so that the initial value for ξ is $\xi_{2000} = 0.062$. After year 2000, we assume that there is gradual financial improvement so ξ falls linearly to zero by year 2024. The motivation for such decline is twofold. First, we believe it is reasonable that banks improve their lending practices over time, so that borrowing-lending spreads will eventually be in line with corresponding spreads in developed economies. Second, a falling ξ will generate capital deepening in F firms and E firms due to cheaper borrowing and higher wages, respectively. Such development helps the model to generate an increasing aggregate investment rate during 2000-2009, which is a clear pattern of aggregate data. If ξ were constant, the model would predict a falling rate (see Song *et al.*, 2011, for further discussion).

We set $\sigma = 0.43$, so that entrepreneurs can borrow 87 cents for each dollar in equity in 2000. This value for σ implies that the growth in the DPE employment share is in line with private employment growth between 2000 and 2008 in urban areas. We set the initial level of productivity, A_{2000}, so that the GDP per capita is 8.3% of the US level in 2000. This yields a GDP per capita equal to 20% of the US level in 2010, in line with the data. Moreover, we set the growth rate of A_t (i.e., the secular exogenous productivity growth) so that the model generates an average labor income growth (controlling for human capital) of 7.5% between 2000-2013. The resulting growth rate in A_t is 2.1% larger than the associated world TFP growth rate during this period. After 2010, the growth rate of A_t in excess of the long-run world rate falls linearly to zero until the TFP level in E firms reaches that of US firms. This occurs in year 2022. Thereafter, the TFP grows at the long-run world rate. Finally, β is calibrated to 1.0164 to match the average aggregate urban household saving rate of 25% in 2000-2010.

In the rural sector, we set $\alpha_R = 0.3$ to match the observed 20% investment rate in the rural area in 2000. The technology gap χ_R is set to 0.75 to capture an observed urban-rural wage gap of 1.84 in

[51] Song et al. (2011) document that manufacturing, domestic private enterprises (DPE) have on average a ratio of profits per unit of book-value capital 9% larger than that of SOEs during the period 1998-2007. A similar difference in rate of return on capital is reported by Islam, Dai, and Sakamoto (2006).

[52] Hsieh and Klenow (2009) estimate TFP across manufacturing firms in China and find that the TFP of DPEs is about 1.65 times larger than the TFP of SOEs.

2000. The rural wage grows over time, due to the exogenous technology growth and to the decreasing lending rate. The rural-urban wage gap implied by the model increases from 1.84 in 2000 to 3.48 in 2040 and stays constant thereafter (see figure VI in the appendix).

The initial conditions are set as follows. Total entrepreneurial wealth in 2000 is set equivalent to 14.6% of urban GDP so that the 2000 DPE employment is 20%. The distribution of that entrepreneurial wealth is obtained by assuming that all entrepreneurs are endowed with the same initial wealth in 1995. The initial wealth for workers, retirees, and managers is set so as to match as the 1995 empirical age distribution of financial wealth for urban households from CHIP. The 2000 distribution of wealth across individuals is then derived endogenously. Finally, the initial government wealth is set to 96% of GDP in 2000 so as to generate a net foreign surplus equal to 12% of GDP in 2000.[53]

E.2 Simulated output trajectories

The calibrated model yields growth forecasts that we view as plausible. Figure II shows the evolution of productivity and output per capita forecasted by our model. The growth rate of GDP per worker remains about 7.5% per year until 2020 (see upper panel). After 2020, productivity growth is forecasted to slow down. This is driven by two forces: (i) the end of the transition from state-owned to private firms and (ii) the slowdown in technological convergence. The growth rate remains above 7.2% between 2020-2030 and eventually dies off in the following decade. Note that the growth of GDP per capita is lower than that of GDP per worker after 2013, due to the increase in the dependency ratio. On average, China is expected to grow at a rate of 6.5% between 2013 and 2040. The contribution of human capital is 0.8% per year, due to the entry of more educated young cohorts in the labor force. In this scenario, the urban GDP per worker in China will be 73% of the US level by 2040, remaining broadly stable thereafter. The corresponding GDP per capita of China is 68% of the US level in 2040. Total GDP in China is set to surpass that in the United States in 2013 and to become more than twice as large in the long run.

The wage sequence that was assumed in section 2 is now an endogenous outcome. Wages are forecasted to grow at an average of 4.9% until 2031 and to slow down thereafter. What keeps wage growth high after 2020 is mostly capital deepening.[54]

E.2.1 Sensitivity: high savings and foreign surplus

Although the growth forecasts are plausible, the calibrated economy generates a very large amount of savings. For instance, in 2065 the economy has a wealth-GDP ratio exceeding 1000%. This is because the model is calibrated to match urban household saving during 2000-2010. In that period, China experienced high growth and yet a very high saving rate (a total savings rate of 48.2%, and a household savings rate of 25%).

Since our stylized model forecasts an eventual decline in growth, the intertemporal motive would suggest that consumption should have been high before 2010. Therefore, the model requires a sufficiently high discount factor ($\beta = 1.0164$) in order to predict the empirical saving rate during the first decade of the 21st century. In our model, a high β is a stand-in for a number of institutional

[53] More precisely, government wealth is calculated as a residual. It is equal to the sum of foreign surplus and domestic capital (from both SOE and DPE) minus the stock of private wealth owned by workers and entrepreneurs.

[54] In Section 4 we held the wage sequence constant across the different policy experiments. However, in the general equilibrium model of this section, the wage sequence is endogenous and would in general be affected by alternative reforms. In particular, pension reforms impact labor supply through a wealth effect, and this influences the capital accumulation dynamics during transition. Since the effects are quantitatively small, the results are omitted and are available upon request.

features that are not explicitly considered and that may explain a high propensity to save over and beyond pure preferences (e.g., large precautionary motives or large downpayment requirements for house purchases).[55]

Since it seems implausible that China will continue to save so much, we consider an alternative scenario, where all cohorts entering the labor market after 2013 have $\beta = 0.97$. In such an alternative scenario China's net foreign position would be zero in the long run. The analysis of the alternative pension arrangements discussed in the previous sections yields essentially the same results as in the high β economy. Thus, the calibration of β is unimportant for the effects of the welfare analysis, which is the main contribution of this paper.

This finding is not surprising since long-term wages and GDP do not hinge on the domestic propensity to save. Although the entrepreneurs' propensity to save determines the speed of the transition, this does not to matter much for welfare (see section 5.1).

E.2.2 Sensitivity: Financial development

The model borrows from SSZ the assumption that E firms are financially constrained. Note that the salience of the financial constraints declines over time as E firms accumulate capital. As the economy enters regime (iii), which occurs in 2040, the financial constraint ceases to bind.

In our baseline calibration, the parameter σ, which regulates borrowing of private firms, is assumed to be constant over time. An exogenous increase in σ – for example, due to financial development – would speed up growth of private firms. Wage growth would accelerate earlier, although the long-run wage level would be unaffected.

To study the effects of financial development on pension reform, we consider a stark experiment in which the borrowing constraint on private firms is completely removed in 2013. This means that state-owned firms vanish, and there is large capital inflow driven by entrepreneurial borrowing. Wages jump upon impact (by 88%) due to the large capital deepening. In 2030, the wage level is still 18.5% above the baseline calibration. In 2040 the wage level is the same as in the benchmark calibration.

Although financial development affects the transition path, it brings little change to the conclusions of the welfare analysis.[56] The benchmark reform requires a slightly smaller reduction of the replacement rate: 39.8% instead of 39.1%. The delayed reform still entails gains for the transition cohorts, albeit these gains decline faster over time. For instance, delaying a reform until 2050 yields a 17% consumption equivalent gain for the cohort retiring in 2013, but only a 10.5% gain for the cohort retiring in 2049. The losses suffered by the cohorts retiring after 2050 are comparable in size to those in the baseline scenario without financial development. The gains accruing to the high- and low-discount planners are, respectively, 5.3% and 0.5% (6.4% and 0.9% in the baseline scenario).

The FF reform yields slightly better outcomes. All generations retiring after 2050 gain from the reform (2060 in the baseline scenario), and the losses of the earlier cohorts only reach 7% (11% in the baseline scenario). The high-discount planner continues to prefer the benchmark reform to the FF reform, whereas the low-discount planner continues to have the opposite ranking. The PAYGO reform yields even larger gains to the earlier cohorts. Both the high- and the low-discount social planners continue to prefer the PAYGO reform to any alternative policy-driven reform. However, the welfare gap between the PAYGO and the fully funded reform is now smaller, since the planners dislike the concentrated nature of the gains under the PAYGO reform. For instance, the consumption

[55] Chamon *et al.* (2013) and Song and Yang (2010) study household savings in calibrated life-cycle models. They incorporate individual risk and detailed institutional features of the pension system and find that their models are qualitatively consistent with the life-cycle profile of household saving rates. However, both studies find that with a conventional choice of β, their models would imply quantitatively too low savings for the young households.

[56] We focus for simplicity on the policy-driven reforms, and we omit an explicit analysis of the optimal policy.

equivalent gain of the low-discount planner relative to the benchmark reform is 1%, compared with 1.7% in the baseline scenario. Since the fully funded reform also entails a 0.5% gain relative to the benchmark reform, the consumption equivalent gain of the PAYGO relative to the FF reform is only 0.5% (although it remains significantly higher, 12.4%, for the high-discount planner).

In conclusion, financial development mitigates but does not change the welfare implications of alternative reforms.

APPENDIX REFERENCES

Liang, Zai, and Zhongdong Ma, 2004. "China's Floating Population: New Evidence from the 2000 Census." *Population and Development Review*, 30(3), 467-488.

Zhao, Yaohui, and Jianguo Xu. 2002. "China's Urban Pension System: Reforms and Problems." *The Cato Journal*. 21.3, pp. 395-414.

APPENDIX FIGURES

In this section, we provide the appendix figures.

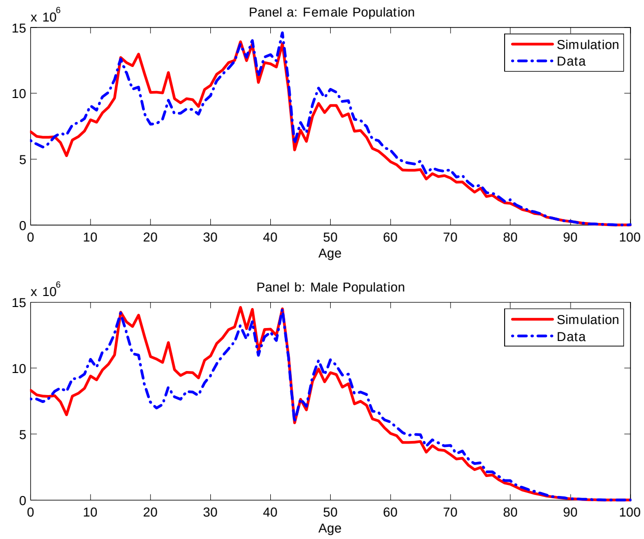

Figure I: The upper panel shows the female population of different ages in 2005, in the survey data (solid line), and in our simulation (dashed line). The lower panel shows the male population in 2005.

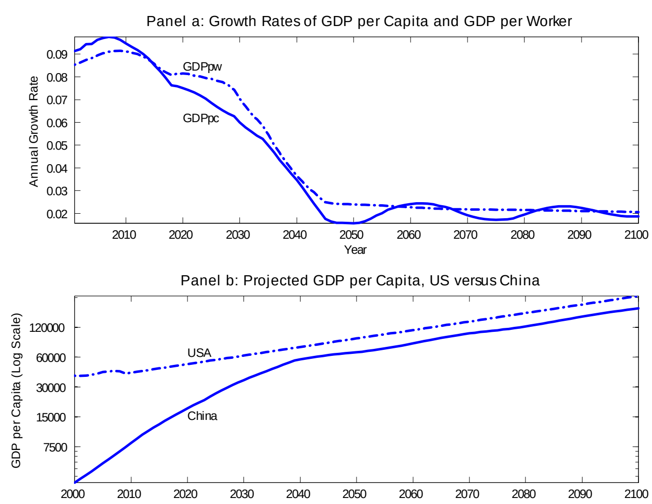

Figure II: The upper panel shows projected annual growth rates in GDP per worker and GDP per capita in the calibrated economy. The lower panel shows projected GDP per capita in levels for China and the US.

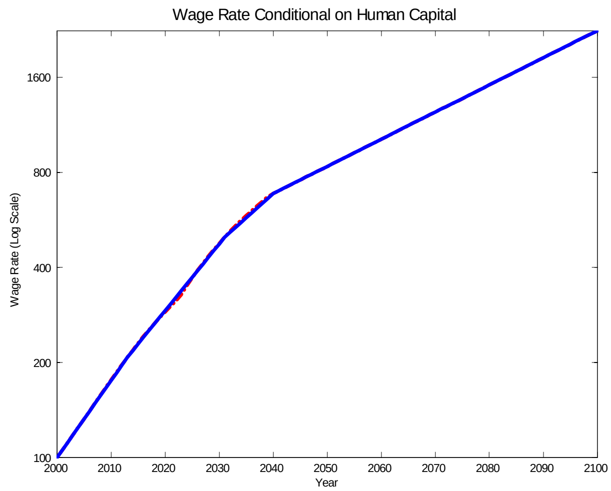

Figure III: The figure shows the assumed hourly wage rate per unit of human capital in urban areas, normalized to 100 in 2000. The solid line is the assumed wage process and the dashed line is the wage process consistent with the endogenous outcome of the general equilibrium model of section E. Note that the two lines are almost indistinguishable.

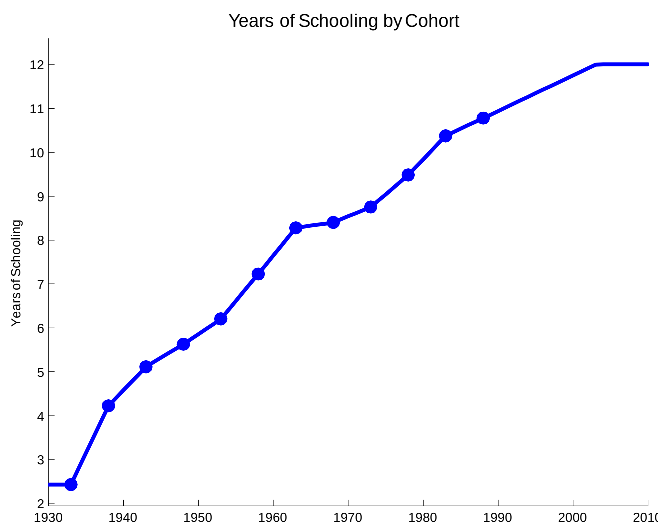

Figure IV: The figure shows the average number of years of schooling for different age cohorts in China. Source: Barro and Lee data set. The values after 1990 are (linearly) extrapolated, assuming the growth in schooling accumulation stagnates at 12 years.

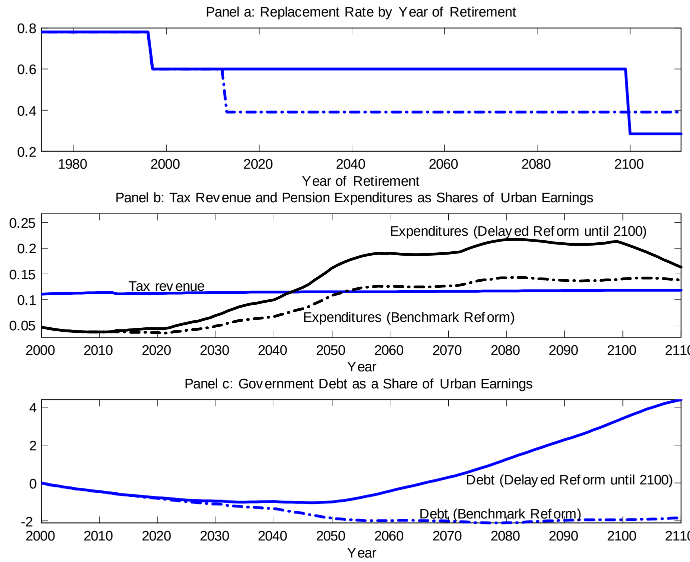

Figure V: Panel (a) shows the replacement rate q_t for the case when the reform is delayed until 2100 (solid line) versus the benchmark reform (dashed line). Panel (b) shows tax revenue and expenditures, expressed as a share of aggregate urban labor income (benchmark reform is dashed and the delay-until-2100 is solid). Panel (c) shows the evolution of government debt, expressed as a share of aggregate urban labor income (benchmark reform is dashed and the delay-until-2100 is solid). Negative values indicate surplus.

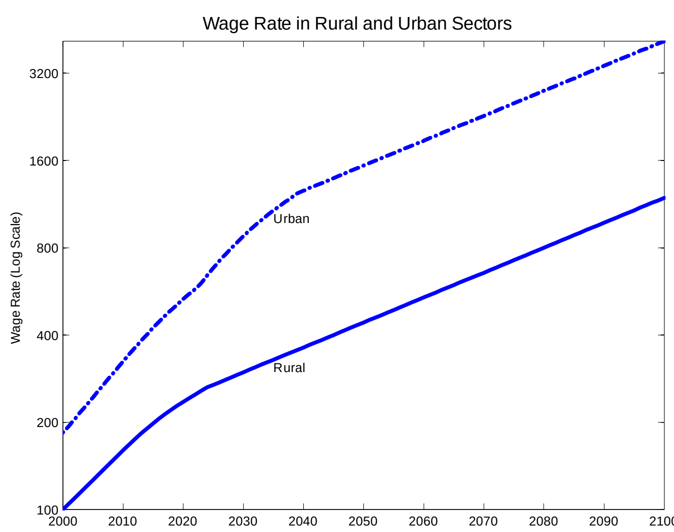

Figure VI: The figure shows the projected hourly wage rate per unit of human capital in urban (dashed line) and rural (continuous line) areas, normalized to 100 in rural areas in 2000. The process is the endogenous outcome of the general equilibrium model of section E.

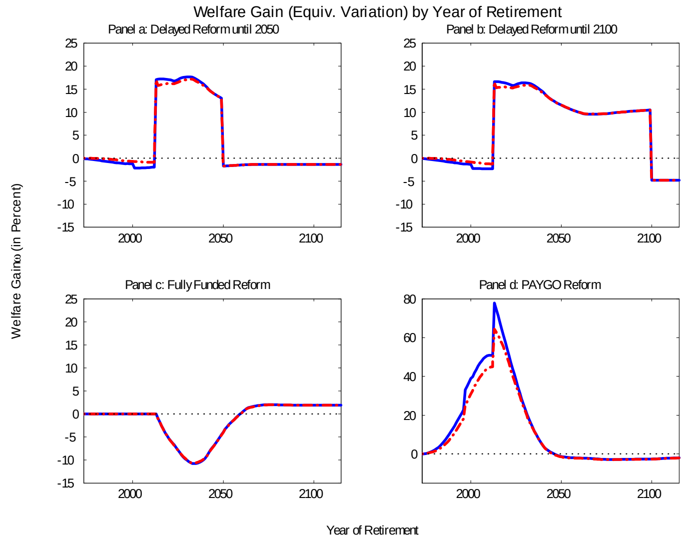

Figure VII: As in figure (6), the solid lines show welfare gains of alternative reforms relative to the benchmark reform for each cohort, but now under the assumption that all the reforms are perfectly anticipated at 2000. The dashed lines are the welfare gains in the baseline scenario, as in figure (6). The gains (ω) are expressed as percentage increases in consumption.

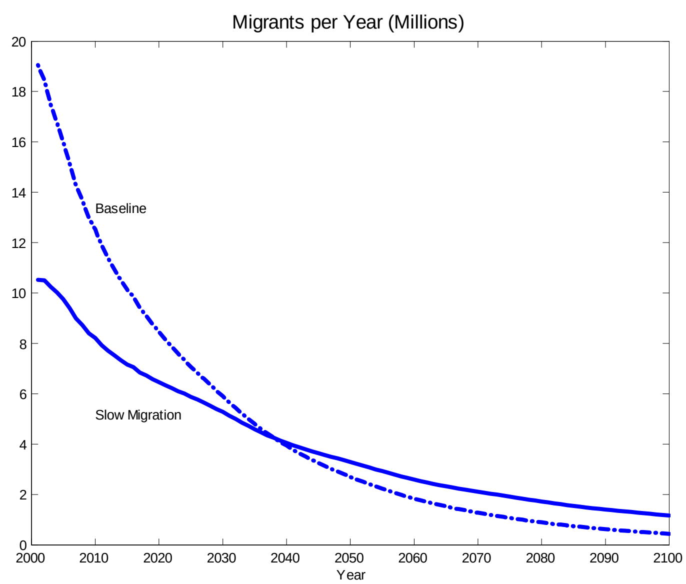

Figure VIII: The migration flow (i.e., the number of migrants per year) in the slow migration and basline scenarios are shown with the solid and dashed lines, resepectively. The migration flow is smaller in the slow migration scenario than in the basline scenario before 2038, but larger afterwards.