

Sharing High Growth Across Generations: Pensions and Demographic Transition in China*

Zheng Song
University of Chicago Booth

Kjetil Storesletten
Federal Reserve Bank of Minneapolis and CEPR

Yikai Wang
University of Zurich

Fabrizio Zilibotti
University of Zurich and CEPR

June 2012

Abstract

The benefits of Chinese growth are unequally distributed across cohorts. China's aging population threatens the sustainability of its pension system, a key vehicle of intergenerational redistribution. We analyze the welfare effects of alternative pension reforms with the aid of a dynamic general equilibrium model incorporating population dynamics and productivity growth. Although a reform is necessary, delaying its implementation implies large welfare gains for the (poorer) current generations, imposing only small costs on (richer) future generations. In contrast, a fully funded reform harms current generations, and yields small gains to future generations. High wage growth is key for these normative results.

JEL Codes. E21, E24, G23, H55, J11, J13, O43, R23.

Keywords: China, Credit market imperfections, Demographic transition, Economic growth, Fully-funded system, Intergenerational redistribution, Labor supply, Migration, Pensions, Rural-urban reallocation, Total Fertility Rate, Wage growth.

*We thank Philippe Aghion, Tim Besley, Martin Eichenbaum, Vincenzo Galasso, Dirk Krueger, Torsten Persson, Richard Rogerson, and seminar participants at the Conference "China and the West 1950-2050: Economic Growth, Demographic Transition and Pensions" (University of Zurich November 21, 2011), London School of Economics, Princeton University, Tsinghua Workshop in Macroeconomics 2011, Università della Svizzera Italiana, University of Frankfurt, University of Mannheim, University of Pennsylvania, University of Toulouse. Yikai Wang acknowledges financial support from the Swiss National Science Foundation (grant no. 100014-122636). Fabrizio Zilibotti acknowledges financial support from the ERC Advanced Grant IPCDP-229883. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

China has grown at stellar rates over the last thirty years. With a GDP per capita still below 20% of the US level, it has still ample scope for further convergence in technology and productivity. However, the success is imbalanced. The labor share of output is low and stagnating, corroborating the perception that the welfare of the majority of the population is not keeping the pace with the high output growth. Intergenerational inequality is also very large, due to the fast productivity growth. For instance, the present value of the income of a young worker who entered the labor force in 2000 is about six times as large as that of a worker who entered in 1970 and is today about to retire. These observations motivate the growing debate about what institutional arrangements can allow more people to share the benefits of high growth.¹

An important aspect of this debate is China's demographic transition. The total dependency ratio has fallen from 75% in 1975 to a mere 37% in 2010. This is due to the combination of a high fertility in the 1960's, and the family planning policies introduced in the 1970's, culminating with the draconian one-child policy of 1978. The expansion of the labor force implied by this transition has contributed to economic growth. However, China is now at a turning point: the old-age dependency ratio will increase from the current 12% to 39% in 2040. The ageing population threatens the viability of redistributive policies, especially pensions, which are arguably the most important institutional vehicle of intergenerational redistribution. In this paper, we analyze the welfare effects of alternative pension reforms.

Our analysis is based on a dynamic general equilibrium model incorporating a public pension system. The standard tool for such analyses is the Auerbach and Kotlikoff (1987) model (henceforth the *Au-Ko model*) – a multiperiod overlapping generations (OLG) model with endogenous capital accumulation, wage growth, and an explicit pension system. Our model departs from the canonical Au-Ko model by embedding some salient structural features of the Chinese economy: the rural-urban transition and a rapid transformation of the urban sector, where state-owned enterprises are declining and private entrepreneurial firms are growing. Such a transition is characterized, following Song *et al.* (2011), by important financial and contractual imperfections.

The model bears two key predictions. First, wage growth is delayed: As long as the transition within the urban sector persists, wage growth is moderate. Yet, as the transition comes to an end, the model predicts an acceleration of the wage growth. Second, financial imperfections cause a large gap between the rate of return to industrial investments and the rate of return to which Chinese

¹For instance, Wen Jiabao, head of the government, declared in the press conference held on March 14, 2012: "*I know that social inequities... have caused the dissatisfaction of the masses. We must push forward the work on promoting social equity...The first issue is the overall development of the reform of the income distribution system.*"

households have access. A calibrated version of the model forecasts that wages will grow at an average 6.2% until 2030, and slow down rapidly thereafter. GDP growth will also slow down, but is expected to remain as high as 6% per year over the next two decades. By 2040, China will have converged to about 70% of the level of GDP per capita of the US.

We use the model to address two related questions: (i) Is a pension system based on the current rules sustainable? (ii) What are the intra- and inter-generational welfare effects of alternative reforms? The answer to the first question is clear-cut: the current system is unbalanced and requires a significant adjustment in either taxes or benefits. We focus on the benefit margin, and consider a benchmark reform reducing the pension payments to all workers retiring after 2011. We assume that the reform does not renege on the outstanding obligations to current retirees, but only changes the entitlements of workers retiring as of 2012 – this is the pattern of most reforms in OECD countries. This reform entails a sharp permanent reduction of the replacement rate, from 60% to 40%. Note that this reform implies that the accumulation of a large pension fund until 2050.

To address the second question, we consider three alternative scenarios. First, we study the effect of a *delayed reform*, by which the current rules remain in place until a future date T , to be followed by a permanent reduction in benefits, so as to balance the pension system in the long run. If the reform is delayed until 2040, our model predicts large welfare gains for the transition generations relative to the draconian benchmark reform in 2012. Quantitatively, the gains accruing to the cohorts retiring before 2040 would be equivalent to an increase of 17% of their lifetime consumption. The generations retiring after 2040 would only suffer small additional losses in the form of an even lower replacement ratio. Second, we consider the effects of switching to a pure pay-as-you-go (*PAYGO*) system where the replacement rate is endogenously determined by the dependency ratio, subject to a balanced budget condition for the pension system. A *PAYGO* reform has a similar, if more radical, welfare effects as a delayed reform. Given the demographic transition of China, the *PAYGO* yields very generous pensions to early cohorts and punishes more severely the generations retiring after 2050. Both reforms share a common feature: they allow the poorer current generations to share the benefits of high wage growth with the richer generations that will enter the labor market when China is a mature economy. Finally, we consider switching to a pure individual account savings-based system, which we label a *fully-funded reform*. In our model, this is equivalent to eliminating the public system altogether. To honor existing obligations, the government issues bonds to compensate current workers and retirees for their past contributions. Since we assume the economy to be dynamically efficient, a standard trade-off emerges: all generations retiring after 2062 benefit from the fully-funded reform, while earlier generations lose.

We aggregate the welfare of different cohorts using a utilitarian social planner who discounts the welfare of future cohorts at a reasonable rate. We show that even a highly forward-looking planner

with an annual discount rate as low as 0.5% would choose to either switch to a PAYGO or to delay the implementation of a sustainable pension reform. Such alternative reforms are preferred to the immediate implementation of the sustainable reform as well as to the fully-funded reform. The motive is the drive to redistribute income from the *rich* cohorts retiring in the distant future to the *poor* cohorts retiring in the next coming years.

These normative predictions run against the common wisdom that switching to a pre-funded pension system is the best response to adverse demographic dynamics. For instance, Feldstein (1999) and Dunaway and Vivek (2007) argue that a fully-funded reform is the best viable option for China. Our findings hinge on two key features of China that are equilibrium outcomes in our model: a high wage growth and a low rate of return on savings.² If we lower the wage growth to an average 2% per year (a conventional wage growth for mature economies), the main results are reversed: the planner who discounts the future at an annual 0.5% would prefer a fully funded (FF) reform, or alternatively the immediate implementation of the draconian sustainable reform, over a PAYGO. Thus, our analysis illustrates a general point that applies to fast-growing emerging economies. Even for economies that are dynamically efficient, the combination of (i) a prolonged period of high wage growth and (ii) a low return to savings to large financial imperfections makes it possible to run a relatively generous pension system over the transition without imposing a large burden to future generations.

The current pension system of China only covers ca. 60% of urban workers. We analyze the welfare effect of making the system universal, extending its coverage to all rural and urban workers. This issue is topical, as the government of China is currently introducing some form of rural pensions. The recurrent question is to what extent this is affordable, and how generous rural pensions can be, since almost half of today's population lives in rural areas, and these workers have not contributed to the system thus far. We find that extending the coverage of the pension system to rural workers would be relatively inexpensive, even though full benefits were paid to workers who never contributed to the system. As expected, this would trigger large welfare gains for the poorest part of the Chinese population. The cost is small, since (i) benefit are linked to local wages, and rural wages are low; (ii) the rural population is shrinking.

The paper is structured as follows. Section 2 outlines the detailed demographic model. Section 3 lays out a calibrated partial equilibrium model *à la* Au-Ko incorporating the main features of the Chinese pensions system. In this section, we assume exogenous paths for wages and interest rate. Section 4 quantifies the effects of the alternative pension reforms. Section 6 provides a full general

²Different from us, Feldstein (1999) assumes that the Chinese government has access to a riskfree annual rate of return on the pension fund of 12%. Unsurprisingly, he finds that a fully funded system that collects pension contributions and invest these funds at such a remarkable rate of return, will dominate a pay-as-you-go pension system that implicitly delivers the same rate of return as aggregate wage growth.

equilibrium model of the Chinese economy based on Song *et al.* (2011) where the wage and interest rate path assumed in Section 3 are equilibrium outcomes. The model allows us to consider reforms that influence the economic transition. Section 7 concludes. An Appendix [missing in this version] contains some technical material.

2 Demographic Model

Throughout the 1950s and 1960s, the total fertility rate (henceforth, TFR) of China was an average about six. Such a high TFR, together with a declining mortality led to a rapid expansion of the total population. The 1982 census estimated a population size of one billion, 70% higher than in the 1953 census. The belief that a booming population is a burden on the development process induced the government to introduce a set of measures to curb fertility during the 1970's, culminating in the one-child policy of 1978. This policy imposes severe sanctions on couples who have more than one child. The policy underwent a few reforms, and is currently more lenient to rural families and ethnic minorities. For instance, rural families are allowed a second birth provided the first child is a girl. In some provinces, *all* rural families are allowed to have a second child provided that a time interval (which varies across provinces) elapses between the first and second birth. Today's TFR is below replacement level, although there is no uniform consensus about its exact level. Estimates based on the 2000 census and earlier surveys in the 1990s range between 1.5 and 1.8 (e.g., Zhang and Zhao, 2006). Recent estimates suggest a TFR of about 1.6 (see Zeng 2007). The demographic outlook is the source of growing concern. Although no Copernican revolution is in the horizon, the Chinese government is gradually loosening the birth control policy, especially in some urban areas.³

2.1 Natural Population Projections

We consider, first, a model without rural-urban migration, which is referred to as the *natural* population dynamics. We break down the population by birth place (rural vs. urban), age and gender. The initial population size and distribution are matched to the adjusted 2000 census data.⁴ There is consensus among demographers that birth rates have been under-reported, causing a deficit of 30 to 37 million children in the 2000 census.⁵ To heed this concern, we take the rural-urban population and

³In 2008, China's National Population and Family Planning Commission stated that the one-child policy would not be lifted for at least ten more years.

⁴The 2000 census data is broadly regarded as a reliable source (see, e.g., Lively, 2001; Goodkind, 2004). The total population was originally estimated to be 1.24 billion, later revised by the NBS to 1.27 billion (see the Main Data Bulletin of 2000 National Population Census). The NBS also adjusted the urban-to-rural population ratio from 36.9% to 36%.

⁵See Goodkind (2004). A similar estimate is obtained by Zhang and Cui (2003) who use primary school enrolments to back out the actual child population.

age-gender distribution from the 2000 census – with the subsequent NBS revisions – and then amend this by adding the missing children for each age group, according to the estimates of Goodkind (2004).

The initial group-specific mortality rates are also estimated from the 2000 census, yielding a life expectancy at birth of 71.1 years, which is very close to the estimate reported by the World Development Indicator in the same year (71.2). It is reasonable to expect that life expectancy will continue to increase as China grows richer. Therefore, we set the mortality rates in 2020, 2050 and 2080 to match the demographic projection by Zeng (2007), and use linear interpolation over the intermediate periods. We assume no further change after 2080. This implies a long-run life expectancy of 81.9 years.

The age-specific urban and rural fertility rates for 2000 and 2005 are estimated using the 2000 census and the 2005 survey, respectively. We interpolate linearly the years 2001-04, and assume the age-specific fertility rates to remain constant at the 2005 level over the period 2006-11. This yields average urban and rural TFR of 1.2 and 1.98, respectively.⁶ Between 2011 and 2050, we assume the age-specific fertility rates to remain constant in rural areas. This is motivated by the observation that, according to the current legislation, a growing share of urban couples (in particular, those in which both spouses are singleton) will be allowed to have two children. In addition, some provinces are discussing a relaxation of the current rule, that would allow even urban couples in which only one spouse is a singleton to have two children.⁷ Zeng (2007) estimates that such a policy would increase the urban TFR from 1.2 to 1.8 (*second scenario* in Zeng (2007)). Accordingly, we assume that the TFR increases to 1.8 in 2012, and then remains constant until 2050.

A long-run TFR of 1.8 implies an ever shrinking population. We follow the United Nations population forecasts, and assume that in the long run the population will be stable. This requires that the TFR converges to 2.078, which is the reproduction rate in our model, in the long run. In order to smooth the demographic change, we assume that both rural and urban fertility rates starts growing in 2051, and use a linear interpolation of the TFRs for the years 2051–99. Since such long-run forecasts are subject to a large uncertainty we also consider an alternative scenarios with a lower fertility.

⁶The acute gender imbalance is taken into account in our model. However, demographers view as unlikely that such imbalance will persist at the current high levels. Following Zeng (2007), we assume that the urban gender ratio will decline linearly from 1.145 to 1.05 from 2000 to 2030, and that the rural gender imbalance falls from 1.19 to 1.06 over the same time interval. No change is assumed thereafter. Our results are robust to plausible changes in the gender imbalance.

⁷In July 2011, Zhang Feng, director of the Guangdong provincial population and family planning commission issued a public request to let his province introduce a looser by which couples would be allowed an extra child if even one parent (as opposed to both) were a single child (The Economist, July 2011). However, in a more recent interview with the Nanfang Daily (October 10, 2011), the same officer declared that there would be no major adjustments to the family planning policy in the near future.

2.2 Rural-Urban Migration

Rural-urban migration has been a prominent feature of the Chinese economy since the 1990s. There are two categories of rural-urban migrants. First, all individuals who physically moved from rural to urban areas. This category include both people who changed their registered permanent residence (i.e., *hukou* workers) and people who reside and work in urban areas but retain an official residence in a rural area (*non-hukou* urban workers).⁸ Second, all individuals who did not move but whose place of registered residence switched from being classified as rural into being classified as urban.⁹ We define as the "net migration flow" (NMF) the sum of the two categories.

We propose a simple model of migration where the age- and gender-specific emigration rates are fixed over time.¹⁰ To this end, it is necessary to estimate the NMF and its associated distribution across age and gender. The estimation will be the backbone of our projection of migration and the implied rural and urban population dynamics. First, we use the 2000 census and construct a projection of the *natural* rural and urban populations until 2005 based on the method described above. Then, we compare our projection to the 2005 survey data. The differences between the *natural* populations and the 2005 survey data yield an estimate of the NMF and its distribution across age groups.¹¹ The technical details of the estimation are deferred to an appendix.

According to our estimates, the overall NMF between 2000 and 2005 was 91 million, corresponding to 11.1% of the rural population in 2000.¹² Survey data show that the urban population grows at

⁸There are important differences across these two subcategories. Most non-resident workers are currently not covered by any form of urban social insurance including pensions. However, there have been relaxations of the system in recent years. The system underwent some reforms in 2005, and in 2006 the central government abolished the hukou requirement for civil servants (Chan and Buckingham, 2008). Since there are no reliable estimates of the number of non-hukou workers, and in addition there is uncertainty about how the legislation will evolve in future years, we decided not to distinguish explicitly between the two categories of migrants in the model. This assumption is of importance with regard to the coverage of different type of workers in the Chinese pension system and we will return to its discussion below.

⁹This was a sizeable group in the 1990s: According to China Civil Affairs' Statistical Yearbooks, a total of 8439 new towns were established from 1990 to 2000 and 44 million rural citizens became urban citizens (Hu, 2003). However, the importance of reclassified areas has declined after 2000. Only 24 prefectures were reclassified as prefecture-level cities in 2000-2009, while 88 prefectures were reclassified in 1991-2000.

¹⁰Although emigration rates likely responds to the urban-rural wage gap, pension and health care entitlements for migrants, the rural old-age dependency ratio, etc., we will abstract from this and maintain that the demographic development is exogenous. It is very difficult to estimate the future migration elasticities given that the migration flows in China have been restricted by legal and administrative regulations. Moreover, even for developed countries the internal migration patterns remain hard to predict (XXXcitationASK_SAM).

¹¹Our method is related to Johnson (2003), who also exploits *natural* population growth rates. Our work is different from Johnson's in three respects. First, his focus is on migration across provinces, while we estimate rural-urban migration. Second, Johnson only estimates the total migration flow, while we obtain a full age-gender structure of migration. Finally, our estimation takes care of measurement error in the census and survey (see discussion above), which were not considered in previous studies.

¹²There are a number of inconsistencies across censuses and surveys. Notable examples include changes in the definition of city population and urban area (see, e.g., Zhou and Ma, 2003; Duan and Sun, 2006).

Such inconsistencies could potentially bias our estimates. In particular, the definition of urban population in the 2005 survey is inconsistent with that in the 2000 census. In the 2000 census, urban population refers to the resident population

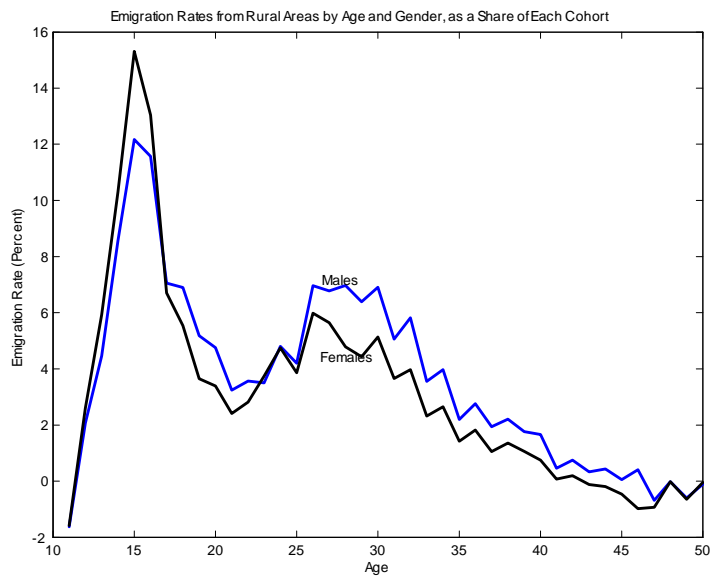


Figure 1: The figure shows rural-urban migration rates by age and gender as a share of each cohort. The estimates are smoothed by 5-year moving averages.

an annual 4.1% rate between 2000 and 2005. Hence, 89% of the Chinese urban population growth during those years appears to be accounted for by rural-urban migration. Our estimates are in line with earlier estimates of the aggregate NMF. For instance, Hu (2003), estimates that the annual NMF ranged between 17.5 and 19.5 millions in the period 1996–2000. Our estimate implies an annual flow of 18.3 million migrants between 2001 to 2005, equal to an annual 2.3% of the rural population.

The estimated age-gender-specific migration rates are shown in Figure ???. Both the female and male migration rates peak at age fifteen, with 16.8% for females and 13.3% for males. The migration rate falls gradually at later ages, remaining above 1% until age thirty-nine for females and until age forty for males. Migration becomes negligible after age forty.

To incorporate rural-urban migration in our population projection, we make two assumptions. First, the age-gender-specific migration rates remain constant after 2005 at the level of our estimates for the period 2000–2005. Second, once the migrants have moved to an urban area, their fertility and mortality rates are assumed to be those of urban residents.

Figure 2 shows the resulting projected population dynamics (solid lines). For comparison, we also plot the natural population dynamics, i.e., the population model without migration (dotted lines). The rural population declines throughout: 263 million rural residents will move to urban areas between

(*changzhu renkou*) of the place of enumeration who had resided there for at least six months on census day. The minimum requirement was removed in the 2005 survey. Therefore, relative to the 2005 survey definition, rural population tends to be over-counted in the 2000 census. This tends to bias our NMF estimates downwards.

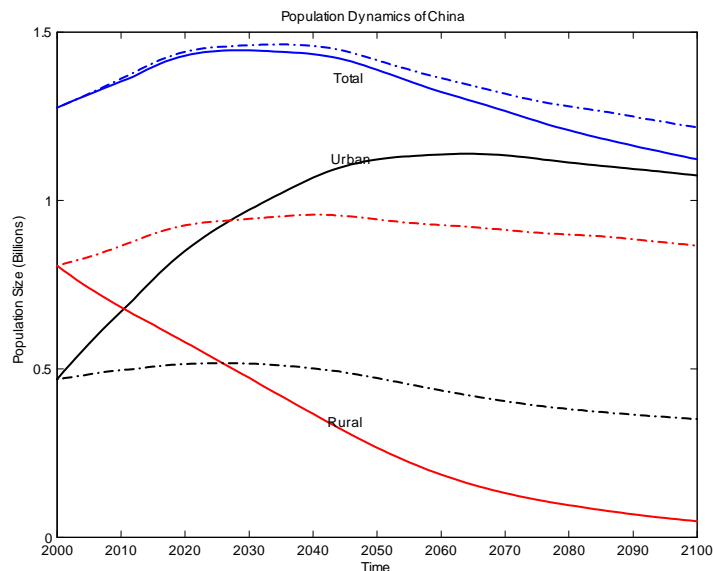


Figure 2: The figure shows the projected population dynamics for 2000-2100 (solid lines) broken down by rural and urban population. The dashed lines show the corresponding natural population dynamics, i.e., the counterfactual projection under a zero urban-rural migration scenario.

2010 and 2050. The urban population share increases from 50% in 2011 to 80% in 2050 and to over 90% in 2100. In absolute terms, the urban population increases from 450 millions in 2000 to its long run 1.2 billion level in 2050. Between 2050-2100 there are two opposing forces that tend to stabilize on net the urban population: on the one hand, fertility is below replacement in urban areas until 2100; on the other hand, there is still sizeable immigration from rural areas. In contrast, if there was no migration in the XXIst Century, the urban population would start declining already in 2008, and it would be a mere third of the total population in 2050.

Figure 3 plots the old-age dependency ratio – i.e., the number of retirees as percentage of individuals in working age (18-60) – broken down by rural and urban areas (solid lines).¹³ We also plot, for contrast, the old-age dependency ratio in the no migration counterfactual (dashed lines). Rural-urban migration is very important for the projection. The projected urban dependency ratio is 50% in 2050, while it would be as high as 80% in the no migration counterfactual. This is an important statistic: The Chinese pension system only covers urban workers, so its sustainability hinges on the urban old-age dependency ratio.

¹³In China, the official retirement age is 55 for females and 60 for males. In the rest of the paper, we ignore this distinction, and assume that all individuals retire at age 60, anticipating that the age of retirement is likely to increase in the near future. We also consider the effect of changes in the replacement ratio.

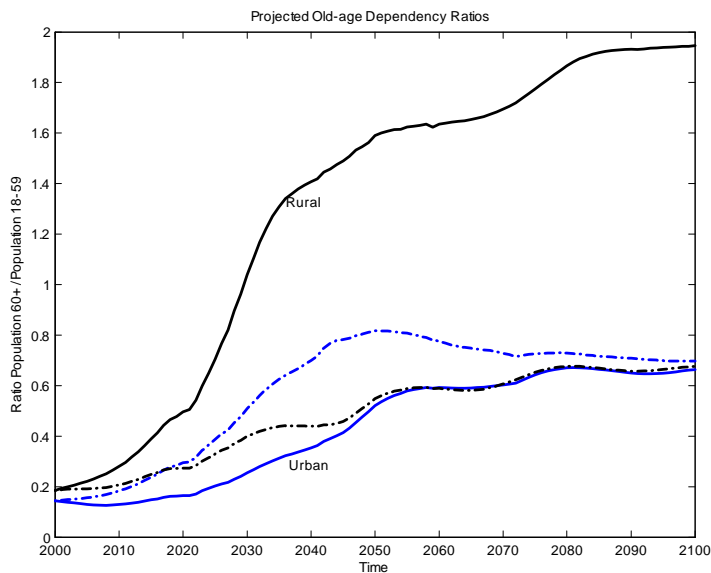


Figure 3: The figure shows the projected old-age dependency ratios, defined as the ratio of population 60+ over population 18-59, for 2000-2100 (solid lines). Blue (black) lines denote urban (rural) dependency ratios. The dashed lines show the corresponding ratios under the natural population dynamics, i.e., under the zero migration counterfactual.

3 A Partial Equilibrium Model

In this section, we construct and calibrate a multiperiod OLG model *à la* Auerbach and Kotlikoff (1987), consistent with the demographic model of section 2. Then, we feed an exogenous wage growth process into the model and use it to assess the welfare effects of alternative sustainable pension reforms. In section 6 we show that the assumed wage process is the equilibrium outcome of a calibrated dynamic general-equilibrium model with credit market imperfections close in spirit to Song *et al.* (2011).

3.1 Households

The model economy is populated by a sequence of overlapping generations of agents. Each agent lives up to $J - J_C$ years and has an unconditional probability of surviving until age j equal to s_j . During their first $J_C - 1$ years (childhood), agents are economically inactive and make no choices. Preferences are defined over consumption and leisure, and represented by a standard lifetime utility function,

$$U_t = \sum_{j=0}^J s_j \beta^j u(c_{t+j}, h_{t+j}),$$

where c is consumption and h is labor supply. Here, t denotes the period when the agent becomes adult, i.e., economically active. Thus, U_t is the discounted utility of an agent born in period $t - J_C$.

Workers earn an hourly wage from age J_C until retirement, which happens at age J_W for all workers. Thereafter, they earn pension benefits until death. Wages are subject to proportional taxes. Adult workers and retirees can borrow and deposit their savings with banks paying a gross annual interest rate R . A perfect annuity market allows agents to insure against the uncertainty about the time of death.

Agents maximize U_t , subject to a lifetime budget constraint:

$$\sum_{j=0}^J \frac{s_j}{R^j} c_{t+j} = \sum_{j=0}^{J_W} \frac{s_j}{R^j} (1 - \tau_{t+j}) \zeta_j \eta_t w_{t+j} h_{t,t+j} + \sum_{j=J_W+1}^J \frac{s_j}{R^j} b_{t,t+j}$$

where $b_{t,t+j}$ denotes the pension accruing in period $t+j$ to a person who became adult in period t , w_{t+j} is the wage rate per efficiency unit at $t+j$, η_t denotes the human capital specific to the cohort turning adult in t (we abstract from within-cohort differences in human capital across workers), and ζ_j is the efficiency units per hour worked for a worker with j years of experience which captures the experience-wage profile.

The government runs a pension system financed by a social security tax levied on labor income and by an initial endowment, A_0 . The government intertemporal budget constraint yields:

$$\sum_{t=0}^{\infty} R^{-t} \left(\sum_{j=J_W+1}^J N_{t-j,t} b_{t-j,t} - \tau_t \sum_{j=0}^{J_W} N_{t-j,t} \zeta_j \eta_{t-j} w_t h_{t-j,t} \right) \leq A_0 \quad (1)$$

where $N_{t-j,t}$ is the number (measure) of agents in period t who became active in period $t-j$.

3.2 The Pension System

The model pension system replicates the main features of China's pension system. The current system was originally introduced in 1986 and underwent a major reform in 1997. Before 1986, urban firms (which at the time were almost entirely state owned) were responsible for paying pensions to their former employees. This system become untenable in an economy where firms can go bankrupt, and workers can change jobs. The 1986 reform introduced a defined benefits system whose administration was assigned to municipalities. The new system came under financial distress, mostly due to firms evading their obligations to pay pension contributions for their workers.

The subsequent 1997 reform tried to make the system more sustainable by reducing the replacement rates for future retirees and by enforcing the social security contributions more strictly. The 1997 system has two tiers. The first is a standard transfer-based system with resource-pooling at the provincial level. The second is an individual accounts system. However, as documented by Sin (2005; p.2) "the individual accounts are essentially 'empty accounts' since most of the cash flow surplus has been diverted to supplement the cash flow deficits of the social pooling account." Given the low

capitalization of the system, it can be regarded as a *de facto* transfer-based system which permits, as does the US Social Security system, the accumulation of a trust fund to smooth the ageing of the population. Therefore, in our analysis we ignore the nominal distinction between the different pension pillars.

We model the pension system as a defined benefits plan, subject to the intertemporal budget constraint, (1). In line with the actual Chinese system, pensions are partly indexed to wage growth. We approximate the benefit rule by a linear combination of the average earnings of the beneficiary at the time of retirement and the current wage of workers about to retire, with weights 60% and 40%, respectively. More formally, the pensions received at period $t + j$ by an agent who retired in period $t + J_W$ (and who became adult in period t) is

$$b_{t,t+j} = q_{t+J_W} \cdot (0.6 \cdot \bar{y}_{t+J_W} + 0.4 \cdot \bar{y}_{t+j-1}), \quad (2)$$

where q_t denotes the replacement rate in period t and \bar{y}_t is the average pre-tax labor earnings for workers about to retire in period t :

$$\bar{y}_t \equiv w_t \eta_{t-J_W} \zeta_{J_W} h_{t-J_W,t}.$$

In line with the 1997 reform (see Sin 2005), we assume that pensioners retiring before 1997 continued to earn a 78% replacement rate throughout their retirement. Moreover, those retiring between 1997 and 2011 are entitled to a 60% replacement ratio.

We assume a constant social security tax (τ) equal to 20%, in line with the empirical evidence.¹⁴ The tax and the benefit rule do not guarantee that the system is financially viable. In fact, we will show that, given our forecasted wage process and demographic dynamics, the current system is not sustainable, so long-run budget balance requires either tax hikes or benefit reductions. In this paper we mainly focus on reducing benefits. As a benchmark (labeled the *benchmark reform*) we assume that in 2012 the replacement rate is lowered permanently to a new level so as to satisfy the intertemporal budget constraint, (1).

The current pension system of China only covers a fraction of the urban workers. The coverage rate has grown from about 40% in 1998 to 57% in 2009.¹⁵ In the baseline model, we assume a constant

¹⁴The statutory contribution rate including both basic pensions and individual account is 28%, of which 20% should be paid by firms and 8% should be paid by workers (see Document 26 issued by the state council, "A Decision on Establishing a Unified Basic Pension System for Enterprise Workers"). However, there is evidence that a significant share of the contributions is evaded, even for workers who formally participated in the system. For instance, in the annual National Business 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 2004-07 amounts to 11% of the average wages, nine percentage points below the statutory rate. In addition, wage appear to be underreported. Most evasion comes from privately owned firms, whose contribution rate is a mere 7%.

¹⁵The coverage rate is equal to the number of employees participated in the pension system divided by the number

coverage rate of 60%. The coverage rate of migrant workers is a key issue. Since we do not have direct information about their coverage, we decided to simply assume that rural immigrants get the same coverage rate as that of urban workers. This seems a reasonable compromise between two considerations. On the one hand, the coverage of migrant workers (especially low-skill non-hukou workers) is probably lower than that of non-migrant urban resident;¹⁶ on the other hand the total coverage has been growing since 1997.

We then consider a set of alternative reforms. First, we assume that the current rules are kept in place until period T (where $T > 2011$), in the sense that the current replacement rate ($q_t = 60\%$) apply for those who retire until period T . Thereafter, the replacement rates are adjusted permanently so as to satisfy (1). Clearly, the size of the adjustment depends on T : since the system is currently unsustainable, a delay requires a larger subsequent adjustment. We label such scenario *delayed reform*.

Next, we consider a reform that eliminates the transfer-based system introducing, as of 2012, a mandatory saving-based pension system. In our stylized model such a FF system is identical to no pension system because agents are fully rational and subject to no borrowing constraints or time inconsistency in their saving decisions. In the *FF reform* scenario, the pension system is abolished in 2012. However, the government does not default its outstanding liabilities: those who are already retired receive a lump sum transfer equal to the present value of the benefits they would have received under the benchmark reform. Moreover, those still working in 2012 are compensated for their accumulated pension rights, scaled by the number of years they have contributed to the system. To cover these lump-sum transfers, the government issues debt. In order to service this debt, the government introduces a new permanent tax on labor earnings, which replaces the (higher) old social security tax.

Next, we consider switching to a pure PAYGO reform system where the tax rate is kept constant at $t=20\%$ and the benefit rate is endogenous and depends on the tax revenue (which is in turn affected by the demographic structure and endogenous labor supply). Finally, we consider two reforms that extend the coverage of the pension system to rural workers. The *moderate* rural reform scenario offers a 20% replacement rate to rural retirees financed by a 6% social security tax on rural workers. Such a rural pension is similarly to a scheme started recently by the government on a limited scale.¹⁷ The

of urban employees. Both numbers are obtained from China Statistical Yearbook 2010. There is a concern that the rapidly growing size of migrant workers might lead to a 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, while 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.

¹⁶In a recent local survey conducted by Shanghai Population and Family planning commission in 2011, only 18% of a total of 24,000 migrants in the sample are covered by the urban pension system.

¹⁷The new program provides a basic pension of RMB55 per month. Since in 2009 the average rural per capita annual net income was RMB5153 (China Statistical Yearbook 2010), this implies a replacement rate of 12.8%. However, provinces and cities are allowed to set higher replacement rates if local governments have the fiscal capacity. For instance, Beijing

radical rural reform scenario introduces a universal pension system with the same benefits and taxes in rural and urban areas.

3.3 Calibration

One period is a year. Agents can live up to 100 years ($J = 100$) and the demographic process (mortality, migration, and fertility) is described in Section 2. Agents become adult (i.e., economically active) at age $J_C = 23$, and retire at age 60, which is the male retirement age in China (so $J_W = 59$). Hence, workers retire after 37 years of work. We set the age-wage profile $\{\zeta_j\}_{j=23}^{59}$ equal to the one estimated by Song and Yang (2011) for Chinese urban workers. This implies an average return to experience of 0.5%. In this section of the paper, we take the hourly wage rate as exogenous. The assumed dynamics of wages per effective unit of labor is shown in Figure 4: Hourly wages (conditional on human capital) grow at approximately 5.7% between 2000 and 2011, 5.1% between 2011 and 2030, and 2.7% between 2030 and 2050. In the long run, wages are assumed to grow at 2% per year, in line with wage growth in the United States over the last century. In section 6, we show that the assumed wage rate dynamics of Figure 4 is the equilibrium outcome of a calibrated version of the model of Song *et al.* (2011).

There has been substantial human capital accumulation in China over the last two decades. To incorporate this, we assume that each generation has a cohort-specific education level, which is matched to the average years of education by cohort according to Barro and Lee (2010) (see Figure 13 in the Appendix). The values for cohorts born after 1990 are extrapolated linearly, assuming the growth in the years of schooling ceases in year 2000 when it reaches an average twelve years, which is the current level for the US. We assume an annual return of 10% per year of education.¹⁸ Since younger cohorts have more years of education, wage growth across cohorts will exceed that shown in Figure 4. However, the education level for an individual remains constant over his/her worklife, so Figure 4 is the relevant time path for the individual wage growth.

The rate of return on capital is very large in China (see e.g. Bai *et al.* 2006). However, these high rates of returns are arguably inaccessible to the government and to the vast majority of workers and retirees. Indeed, in addition to housing and consumer durables, bank deposits is the main asset for saving for Chinese households. For example, in 2002 more than 68% of households' financial assets

and Shanghai have set higher pension benefits (RMB280 in Beijing and RMB150-300 in Shanghai). Since the average 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%. We set the replacement rate to 20% to match the average of the basic level of 12.8% and the high level of 27.2%. The new program asks rural residents to contribute 4% to 8% of the local average income per capita in the previous year. We then set the contribution rate to 6%.

¹⁸Zhang et al. (2005) estimated returns to education in urban areas of six provinces from 1988 to 2001. The average returns were 10.3% in 2001.

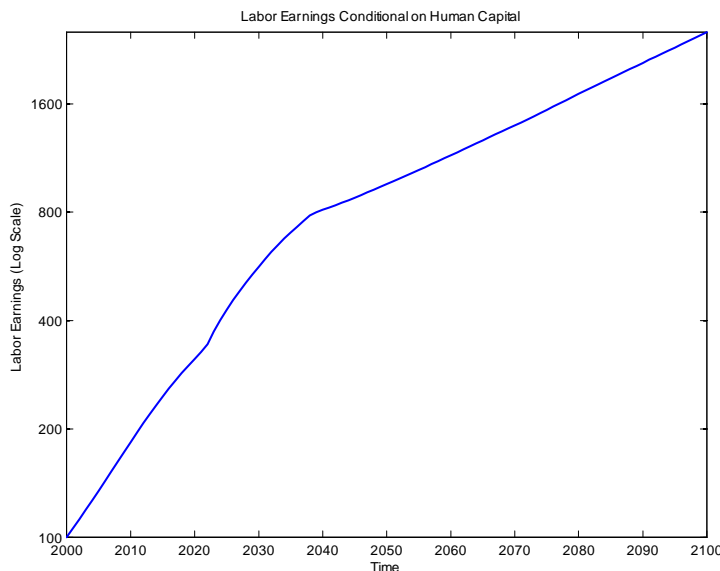


Figure 4: The figure shows the projected hourly wage rate per unit of human capital in urban areas, normalized to 100 in 2000. The process is the endogenous outcome of the general equilibrium model of section 6.

were held in terms of bank deposits and bonds, and for the median decile of households this share is 75% (source: CHIP 2002). Moreover, aggregate household deposits in Chinese banks amounted to 76.6% of GDP in 2009 (source: CSY 2010). High rates of return on capital do not appear to have been available to the government either.¹⁹

Building on Song *et al.* (2011), the model of Section 6 provides an explanation – based on large credit market imperfections – for why neither the government nor the workers have access to the high rates of return of private firms. In this section, we simply assume that the annual rate of return for private and government savings is $R = 1.025$. This is slightly higher than the empirical one-year real deposit rate in Chinese banks, which was 1.75% during 1998-2005 (nominal deposit rate minus CPI inflation). The choice of 2.5% per year is in our view a conservative benchmark, and reflects the possibility for some households to access to savings instruments that yields higher return. Moreover, this rate of return seems like a reasonable long-run benchmark as China becomes a developed country.²⁰

¹⁹[preliminary] The balance sheet of the Chinese government consists mainly of three items: foreign government bonds (XXX60% of GDP in 2009), foreign reserves GDP ratio is 48% in 2009 (CSY, 2010) ownership of SOEs, and RMB-denominated debt (XXX55% of GDP in 2009). Government debt GDP ratio is 17.7% in 2009 (CSY, 2010). In addition, the government has some small amounts in investment funds (4.8% of GDP in 2009, CSY 2010). As documented in Dollar and Wei (2007), the rate of return on capital in SOEs is substantially lower than the average rate of return in the economy. We conclude that the relevant marginal rate of return on government savings is the world-market rate of return on government bonds.

²⁰Assuming a very low R would also imply that the rate of return is lower than the growth rate of the economy, implying dynamic inefficiency. In such a scenario, there would be no need for a pension reform due to a well-understood mechanism (cf. Abel et al. 1989).

Consider, finally, preference parameters: The discount factor is set to $\beta = 1.0175$ to capture the large private savings in China. This is slightly higher than the value (1.011) that Hurd (1989) estimated for the United States. As a robustness check we also consider an alternative economy where β is lower for all people born after 2012 (see Section 5). In Section 6 we document that with $\beta = 1.0175$ the model economy matches China's average aggregate saving rate during 2000-2010.

We assume that preferences are represented by the following standard utility function

$$u(c, h) = \log c - h^{1+\frac{1}{\phi}},$$

where ϕ is the Frisch elasticity of labor supply. We set $\phi = 0.5$, in line with standard estimates in labor economics (Keane, 2011). Note that both the social security tax and pensions in old age distort labor supply.

Finally, we obtain the initial distribution of wealth in year 2000 by assuming that all agents alive in 1992 had zero wealth (since China's market reforms started in 1992). Given the 1992 distribution of wealth for workers and retirees, we simulate the model over the 1992-2000 period assuming an annual wage growth of 5.7%, excluding human capital growth. The distribution of wealth in 2000 is then obtained endogenously. The initial government wealth in 2000 is set to 71% of GDP. As we explain in detail below this is consistent with the observed foreign surplus in year 2000 given the calibration of the general equilibrium model in section 6.

4 Results

Under our calibration of the model, the current pension system is not sustainable. In other words, the intertemporal budget constraint, (1), would not be satisfied if the current rules were to remain in place forever. For the intertemporal budget constraint to hold, it is necessary either to reduce pension benefits, or to increase contributions.

4.1 The benchmark reform

We define as the *benchmark reform* a pension scheme such that: (i) the existing rules apply to all cohorts retiring earlier than 2012; (ii) the social security tax is set to a constant $\tau = 20\%$ for all cohorts; (iii) the replacement rate q which applies to all individuals retiring after 2011 is set to the highest constant level consistent with the intertemporal budget constraint, (1). All households are assumed to anticipate the benchmark reform.²¹

²¹When we consider alternative policy reforms below, we introduce them as "surprises", i.e., agents expect the benchmark reform, but then, unexpectedly, a different reform occurs. After the surprise, perfect foresight is assumed. This assumption is not essential. The main results of this section are not sensitive to different assumption, such as assum-

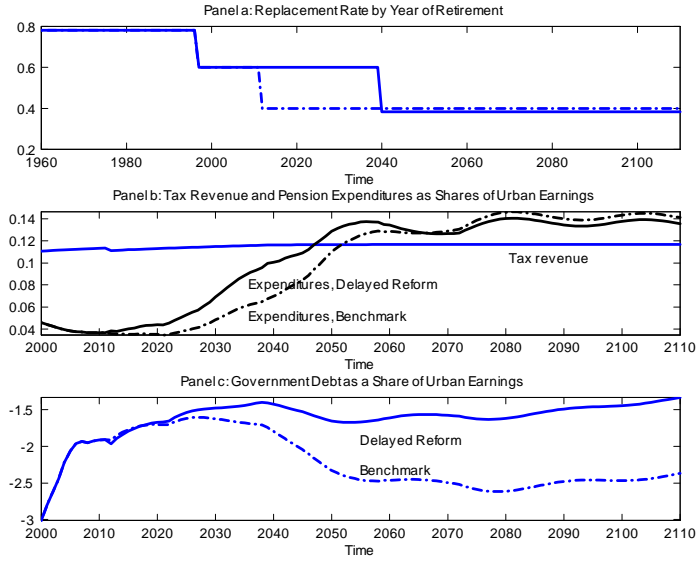


Figure 5: Panel (a) shows the replacement rate q_t for the benchmark reform (dashed line) versus the case when the reform is delayed until 2040. Panel (b) shows tax revenue (blue) and expenditures (black), expressed as a share of aggregate urban labor income (benchmark reform is dashed and the delay-until-2040 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-2040 is solid). Negative values indicate surplus.

The benchmark reform entails a large reduction in the replacement rate, from 60% to 40%. Namely, pensions must be cut by a third in order for the system to be financially sustainable. Such an adjustment is consistent with the existing estimates of the World Bank (see Sin 2005, p.30). Alternatively, if one were to keep the replacement ratio constant at the initial 60% and to increase taxes permanently so as to satisfy (1), then τ should increase from 20% to 30.1% as of year 2012.

Figure 5 shows the evolution of the replacement rate by cohort under the benchmark reform (panel a, dashed line). The replacement rate is 78% until 1997, and then falls to 60%. Under the benchmark reform, it falls further to 40% in 2012 remaining constant thereafter. Panel b (dashed line) shows that such a reform implies that the pension system runs a surplus until 2051. The government builds up a government trust fund amounting to 261% of urban labor earnings by 2080 (panel c, dashed line). The interests earned by the trust fund are used to finance the pension system deficit after 2051.²²

ing that all reforms (including the benchmark reform) come as a surprise, or assuming that all reforms are perfectly anticipated.

²²Note that in panel c the government net wealth (i.e., minus the debt) is falling sharply between 2000 and 2020 when expressed as a share of urban earnings, even though the government is running a surplus. This is due to the fact that urban earnings is rising very rapidly due to both high wage growth and growth in the number of urban workers.

4.2 Alternative reforms

Having established that a large adjustment is necessary to balance the pension system, we address the question of whether the reform should be implemented urgently (as suggested, e.g., by Feldstein (1999)), or whether it could be deferred. In addition, we consider two more radical alternative reforms: a move to a FF, pure contribution-based system, and a move in the opposite direction to a pure PAYGO system.

We compare the welfare effects of each alternative reform by measuring, for each cohort, the equivalent consumption variation of each alternative reform relative to the benchmark reform. Namely, we calculate what (percentage) change in lifetime consumption would make agents in each cohort indifferent between the benchmark and the alternative reform.²³ We also aggregate the welfare effects of different cohorts by assuming a social welfare function based on a utilitarian criterion, where the weight of the future generation decay at a constant rate ϕ . More formally, the planner's welfare function (evaluated in year 2012) is given by:

$$U = \sum_{t=1935}^{\infty} \phi^t \sum_{j=0}^J \beta^j u(c_{t,t+j}, h_{t,t+j}). \quad (3)$$

Then, the equivalent variation is given by the value ω solving

$$\sum_{t=1935}^{\infty} \sum_{j=0}^J \beta^j u((1 + \omega) c_{t,t+j}^{BENCH}, h_{t,t+j}^{BENCH}) = \sum_{t=1923}^{\infty} \phi^t \sum_{j=0}^J \beta^j u(c_{t,t+j}^*, h_{t,t+j}^*), \quad (4)$$

where superscripts *BENCH* stand for the allocation in the benchmark reform and stars stand for the allocation in the alternative reform.²⁴

The planner experiences a welfare gain (loss) from the alternative allocation whenever $\omega > 0$ ($\omega < 0$). We shall consider two particular values of the intergenerational discount factor, ϕ . First, $\phi = R$, i.e., the planner discounts future utilities at the market interest rate, as suggested, e.g., by Nordhaus (2007). We label such a planner as the *high-discount planner*. Second, $\phi = R/(1 + g)$, where g is the long-run wage growth rate (recall that in our calibration $R = 1.025$ and $g = 0.02$). Such a lower intergenerational discount rate is an interesting benchmark, since it implies that the planner would not want to implement any intergenerational redistribution in the steady state. We label a planner endowed with such preferences as the *low-discount planner*.

²³Note that we measure welfare effects relative to increases in *lifetime* consumption even for people who are alive in 2012. This makes it easier to compare welfare effects across generations.

²⁴Note that we sum over agents alive or yet unborn in 2012. The oldest person alive became an adult in 1935, which is why the summations over cohorts indexed by t start from 1935.

4.2.1 "Delayed" Reform

We start by evaluating the welfare effects of delaying the reform. Namely, we assume that the current replacement rate remains in place until some future date T , when a reform similar to the benchmark reform is conducted (i.e., the system provides a lower replacement rate which remains constant for ever). A delay has two main effects: On the one hand, the generations retiring shortly after 2012 receive higher pensions which increase their welfare. On the other hand, the fund accumulates a lower surplus between 2012 and the time of the reform, making necessary an even larger reduction of the replacement rate thereafter. Thus, the delay shifts the burden of the adjustment from the current (poorer) generations to (richer) future generations.

Figure 5 describes the positive effects of delaying the reform until 2040. Panel (a) shows that the post-reform replacement rate falls now to 38.4%, which is only 1.6 percentage points lower than the replacement rate granted by the benchmark reform. Panel (b) shows that the pension expenditure is higher than in the benchmark reform until 2066. Moreover, the system starts running a deficit already in 2048. As a result, the government accumulates a smaller trust fund during the years in which the dependency ratio is low. The reason why the difference in the replacement rate is small, is threefold. First, the urban working population continues to grow until 2040, due to internal migration. Second, wage growth is high between 2012 and 2040. Third, the trust fund has only access to a 2.5% interest rate, well below the average wage growth. The second and third factor, which are exogenous in this section, will be derived as the endogenous outcome of a calibrated general equilibrium model with credit market imperfections in section 6.

Consider, next, deferring the reform until 2100 (see Figure 6). In this case, the pension system starts running a deficit as of year 2043 (panel b). The deficit grows fast thereafter, and the government debt reaches 200% of the aggregate urban labor earnings in 2094. Consequently, a sizeable adjustment is required in 2100: the replacement rate must fall to 29.7% to balance the intertemporal budget (panel a).

Figure 7 shows the equivalent variations, broken down by the year of retirement for each cohort. Panel (a) shows the case in which the reform is delayed until 2040. The consumption equivalent gains for agents retiring between 2012 and 2039 are large: on average over 17% of their lifetime consumption! The main reason is that delaying the reform enables the transition generation to share the gains from high wage growth after 2012, to which pension payments are (partially) indexed. The welfare gain declines over the year of cohort retirement, since wage growth slows down. Yet, the gains of all cohorts affected are large, being bounded from below by the 15.5% gains of the generation retiring in 2039. On the contrary, all generations retiring after 2039 lose, though their welfare losses are quantitatively

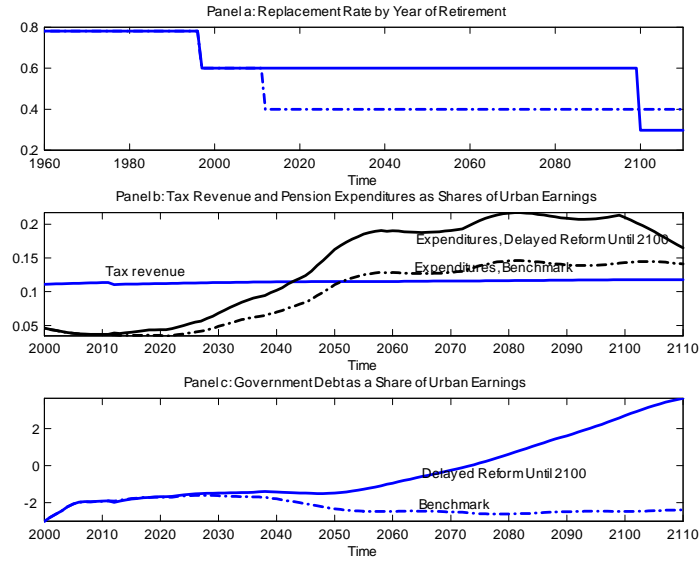


Figure 6: Panel (a) shows the replacement rate q_t for the case when the reform is delayed until 2100 (solid line) versus the benchmark reform (dashed line). Panel (b) shows tax revenue (blue) and expenditures (black), expressed as a share of aggregate urban labor income (benchmark reform is dashed and 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.

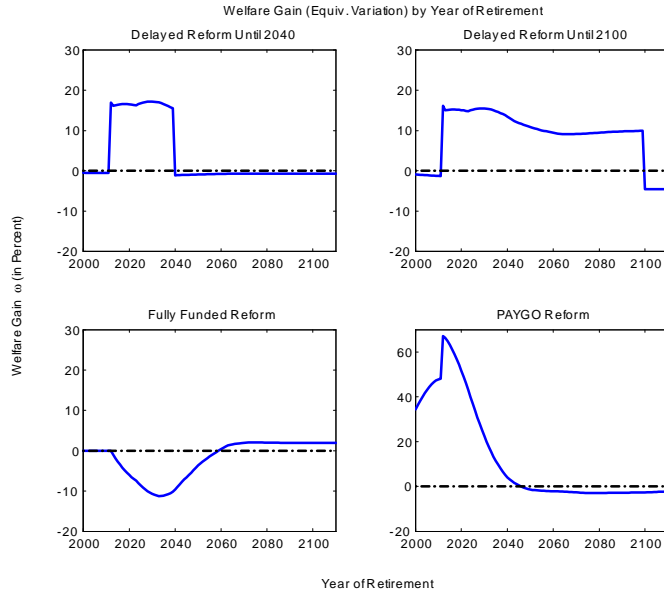


Figure 7: The graph shows welfare gains of alternative reforms relative to the benchmark reform for each cohort. The gains (ω) are expressed as percentage increase in consumption (see eq. 4).

small, being less than 1.1% of their lifetime consumption. The difference between the large welfare gains accruing to the first twenty-nine cohorts and the small losses suffered by later cohorts is stark.

A similar trade off can be observed in panel (b) for the case in which the reform is delayed until 2100. In this case, sizeable gains accrue to a larger number of cohorts. As in the previous case, the welfare gains decline over cohorts, falling below 10% for all generations retiring after 2045. The losses accruing to the future generations are now significantly larger. All agents retiring after 2100 suffer a loss equivalent to 4.6% of their lifetime consumption.

Figure 8 shows the welfare gains/losses of delaying the reform until year T , according to the utilitarian social welfare function. The figure displays two curves: In the upper curve, we have the consumption equivalent variation of the high-discount planner, while in the lower curve we have that of the low-discount planner.

Consider, first, delaying the reform until 2040. The delayed reform yields $\omega = 5\%$ for the high-discount planner, i.e., the delayed reform is equivalent to a permanent 5% increase in consumption in the benchmark allocation. The gain is partly due to the fact that future generations are far richer, and, hence, have a lower marginal utility of consumption. For instance, in the benchmark reform scenario the average pension earned by an agent retiring in 2050 is 5.28 times larger than that of an agent retiring in 2012. Thus, delaying the reform has a strong equalizing effect that increases the utilitarian planner's utility. The welfare gain of the low-discount planner remains positive, albeit smaller, $\omega = 0.8\%$.

The figure shows that the high-discount planner would maximize her welfare gain by a long delay of the reform (the curve is uniformly increasing in the range shown in the figure, and reaches a maximum in year 2480. In contrast, the low-discount planner would maximize her welfare gain by delaying the reform until year 2049.

4.2.2 Fully Funded Reform

Consider, next, switching to a FF system, i.e., a pure contribution-based pension system featuring no intergenerational transfers, where agents are forced to save for their old age in a fund which has access to the same rate of return to which private savers have access. As long as agents are rational and have time-consistent preferences, and mandatory savings do not exceed the savings that agents would make privately in the absence of a pension system, a FF system is equivalent to no pension system. However, switching to a FF system does not cancel the outstanding liabilities, i.e., payments to current retirees and entitlements of workers who have already contributed to the system. We will therefore design a reform such that the government does not default on existing claims. In particular,

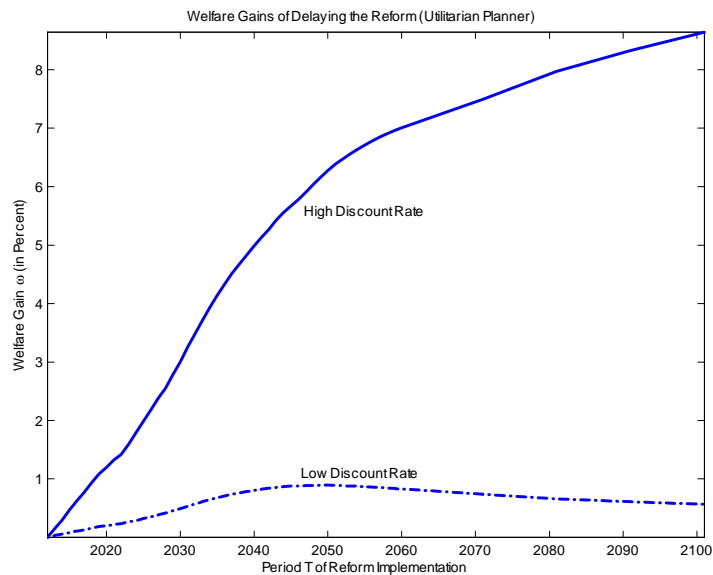


Figure 8: The figure shows the consumption equivalent gain/loss accruing to a high-discount planner (solid lines) and to a low-discount planner (dashed lines) of delaying the reform until time T relative to the benchmark reform. When $\omega > 0$, the planner strictly prefers the delayed reform over the benchmark reform.

we assume that all workers and retirees who have contributed to the pension system are refunded the present value of the pension rights they have accumulated.²⁵ Since the social security tax is abolished, the existing liabilities are financed by issuing government debt, which in turn must be serviced by a new tax.

Figure 9 shows the outcome of this reform. The old system is terminated in 2011, but people with accumulated pension rights are compensated as discussed above. To finance such pension buy-out scheme, government debt must increase to over 87% of total labor earnings in 2011. A permanent 0.3% annual tax is needed to service such a debt. The government debt first declines as a share of total labor earnings, due to high wage growth in that period, and then stabilizes at a level about 30% of labor earnings around 2040. Agents born after 2040 live in a low-tax society with no intergenerational transfers.

Panel (c) of Figure 7 shows the welfare effects of the FF reform relative to the benchmark. The welfare effects are now opposite to those of the delayed reforms. The cohorts retiring between 2012 and 2058 are harmed by the FF reform relative to the benchmark. There is no effect on earlier generations, since those are fully compensated by assumption. The losses are also modest for cohorts

²⁵In particular, people who have already retired are given an asset worth the present value of the pensions according to the old rules. Since there are perfect annuity markets, this is equivalent for those agents to the pre-reform scenario. People who are still working and have contributed to the system are compensated in proportion to the number of years of contributions.

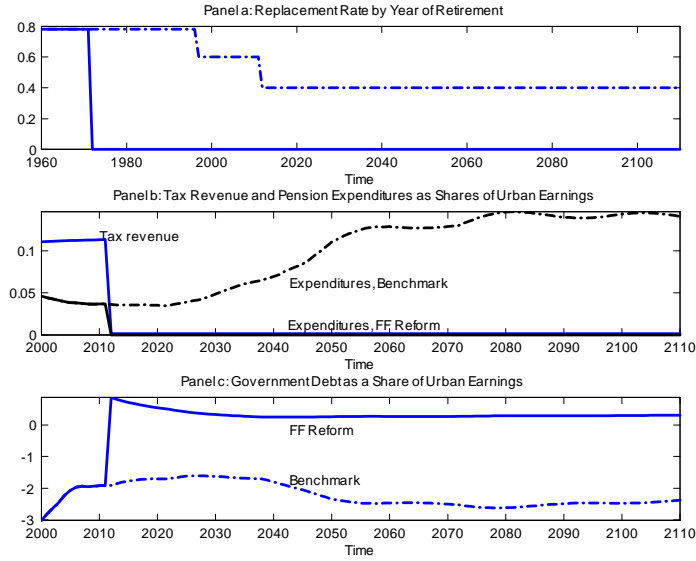


Figure 9: The figure shows outcomes for the fully funded reform (solid lines) versus the benchmark reform (dashed lines). Panel (a) shows the replacement rate, Panel (b) shows taxes (blue) and pension expenditures (black) expressed as a share of aggregate urban labor income, and Panel (c) the government debt as a share of aggregate urban labor income.

retiring soon after 2012, since these have earned almost full pension rights by 2012. However, the losses increase for later cohorts and become as large as 11% for those retiring in 2030-35. For such cohorts, the system based on intergenerational transfer is attractive, since wage growth is high during their retirement age (implying fast-growing pensions), whereas the returns on savings are low. Losses fade away for cohorts retiring after 2050, and turn into gains for those retiring after 2058. The fact that generations retiring sufficiently far in the future gain is guaranteed by the assumption that the economy is dynamically efficient. However, the long-run gains are modest. The high-discount planner strictly prefers the benchmark over the FF reform, the consumption equivalent discounted loss being 3.5%. In contrast, the low-discount planner makes a 0.2% consumption equivalent gain. This small gain arises from the labor supply adjustment triggered by the lower tax distortion. If labor supply were inelastic, even the low-discount would lose by moving to a fully-funded system.

4.2.3 Pay as you go reform

We now analyze the effect of moving to a pure PAYGO. In particular, we let the contribution rate be fixed at $\tau = 20\%$ and assume that each year the benefits equal the total contributions. Therefore, the

pension benefits b_t in period t are endogenously determined by the following formula:²⁶

$$b_t = \frac{\tau \sum_{j=0}^{J_W} N_{t-j,t} \zeta_j \eta_{t-j} w_t h_{t-j,t}}{\sum_{j=J_W+1}^J N_{t-j,t}}.$$

Figure 10 shows the outcome of this reform. Panel (a) reports the pension benefits as a fraction of the average earnings by year. Note that this notion of replacement rate is different from that used in the previous experiments (panel a of Figures 5 to 6); there the replacement rate was cohort specific and was computed according to equation (2) by the year of retirement of each cohort. Until 2050, the PAYGO reform implies larger average pensions than under the benchmark reform.

Panel (b) shows the lifetime pension as a share of the average wage in the year of retirement, by cohort. This is also larger than in the benchmark reform until the cohort retiring in 2044. We should note that, contrary to the previous experiments which were neutral *vis-a-vis* cohorts retiring before 2012, here even earlier cohorts benefit from the PAYGO reform, since the favorable demographic balance yields them higher pensions than what they had been promised. This can be seen clearly in panels (b) and (c). Welfare gains are very pronounced for all cohorts retiring before 2044, especially so for those retiring in 2012 and in the few subsequent years, who would suffer a significant pension cut in the benchmark reform. These cohorts retire in times when the old-age dependency ratio is still very low, and therefore would benefit the most from a pure PAYGO system. On the other hand, generations retiring after 2045 suffer a loss relative to the benchmark reform.

Due to the strong redistribution in favor of poorer early generations, the utilitarian welfare is significantly higher under the PAYGO reform than in the benchmark reform, for both a high- and low-discount planner. The consumption equivalent gains relative to the benchmark reform are, respectively, 13.5% and 1.8% for urban workers. These gains are larger than under all alternative reforms (including delayed and FF reform). These results underline that the gains for earlier generations come at the expenses of only small losses for the future generations.

4.2.4 Increasing retirement age

An alternative to reducing pension benefits would be to increase the retirement age. Our model allows one to calculate the increase in retirement age that would be required to balance the intertemporal budget, (1), given the current social security tax and replacement rate. We find such an increase to be equal to approximately six years, i.e., retirement age would have to increase from 60 to 66 years.

²⁶Note that the pension system has accumulated some wealth before 2011. We assume that this wealth is rebated to the workers in a similar fashion as the implicit burden of debt was shared in the fully funded experiment. In particular, the government introduces a permanent reduction δ in the labor income tax, in such a way that the present value of this tax subsidy equals the 2011 accumulated pension funds. In our calibration, we obtain $\delta = 0.54\%$.

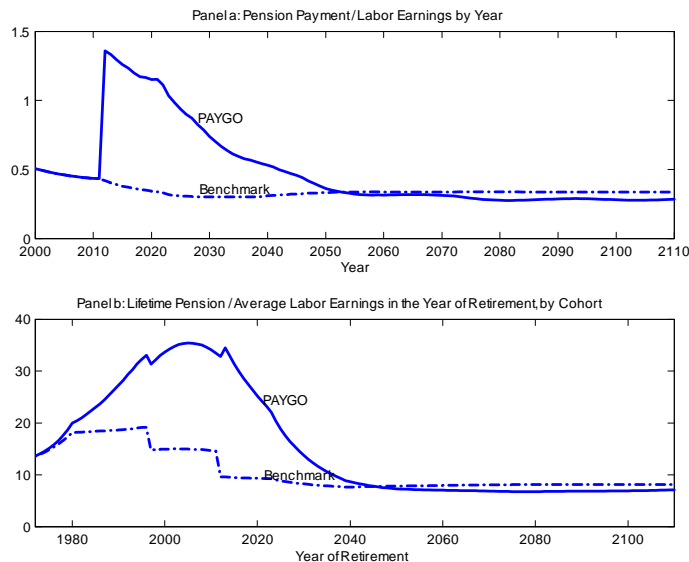


Figure 10: Panel (a) shows the average pension payments in year t as a share of average wages in year t for the PAYGO (solid) and the benchmark reform (dashed line). Panel (b) shows the ratio of the lifetime pensions (discounted to the year of retirement) to the average labor earnings just before retirement for each cohort.

This shows that a draconian reduction in pension entitlements may not be necessary if the retirement age can be increased.

Our model misses important dimensions of the labor supply decision, such as declining health and productivity at a late age and non-convexities in labor supply that could justify a retirement decision (see, e.g., Rogerson and Wallenius 2011). Therefore, we do not emphasize the welfare effects of policies affecting retirement age.

4.2.5 Rural Pension

The vast majority of people living in rural areas are not covered by the current Chinese pension. In accordance with this fact, we have so far maintained the assumption that only urban workers are part of the pension system. In this section, we consider extending the system to rural workers.

While a rural and an urban pension system could in principle be separate programs, we assume that there is a consolidated intertemporal budget constraint, namely, the government can transfer funds across the rural and urban budget. This is consistent with the observation that the modest rural pension system that China is currently introducing is heavily underfunded (see footnote 27), suggesting that the government implicitly anticipates a resource transfer from urban to rural areas.

The modified consolidated government budget constraint then becomes:

$$A_0 + \sum_{t=0}^{\infty} R^{-t} \left(\sum_{j=0}^{J_W} \zeta_j [\tau_t N_{t-j,t} w_t h_{t-j,t} + \tau_t^r N_{t-j,t}^r w_t^r h_{t-j,t}^r] - \sum_{j=J_W+1}^J [N_{t-j,t} b_{t-j,t} + N_{t-j,t}^r b_{t-j,t}^r] \right) \geq 0, \quad (5)$$

where superscripts r denote variables pertaining to the rural areas while urban variables are defined, as above, without any superscript. We assume that the rural wage rate is 54% of the urban wage, consistent with the empirical observation since 2000 (source: China Health and Nutrition Survey).

We consider two experiments. In the first (*low-scale reform*), we introduce in 2012 a rural pension system with different rules from those applying to urban areas. This experiment mimics the rule of the new old-age programs that the Chinese government is currently introducing for rural areas.²⁷ The replacement rate is $q_t^r = 20\%$ and the contribution rate is $\tau_t^r = 6\%$. These rates are assumed to remain constant forever. Moreover, we assume that all rural inhabitants older than retirement age are eligible for this pension already in 2012. The introduction of such a scheme in 2012 is the source of a fiscal imbalance. Restoring the balance through a reform in 2012 requires a larger cut in the replacement rate of urban workers to $q_t = 38.8\%$, which is 1.2 percentage points lower than in the benchmark reform without rural pensions. Hence, the rural pension implies a net transfer from urban to rural inhabitants.

A low-discount planner who only cares for urban households participating in the pension system would incur a welfare loss of less than 0.57% from expanding the pension system to rural inhabitants. In contrast, a low-discount planner who only cares for rural households would incur a welfare gain of 11.9%. When weighting rural and urban households by their respective population shares one obtains an aggregate welfare gain of 2% relative to the benchmark reform.²⁸

The second experiment (*drastic reform*) consists of turning the Chinese pension system universal,

²⁷This benchmark version of a prospective rural pension is motivated by two observations. On the one hand, China has already put in place a new nationwide program paying a basic pension of RMB55 (\$8.7) per month (“Instructions on New Rural Pension Experiments,” State Council, 2009). This corresponds to an average replacement rate of approximately 9% of the average rural wage. However, provinces are allowed to choose more generous rural pensions. For example, Beijing and Shanghai are paying lump-sum rural pensions of RMB280 and RMB 150-300, respectively (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). This amounts to replacement rates of approximately 19% of the rural wages in these provinces.

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

²⁸A high-discount planner who only cares for urban households participating in the pension system would incur a welfare loss of less than 0.64% from expanding the pension system to rural inhabitants. A high-discount planner who only cares for rural households would incur a welfare gain of 12.4%. When weighting rural and urban households by their respective population shares one obtains an aggregate welfare gain of 2% relative to the benchmark reform.

pooling all Chinese workers and retirees – in both rural and urban areas – into a system with common rules. As of 2012, all workers contribute 20% of their wage. In addition, the system bails out all workers who did not contribute to the system in the past. Namely, all workers are paid benefits according to the new rule even though they had not made any contribution in the past. While rural and urban retirees have the same replacement rate, pension benefits are proportional to the group-specific wages, i.e., rural (urban) wages for rural (urban) workers. As in the benchmark reform above, the replacement rate is adjusted in 2012 so as to satisfy the intertemporal budget constraint of the universal pension system. Although we ignore issues with the political and administrative feasibility of such a radical reform, this experiment provides us with an interesting upper bound of the effect of a universal system.

The additional fiscal imbalance from turning the system universal is limited: The replacement rate must be reduced to $q_t = 38.7\%$ from 2012 and onwards, relative to 40% in the benchmark reform. The welfare loss for urban workers participating into the system is very limited – the high-discount planner would suffer a 0.53% loss relative to the benchmark (only marginally higher than in the low-scale reform). In contrast, the welfare gains for rural workers are very large (+23.5% if evaluated by the high-discount planner). Urban workers not participating in the system would also gain substantially (+13.4% if evaluated by the high-discount planner). The average effect (assessed from the standpoint of the high-discount planner weighting equally all inhabitants) is 8.1%.

To understand why this reform can give so large gains with such a modest additional fiscal burden, it is important to emphasize that (i) the earnings of rural workers are on average much lower than those of urban workers; (ii) the rural population is declining fast over time. Both factors make pension transfers to the rural sector relatively inexpensive. It is important to note that our calculations ignore any cost of administering and enforcing the system. In particular, the benefit would decrease if the enforcement of the social security tax in rural areas proved more difficult than in urban areas.

5 Sensitivity analysis:

In this section, we study how the main results of the previous section depend on key assumptions about structural features of the model economy: wage growth, population dynamics and interest rate. We focus for simplicity on the urban pension system (no payments to rural workers). We refer to the calibration of the model used in the previous section as the *baseline economy*.

5.1 Low wage growth

In this section, we consider a low wage growth scenario. In particular, we assume wage growth to be constant and equal to 2%. In this case, the benchmark reform implies a replacement rate of 40.5%. Note that in the low wage growth economy the present value of the pension payments is lower than in the baseline economy, since pensions are partially indexed to the wage growth. Thus, pensions are actually lower, in spite of the slightly higher replacement rate.

Consider, next, the welfare effects of the alternative reforms. The top-left panel of Figure 11 plots the welfare gains/losses of generations retiring between 2000 and 2110 in the case of a delay of the reform till 2040 (dashed line) and 2100 (continuous line). The top-center and top-right panels of Figure 11 yield the welfare gains/losses in the case of a FF reform (center) and PAYGO (right). Recall that gains and losses are expressed relative to the benchmark reform, and thus a cohort gains (loses) when the curve is above (below) unity.

Delaying the reform until 2040 (2100) yields a replacement rate of 40.5% (38.4%). The welfare gains of the earlier generations relative to the benchmark reform are significantly smaller than in the baseline economy. For instance, if the reform is delayed until 2040 the cohorts retiring between 2012 and 2039 experience a consumption equivalent welfare gain ranging between 8% and 9%. The cost imposed on the future generations is similar in magnitude to that of the baseline economy. The high-discount planner enjoys a consumption equivalent gain of 2.4%, which is significantly lower than the 5% gain found in the baseline economy. In the case of the low-discount planner, the gain almost 0. Thus, more than half of the welfare gains of delaying the reform accrues due to the high wage growth. In the alternative of a delayed reform until 2100, the high-discount planner enjoys a welfare gain of less than 5.6%, compared with 8.6% in the baseline economy. Moreover, the low-discount planner now prefers the benchmark reform over a reform delayed until 2100.

As in the baseline case, the FF alternative reform harms earlier cohorts, whereas it benefits all cohorts retiring after 2046. However, the relative losses of the earlier cohorts are significantly smaller than in the baseline economy. For instance, the cohort which is most negatively affected by the FF reform suffers a loss of 3.9% in the low wage growth economy, compared to a 11.3% loss in the baseline economy. Accordingly, the high-discount planner suffers a smaller welfare loss (0.5%) than in the baseline economy (3.5%). Thus, about 85% of the loss accruing to the utilitarian planner arises from the high implicit return of intergenerational transfers due to high wage growth in the baseline economy. Interestingly, the low-discount planner would now prefer the FF reform over any of the alternatives. She would also prefer no delay to any of the delayed reforms.

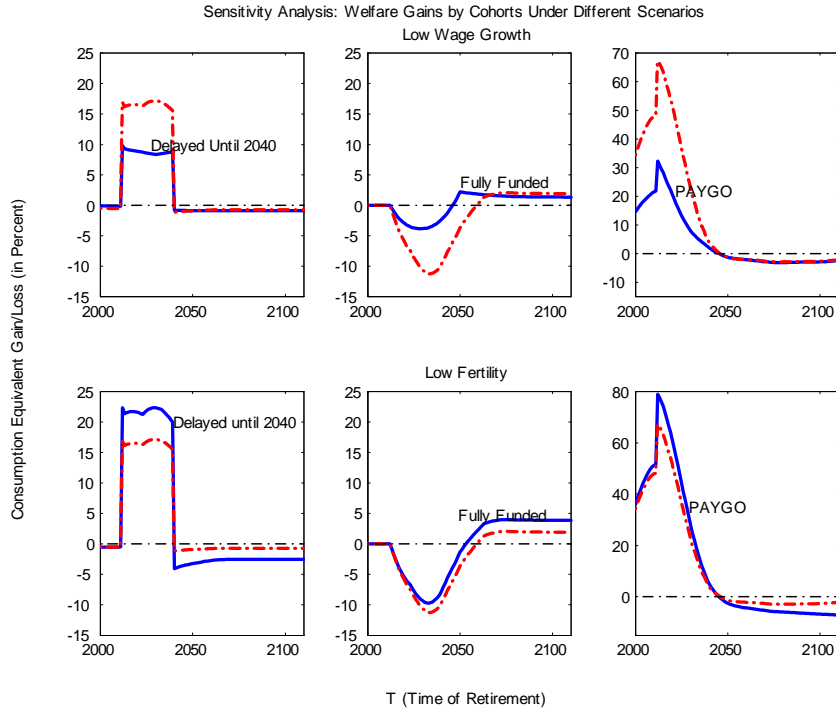


Figure 11: The figure shows consumption equivalent gains/losses accruing to different cohorts in two alternative scenarios. The top panels refer to the low wage growth scenario of section 5.1. The bottom panels refer to the low fertility scenario of section 5.2. In each panel, the dashed red lines refer to the welfare gains under the benchmark calibration (see Section 4). The left-hand panels show the consumption equivalent gains/losses associated with delaying the reform until 2040 (solid blue lines). The center panels show the consumption equivalent gains/losses associated with a fully funded reform (solid blue lines). The right-hand panels show the consumption equivalent gains/losses associated with a PAYGO reform (solid blue lines).

Finally, the large welfare gains from the PAYGO alternative reform by and large vanish. While the high-discount planner would still prefer the PAYGO reform to the benchmark reform, the consumption equivalent gain would be about a third than in the high growth scenario. Perhaps more interesting, the low discount planner who has no built-in preference for transfers to the earlier generations at a given interest rate would now prefer the benchmark reform to the PAYGO reform. Thus, the welfare ranking order of the low discount planner is: FF reform first, then benchmark reform, and last PAYGO reform.

In summary, high wage growth magnifies the welfare gains of delaying a reform (or of switching to PAYGO) and increases the welfare costs of a FF reform relative to the benchmark reform. This is not unexpected since high wage growth increases the implicit return of a system based on intergenerational transfers. The comparison with a constant 2% wage growth scenario is especially revealing since it is consistent with the standard assumption for pension analysis of developed economies.

5.2 Lower fertility

Our forecasts are based on the assumption that the TFR will increase to 1.8 already in 2012. This requires a reform or a lenient implementation of the current one-child policy rules. In this section, we consider an alternative lower fertility scenario along the lines of scenario 1 in Zeng (2007). In this case, the TFR is assumed to be 1.6 forever, implying an ever-shrinking total population. We view this as a lower bound to reasonable fertility forecasts. Consider, next, the welfare effects of the two alternative reforms. The three bottom panels of Figure 11 plot the welfare gains/losses of generations retiring between 2000 and 2110 in the case of a delayed, FF reform and PAYGO respectively.

Under this low-fertility scenario, the benchmark reform requires an even more draconian adjustment. The replacement rate must be set equal to 35.6% as of 2012. Delaying the reform is now substantially more costly. A reform in 2040 requires a replacement rate of 29.8%, whereas a reform in 2100 requires a negative replacement rate of -45.7%. The trade off between current and future generations becomes sharper than in the baseline economy. Consider delaying the reform until 2040. On the one hand, there are larger gains for the cohorts retiring between 2012 and 2039 relative to the benchmark reform (with gains ranging between 16% and 17%). On the other hand, the delay is more costly for the future generations. Aggregating gains and losses using a utilitarian welfare function yields a gain for the high-discount planner of 6.4% which is larger than in the benchmark economy. This large gain is partly due to the fact that the population size is declining, so the planner attaches a higher weight on more numerous earlier generations relative to the baseline economy. The gain is as large as 10.5% if the reform is delayed until 2100. However, the welfare loss for the future generations is also large, equal to ca. 39%. The results are similar, albeit less extreme, for the low-discount

planner. For instance, delaying a reform until 2040 (2100) yields a welfare gain for the low-discount planner of 2.6% (6.5%). In all cases the gains are larger than in the baseline model. The FF reform exhibit larger losses than in the baseline model (even the low-discount planner prefers the benchmark to a fully-funded reform). Moreover, the PAYGO reform yields larger gains than in the benchmark reform (16.5% with the high-discount and 5.3% with the low-discount planner, respectively). Part of the reason is that with a low population growth the planner attaches a higher relative weight to the early generations, who are the winners in this scheme.

In summary, a lower fertility increases the magnitude of the adjustment required to restore the intertemporal balance of the pension system. It also widens the gap between the losses and gains of different generations in the alternative reforms.

5.3 High interest rate

In the macroeconomic literature on pension reforms in developed economies, it is common to assume that the return on the assets owned by the pension fund is equal to the marginal return to capital (cf. Auerbach and Kotlikoff 1984). In this paper, we have calibrated the return on assets to world market interest rate (2.5%). However, the empirical rate of return on capital in China has been argued to be much higher than the world market interest rate (see discussion above). To get a sense of the role of this assumption, we now consider a scenario in which the interest rate is much higher – equal to 6% – between 2012 and 2050. We assume that the period of high interest rate will eventually come to an end as China becomes fully industrialized. According to the macroeconomic model laid out in Section 6 below, the year 2050 is roughly the end of this transition.

There are two main differences. First, delaying the reform yields much smaller gains for the transitional generations, and in fact the low discount planner is essentially indifferent between the benchmark reform and a delay till 2040, which she strictly prefers over delaying until 2100. Second, the FF reform entails larger gains for the future generations and smaller losses for the current generations relative to the baseline calibration. As it should be expected, when the interest rate is significantly higher than the average growth rate the PAYGO system becomes less appealing, because the gains to current generations are smaller. In particular, the low discount planner prefers the FF to the PAYGO reform, although both are dominated by the benchmark reform.

6 A dynamic general equilibrium model

Up to now we have taken the wages and the rate of return on savings as exogenous. As we demonstrated in section 5 the welfare effects depend significantly on the wage growth. 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 3 as an equilibrium outcome. These prices are sufficient to compute the optimal decisions (consumption and labor supply) of workers and retirees as well as the sequence of budget constraints faced by the government. Therefore, the allocations and welfare analysis of the previous section carry over to the general equilibrium environment.

The model is closely related to the model of economic transition of Song *et al.* (2011), augmented with the demographic model and the pension system of Section 3.

6.1 The production sector

The production sector consists of two types of firms: (i) *financially integrated* (F) firms, modelled as standard neoclassical firms; and (ii) *entrepreneurial* (E) firms, owned by (old) entrepreneurs. These are residual claimants on the profits generated by E firms, and delegate their management to specialized agents called *managers*. E firms can run more productive technologies (see Song *et al.* 2011 for microfoundations of this assumption). However, they are subject to credit constraints that limit their size and their growth. In contrast, the less productive F firms are unconstrained. Motivated by the empirical evidence 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:

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

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 capital is *putty-putty*, i.e., there are no irreversibilities in investment decisions. Thus, F-firms can adjust the desired level of capital every period, irrespective of their past productive capacity. Let r^l denote the net interest rate at which F firms can raise external funds. Let w denote the market wage. Profit maximization implies that $K_F = AN_F (\alpha / (r^l + \delta))^{-\frac{1}{1-\alpha}}$, where δ is the depreciation rate. The capital-labor ratio and the equilibrium are determined by r^l . Thus,

$$w_t \geq (1 - \alpha) \left(\frac{\alpha}{r^l + \delta} \right)^{\frac{\alpha}{1-\alpha}} A_t. \quad (6)$$

As long as there are active F firms in equilibrium ($N_F > 0$), equation (6) holds with strict equality.

E firms are subject to a collateral constraint. In particular, a positive share of the capital stock must be financed out of the personal wealth of the entrepreneurs. We denote by $\Omega_{E,t}$ the stock of entrepreneurial wealth at t. Then, the credit constraint imposes that

$$K_{Et} \leq (1 + \sigma) \Omega_{E,t}, \quad (7)$$

where $\sigma/(1 + \sigma)$ is the maximum shares of external financing of E firms capital.

Three regimes are possible: (i) during the first stage of the transition the credit constraint (7) is binding and F firms are active (hence, the wage is pinned down by (6)); (ii) during the mature stage of the transition the credit constraint (7) is binding and F firms are inactive; (iii) eventually, the credit constraint (7) ceases to bind (F firms remain inactive). In regimes (ii) and (iii), (6) holds with strict inequality.

Consider, first, scenario (i), which is the case emphasized in Song *et al.* (2011). Then,

$$K_{Et} = (1 + \sigma) \Omega_{E,t}, \quad (8)$$

implying that K_{Et} is determined by past savings and investment decisions of entrepreneurs, and is a state variable.

In addition to the financial frictions, 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.²⁹ Profit maximization yields, then, the following optimal labor hiring decision:

$$\begin{aligned} N_{Et} &= \arg \max_{\tilde{N}_t} \left\{ (1 - \psi) (K_{Et})^\alpha \left(\chi A_t \tilde{N}_t \right)^{1-\alpha} - w_t \tilde{N}_t \right\} \\ &= ((1 - \psi) \chi)^{\frac{1}{\alpha}} \left(\frac{r^l + \delta}{\alpha} \right)^{\frac{1}{1-\alpha}} \frac{K_{Et}}{\chi A_t}. \end{aligned} \quad (9)$$

Consider, next, the gross rate of return on entrepreneurial wealth $\Omega_{E,t}$. In regime (i), this is given by

$$\begin{aligned} R_{E,t} &= \left((1 - \psi) K_{Et}^\alpha (\chi A_t N_{Et})^{1-\alpha} - w_t N_{Et} - \sigma \left(1 + r_t^l \right) \Omega_{E,t} + (1 - \delta) K_{Et} \right) / \Omega_{E,t} \\ &= \left(r_t^l + \delta \right) \left((1 - \psi)^{\frac{1}{\alpha}} \chi^{\frac{1-\alpha}{\alpha}} (1 + \sigma) - \sigma \right) + 1 - \delta, \end{aligned}$$

where the second expression follows from substituting N_{Et} and w_t by their equilibrium expressions, (6) and (9). We assume that $(1 - \psi)^{\frac{1}{\alpha}} \chi^{\frac{1-\alpha}{\alpha}} > 1$ ensuring that the return to capital is higher in E

²⁹Managers 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.

firms than in F firms ($R_{E,t} > r_t^l + 1$). Note that in regime (i) the rate of return to capital is a linear function of r_t^l in both E and F firms. The equilibrium is closed by the condition that employment in the F sector is determined residually, namely,

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

Consider, next, regime (ii), where only E firms are active ($N_{E,t} = N_t$), and the borrowing constraint is binding, so (8) holds. In this case, the rate 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

$$\begin{aligned} R_{E,t} &= \left((1 - \psi) K_{E,t}^\alpha (\chi A_t N_t)^{1-\alpha} - w_t N_t - \sigma (1 + r_t^l) \Omega_{E,t} + (1 - \delta) K_{E,t} \right) / \Omega_{E,t} \\ &= \left(\alpha (1 - \psi) (1 + \sigma) \left(\frac{\chi A_t N_t}{(1 + \sigma) \Omega_{E,t}} \right)^{1-\alpha} - \sigma (r_t^l + \delta) + (1 - \delta) \right) \end{aligned}$$

In regime (ii), the stock of capital continues to be a state variable determined by the accumulation of entrepreneurial wealth.

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 $w_t = (1 - \alpha) (\alpha / (r_t^l + \delta))^{\alpha/(1-\alpha)} \chi A_t$.

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

6.2 Banks

Competitive financial intermediaries (*banks*) with an 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 operational costs, red tape, etc., 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.

6.3 The households' saving decisions

Workers and retirees face the problem discussed in Section 3, given the equilibrium wage sequence, and having defined $R \equiv 1 + r$. For the sake of realism, we assume that an exogenous share of workers are not in the pension system. These workers pay no taxes and receive no pensions.

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). As they reach age $J_E + 1$, they must retire – i.e., quit as managers – and can become entrepreneur. 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 equals is then given by:

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

6.4 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:

- a high β implies a high propensity to saving of managers and entrepreneurs and a high speed of transition;
- 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 returns of E firms, a high managerial compensation and, hence, a high speed of transition;
- a high productivity differential (χ) implies a high rate of returns of E firms, a high managerial compensation and, hence, a high speed of transition;
- a high σ implies that entrepreneurs can leverage up their wealth and earn a higher return on their savings. This will speed up the transition.

- a high managerial rent (ψ) implies a low rate of returns of 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 only depend on the world market interest rate and on the iceberg cost.

6.5 Calibration

We must calibrate two parameters related to the financial system, ξ and σ , and four technology parameters, $\alpha, \delta, \chi, \psi$. 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%. Like in the partial equilibrium model, we set $\beta = 1.0175$.

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 Song *et al.* (2011) 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.³⁰ The implied parameter values are $\psi = 0.27$ and $\chi = 2.73$. This implies that TFP of an E-firm is 1.65 times larger than TFP of an E-firm.³¹

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 over time improve their lending practices, so that borrowing-lending spreads eventually will 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 to higher wages, respectively. Such development helps the model 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 DPE employment share is in line with the private employment

³⁰Song et al. (2011) document that in manufacturing, DPEs 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).

³¹Hsieh and Klenow (2009) estimate the TFP across manufacturing firms in China and find that the TFP of DPEs is about 1.65 times larger than the TFP of SOEs.

growth between 2000 and 2008 in urban areas. We set the initial level of productivity, A_{2000} , so that the urban GDP per capita is 20% of the US level in 2011. Moreover, we set the growth in A_t , i.e., the secular exogenous productivity growth, so that the model generates an aggregate growth in GDP per capita of 9.7% for China during 2000-2011. The resulting growth rate in A_t is 2% larger than the associated world growth rate during this period. After 2011, this excess growth in A_t falls linearly to zero until the TFP level in E-firms is equal to that of US firms. This occurs in year 2022.

The initial conditions are set as follows. The 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 in 1992 all entrepreneurs are endowed with the same initial wealth (1992 is the year when free-market reforms in China accelerated). Moreover, all managers are assumed to start with zero wealth in 1992. Initial wealth for workers and retirees is also set to zero in 1992. The 2000 distribution of wealth across individuals is then derived endogenously. Finally, the initial government wealth is set to 71% of GDP in 2000 so as to generate a net foreign surplus equal to 12% of GDP in 2000.

6.6 Simulated output trajectories

The calibrated model yields growth forecast that we view as plausible. Figure 12 shows the evolution of productivity and output per capita forecasted by our model. The growth rate of GDP per worker remains about 8.5% per year until 2020 (see upper panel). After 2020, productivity growth is forecasted to slow down. This is due to 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 6.9% between 2020-30, and eventually dies off in the following decade. Note that the growth of GDPpc is lower than that of GDPpw after 2015, due to the increase in the dependency ratio. On average, China is expected to grow at a 6.5% rate between 2012 and 2040. The contribution of human capital is 0.8% per year, due to the entry in the labor force of more educate young cohorts. In this scenario, the GDP per worker of China will be 73% of the US by 2039, remaining broadly stable thereafter. The total GDP of China is set to surpass that of 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 3 is now an endogenous outcome. Wages are forecasted to grow at an average 5.1% until 2030, and to slow down thereafter. What keeps wage growth high after 2020 is mostly capital deepening.

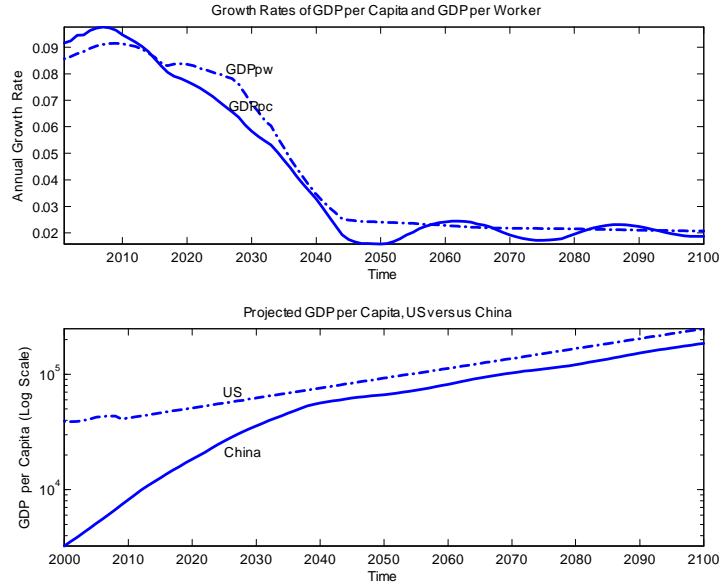


Figure 12: Upper panel shows projected annual growth rates in GDP per worker and GDP per capita in the calibrated economy. Lower panel shows projected GDP per capita in levels for China versus the US.

6.7 Sensitivity analysis

6.7.1 High savings and foreign surplus

Although the growth forecasts are plausible, the calibrated economy generates a very large amount of savings. For instance, by 2070 the economy has a wealth-GDP ratio equal to 1169%. The reason for this is that the model is calibrated to match the aggregate savings during 2000-2010. In that period, China experienced high growth, and yet a very high saving rate (48.2% on average).

Since our stylized model forecasts and 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 ($\beta = 1.0175$) in order to predict the empirical saving rate during the first decade of the XXth century. According to our model, the future savings rate will be even higher than today once the wage growth declines – provided that the discount factor remains constant. In our model, a high β is a stand-in for a number of institutional features that are not explicitly considered and that may explain a high propensity to save over and beyond pure preferences. For instance, the high savings could be due to a large precautionary motive or large downpayment requirements for house purchases.³²

³²[preliminary] Chamon et al. (2010) and Song and Yang (2010) study household savings in calibrated life-cycle models incorporating individual risk and detailed institutional features of the welfare system. Both studies find that with a conventional choice of β , their models would imply too low savings, especially for the young.

It is important to note that the 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). The entrepreneurial firms grow out of their financial constraint by year 2039. Thereafter, domestic capital accumulation and wages are determined by the world interest rate. Thus, β only determines the foreign position, which is predicted to reach 13.7 times GDP by 2070.

It seems implausible that China will accumulate such a large foreign surplus. One might be also concerned that the high discount factor could affect our quantitative welfare results. To address such concerns, we consider an alternative scenario where all cohorts entering the labor market after 2012 have $\beta = 0.97$. In such an alternative scenario China's net foreign position would be zero in the long run. The results are shown in the Appendix. The analysis of the alternative pension arrangements 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.

6.7.2 Financial development

The model borrows from Song *et al.* (2011) 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 2038, 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 – due, e.g., to financial development – would speed up the 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 2012. This means that state owned firms vanish, and there is large capital inflow driven by entrepreneurial borrowing. Wages jump upon impact (by 85%) due to the large capital deepening. In 2030, the wage level is still 15.8% above the baseline calibration. By 2038 the wage level is the same as in the benchmark calibration.

While financial development affects the transition path, it brings little change to the conclusions of the welfare analysis. The benchmark reform requires a slightly smaller reduction of the replacement rate: 40.7% instead of 40%. The delayed reform still entails gains for the transition cohorts, albeit these decline faster over time. For instance, delaying a reform till 2040 yields a 17% consumption equivalent gain for the cohort retiring in 2012, but only a 12% gain for the cohort retiring in 2039. The loss suffered by the cohorts retiring after 2040 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, 4.1% and 0.5% (5% and 0.8% in the baseline scenario).

The FF reforms yields slightly better outcomes. All generations retiring after 2050 gain from reform (2058 in the baseline scenario), and the loss of the earlier cohorts only reach 8% (11% in the baseline scenario). The high-discount planner continues to prefer the benchmark reform to the FF reform, while the low-discount reform 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 planner continue to prefer the PAYGO reform to any alternative reform considered. However, the welfare gap between the PAYGO and the fully-funded reform is now smaller, since the planner dislikes the concentrated nature of the gains under the PAYGO. For instance, the consumption-equivalent gain of the low-discount planner relative to the benchmark reform is 1.1%, compared with 1.8% in the baseline scenario. Since the fully-funded reform also entails a 0.6% 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, 11.6%, for the high-discount planner).

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

7 Conclusions

China faces a major dilemma concerning the inclusive nature of its institutions and the extent to which all its citizens and generations get to share the benefits of high growth. In this paper, we have studied the welfare effects of alternative pension reforms with the aid of a dynamic general equilibrium model. Our model – based on Song *et al.* (2011) – is quantitatively consistent with the aggregate trends of the Chinese economy in the first decade of the XXIst Century. In addition, it delivers broadly plausible forecasts: wages will remain high (and possibly increase) until ca. 2030; growth will eventually slow down, and China will become a mature economy by ca. 2040.

A number of studies, mostly based on aggregate demographic models, have argued that China must reform its pension system to achieve long run balance in response to a sharp increase in the dependency ratio (see, e.g., Sin (2005), Dunaway and Vivek (2007), Salditt *et al.* (2007), and Lu (2011)). Our analysis concurs with this view, but shows that rushing into a draconian reform would have large adverse effects on inequality: it would harm significantly current generations and benefit only mildly future generations. In a fast-growing society like China, this would imply dispensing with a powerful institution that redistributes resources from richer future generations to poorer current generations. Under standard welfare criteria, a straight pay-as-you-go system would be preferred to both the draconian reform and to a reform aiming at pre-funding a pension system.

Our model would yield very different predictions in a mature economy with low wage growth and

perfect capital markets, where a fully funded system may in fact be preferred to a pay-as-you-go system. The result highlights the general principle (see, e.g. Acemoglu *et al.* 2006) that mechanically transposing policy advises from mature to developing or emerging economies may be misleading.

REFERENCES

- Abel, Andrew, N. Gregory Mankiw, Lawrence H. Summers, and Richard J. Zeckhauser, 1989, "Assessing Dynamic Efficiency: Theory and Evidence", *Review of Economic Studies*, 56 (1), 1-19.
- Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti, 2006. "Distance to Frontier, Selection and Economic Growth." *Journal of the European Economic Association*, 4 (1), 37-74.
- Auerbach, Alan and Laurence Kotlikoff, 1987. *Dynamic Fiscal Policy*. Cambridge University Press, Cambridge and New York.
- Bai, Chong-En, Chang-Tai Hsieh, and Yingyi Qian, 2006, "The Return to Capital in China." *Brookings Papers on Economic Activity*, 2: 61–88.
- Chan, Kam Wing and Will Buckingham, 2008, "Is China Abolishing the Hukou System", *The China Quarterly*, 195, September 2008, 582-606.
- Chamon, Marcos, Kai Liu, and Eswar S. Prasad, 2010. "Income Uncertainty and Household Savings in China," NBER Working Paper 16565.
- Dollar, David, and Shang-Jin Wei. 2007. "Das (Wasted) Kapital: Firm Ownership and Investment Efficiency in China." National Bureau of Economic Research Working Paper 13103.
- Dunaway Steven and Vivek Arora, 2007. "Pension Reform in China: The Need for a New Approach," IMF Working Paper WP/07/109.
- Duan, Chengrong and Yujing Sun, 2006, "Changes in the Scope and Definition of the Floating Population in China's Censuses and Surveys." (in Chinese), *Population Research*, 30(4), 70-76.
- Feldstein, Martin. 1999. "Social security pension reform in China." *China Economic Review*, 10 (2), 99-107.
- Goodkind, Daniel M., 2004, "China's Missing Children: The 2000 Census Underreporting Surprise." *Population Studies*, 58(3), 281-295.
- Gustafsson, Björn , Li Shi, and Wei Zhong, 2006, "The Distribution of Wealth in Urban China and in China as a Whole in 1995." *Review of Income and Wealth*, 52 (25), 173-188.
- Hu, Ying, 2003, "Quantitative Analysis of the Population Transfers from Rural Area to Urban Area." (in Chinese) *Statistical Research*, 7, 20-24.
- Hurd, Michael D., 1989, "Mortality Risk and Bequests." *Econometrica*, 57, 779-813.
- Johnson, Gale D., 2002, "Provincial Migration in China in the 1990s." *China Economic Review*, 14, 22-31.
- Keane, Michael. P., 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature*, 49(4), 961-1075.
- Lavelly, William, 2001, "First Impressions from the 2000 Census of China." *Population and Development Review*, 27(4), 755-769.
- Li, Si-ming, 2004, "Population Migration and Urbanization in China: A Comparative Analysis of the 1990 Population Census and the 1995 National One Percent Sample Population Survey." *International Migration Review*, 38 (2), 655-685.

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

Lu Jiehua, 2011, "Impacts of Demographic Transition on Future Economic Growth: China's Case Study," paper presented at the Conference: "China and the West 1950-2050: Economic Growth, Demographic Transition and Pensions," University of Zurich, 21 Novembre 2011.

Morgan, Philip S., Zhigang Guo and Sarah R. Hayford, 2009, "China's Below-Replacement Fertility: Recent Trends and Future Prospects." *Population and Development Review*, 35(3), 605-629.

William Nordhaus, 2007, "Critical Assumptions in the Stern Review on Climate Change," *Science*, 317, 201-202.

Retherford, Robert D., Minjia Kim Choe, Jiajian Chen, Xiru Li and Hongyan Cui, 2005, "How Far Has Fertility in China Really Declined?" *Population and Development Review*, 31 (1), 57-84.

Salditt Felix, Peter Whiteford and Willem Adema, 2007. "Pension Reform in China: Progress and Prospects, OECD Social, Employment and Migration Working Paper 53"

Sin, Yvonne, 2005, "China: Pension Liabilities and Reform Options for Old Age Insurance," World Bank Working Paper No. 2005-1.

Song, Zheng, Kjetil Soresletten and Fabrizio Zilibotti, 2011, "Growing Like China." *American Economic Review*, 101 (1), 196-233.

Song, Zheng and Dennis T. Yang, 2010, "Life Cycle Earnings and the Household Saving Puzzle in a Fast-Growing Economy," mimeo, Chinese University of Hong Kong.

Zeng, Yi, 2007. "Options of Fertility Policy Transition in China." *Population and Development Review*. 33 (2), 215-246.

Zhang, Guangyu and Zhongwei Zhao, 2006, "Reexamining China's Fertility Puzzle: Data Collection and Quality over the Last Two Decades." *Population and Development Review*, 32 (2), 293-321.

Zhang, Junsen, Yaohui Zhao, Albert Park and Xiaoqing Song, 2005, "Economic Returns to Schooling in Urban China, 1988 to 2001." *Journal of Comparative Economics*, 33 (4), 730-752.

Zhang, Weimin and Hongyan Cui, 2003, "Estimation of Accuracy of 2000 National Population Census Data." (in Chinese), *Population Research*, 2003 (4), 25-35.

Zhou, Yixing and Laurence J. C. Ma, 2003, "China's Urbanization Levels: Reconstructing a Baseline from the Fifth Population Census." *China Quarterly*, 176-196.