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

March 2012
preliminary and incomplete version

Abstract

A major dilemma facing China is the extent to which the benefits of high growth can be shared more equally within and across cohorts. China will experience a sharp increase in the old-age dependency ratio which may undermine the sustainability of its pension system, arguably the most important institutional vehicle of inter-generational redistribution. In this paper, we analyze the welfare effects of alternative reforms of the Chinese pension system, with the aid of a dynamic general equilibrium model which incorporates the salient structural factors, namely, the population dynamics (including internal migration) and productivity growth. We show that the current system is unbalanced and that either higher taxes or lower benefits are necessary to restore its intertemporal balance. However, delaying such a reform by a few decades has a major positive impact on the welfare of (poorer) workers retiring in the next few years, and imposes only a small additional cost on (richer) future generations. In contrast, a fully funded reform has negative welfare implications on the earlier generations, and yields small gains to the generations retiring in the far future. The high wage growth predicted by our model is key for these normative results.

JEL Codes. F43, G21, O16, O47, O53, P23, P31.

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 Tim Besley, Martin Eichenbaum, Dirk Krueger, Torsten Persson, Richard Rogerson, and seminar participants at the Conference "China and the West 1950-2050: Economic Growth, Demographic Transition and Pensions" (Univeristy of Zurich November 21, 2011), London School of Economics, Princeton University, Tsinghua Workshop in Macroeconomics 2011, University of Frankfurt, University of Mannheim, University of Pennsylvania, University of Zurich. 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 for 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 large and growing, partly due to the fast productivity growth that makes the labor market perspective of a worker entering the labor force today substantially better than that of a worker who entered in, e.g., 1970. These observations motivate a debate about what institutional arrangements can allow more people to share the benefits of high growth.

An important feature of China's development is the 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, followed by 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 and faces a long period of ageing population: the old-age dependency ratio will increase from the current 12% to 39% in 2040. This evolution threatens the viability of redistributional policies. The focus of this paper is on pensions, arguably the most important institutional vehicle to achieve intergenerational redistribution. In particular, we analyze the welfare effects of alternative reforms of the Chinese pension system in the face of the demographic transition.

The analysis is based on a dynamic general equilibrium model incorporating the salient structural factors affecting the welfare cost and benefits of a public pension system. The standard tool for such analysis is the Auerbach and Kotlikoff (1987) model (henceforth the *Au-Ko model*). This is a multiperiod neoclassical overlapping generations (OLG) model with endogenous capital accumulation and wage growth, and a pension system. While sharing the structure of the Au-Ko model, our model embeds a number of structural features specific to the Chinese economy and central for our results. These include the rural-urban transition and a rapid transformation of the urban sector where state-owned enterprises are taken over by entrepreneurial firms. Such a transition is characterized, as in Song et al. (2011), by important financial and contractual imperfections.

The model bears two key predictions for the questions at hand. First, it predicts *delayed* wage growth: As long as transitions within the urban sector persists, wage growth is moderate. Yet, as this transition comes to an end, the model predicts an acceleration in wage growth, which we expect will occur over the two coming decades. Second, financial imperfections result in a large gap between the rate of return to capital and the rate of return to which Chinese savers (except of entrepreneurs) have access to. A calibrated version of the model provides forecasts of China's future economic performance: Wages will grow at an average 6.2% until 2030, and slow down thereafter; GDP growth will slow down gradually, remaining high over the next twenty years (an average 6.1% per year). By 2050, China will converge to 73% of the level of GDP per capita of the US.

We use the model to address two questions: (i) is a pension system based on the current rules sustainable? (ii) what are the intra- and intergenerational welfare effects of alternative reforms? The answer to the first question is clear-cut: the current system is unbalanced and requires a significant adjustment, either in the form of a reduction of pension benefits or in the form of an increase in the social security tax. We focus on adjustments on the benefit margin, and consider as a benchmark a reform cutting the benefit of all workers retiring after 2011. Reflecting a common feature of most actual reforms in OECD, 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. Such a reform entails a sharp reduction of the replacement rate, from 60% to 40%.

To address the second question, we consider two alternative scenarios. First, we consider a *delayed reform*, such that the current rules are retained by a certain number of years, and then followed by a permanent reduction in the replacement ratio so as to balance the pension system in the long run. The results are interesting: If the reform is delayed until 2040, the transition generations experience enormous welfare gains, while the extra cost borne by the generations retiring after 2040 (who perceive a lower replacement ratio) is very small. If a reform is delayed until 2100, the transition generations also gain a lot, but the extra cost they impose on future generations is sizeable. Second, we consider switching to a pure individual account savings-based system, which in our model 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 born in a sufficiently far future benefit from a fully-funded reform, because the current pension system is a net burden for them. In contrast, earlier generations are harmed by the reform. Interestingly, all cohorts retiring before 2062 lose in net terms, and the losses are especially large for cohorts retiring before 2040. Third, we consider switching to a pure pay-as-you-go system where the replacement rate is endogenously determined by the dependency ratio subject to a balanced budget condition for the pension system. Given the demographic transition, this system yields the most generous pensions to early cohorts and treats generations retiring after 2050 the worst.

If one aggregates the welfare of different cohorts using a utilitarian social planner, the fully funded reform turns out to be unattractive, while the utilitarian planner would typically like to delay the reform in order to favor the *poor* cohorts retiring in the next few decades over the *rich* cohorts retiring in the distant future. In other words, running a temporarily unbalanced transfer-based pension system by, for example, delaying a reform until 2040, is an effective way to achieve intergenerational redistribution without imposing large costs on future generations. For the same reason, switching to a pay-as-you-go yields large welfare gains to the planner.

Our results may sound surprising to readers familiar with the debate on pension reforms in developed economies. They 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. For instance, if we lower the wage growth from the average 6.3% annual rate predicted by our model over the 2010-30 period to an average 2% (a conventional wage growth for mature economies), the welfare gains are quantitatively much smaller and in some cases reversed. The pay-as-you go is in some cases dominated by a fully funded system. 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 pay-as-you go system over the transition without imposing a large burden to future generations. This is attractive from the point of view of intergenerational fairness, insofar as it allows the old who earned relatively low wages when the country was less developed to share the fruits of economic development with the richer young generations. For the same reason, a fully funded reform would be a small gain for future generation at a large welfare cost for the current old workers.

Finally, the model can be used to address additional policy questions pertaining to taxation and redistribution. The current pension system of China only covers a subset of urban workers. The government is contemplating introducing a limited rural pension. We show that extending the coverage of the current urban pension system to rural workers would be relatively inexpensive, even though benefits were paid to workers who never contribute to the system. Moreover, this would trigger large welfare gains for the poorest part of the Chinese population.

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 5 provides a full general 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 6 concludes.

2 Demographic Model

Throughout the 1950s and 1960s, China had an average total fertility rate (TFR, henceforth) 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 people, 70% higher than in the 1953 census. The view that a booming population was a burden for the development process of China induced the government to introduce a set of measures to curbe fertility during the 1970's, culminating in the one-child policy of 1978. This policy imposes severe sanctions on couples resident in urban areas who have more than one child. The policy underwent a few reforms, and is 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 (that varies across provinces) elapses between the first and second birth. Today's TFR is below replacement level, although there is no uniform consensus as to the 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). Current estimates suggest a TFR of about 1.6 (see Zeng Yi 2007). The demographic outlook is the source of growing concern. The Chinese government has already loosened the policy in some urban areas, and further adjustments are likely to be introduced in future.[1]

2.1 Natural Population Projections

We consider, initially, a model without rural-urban migration, which is referred to as the *natural* population dynamics. Rural-urban migration is introduced in the next section. 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.[2] 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.[3] To heed this concern, we take the rural-urban population and 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 with development, as it did in the past. 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 periods 2000-2020, 2020-2050 and 2050-2080. 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 that the age-specific fertility rates 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.[4] 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.[5] Zeng (2007) estimates that such a policy would increase

[1] No drastic change is in the horizon, though. In 2008, China's National Population and Family Planning Commission stated that the one-child policy policy would not be lifted for at least ten more years.

[2] The 2000 census data is regarded as a reliable source (see, e.g., Lavely, 2001; Goodkind, 2004). The total population was originally estimated to be 1.24 billion, but the figure was revised later 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, reducing the urban share from 36.9% to to 36%.

[3] Using intercensal methods, Goodkind (2004) estimated a total of 37 million children under age 9 missing in the 2000 census. A similar estimate of 30 million is obtained by Zhang and Cui (2003) who use primary school enrolments to back out the actual child population.

[4] China has currently an acute gender imbalance which is taken into account in our model. It is unlikely that the gender balance will persists at such a high level, though. 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 main results are robust to plausible changes in the gender imbalance.

[5] In July 2011, Zhang Feng, director of the Guangdong provincial population and family planning commission issued

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 some alternative scenarios.

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).[6] Second, all individuals who did not move but whose place of registered residence switched from being classified as rural into being classified as urban. This is a sizeable group: According to China Civil Affair's Statistical Yearbooks, a total of 8439 new towns were established from 1990 to 2000 and 44 million rural citizens became urban citizens (Hu, 2003). 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.[7] 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.[8] The

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.

[6] There 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.

[7] Although 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).

[8] Our method is related to Johnson (2003), who also exploits *natural* population growth rates. Our work is different

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.[9] Survey data show that the urban population grows at 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 2.2. 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.

---

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.

[9] There 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 (*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 downards.

The figure shows rural-urban migration rates by age and gender, estimated as explained in the text. The estimates are smoothed by 5-years moving averages.

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 1 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: XXX million rural residents will move to urban areas between 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

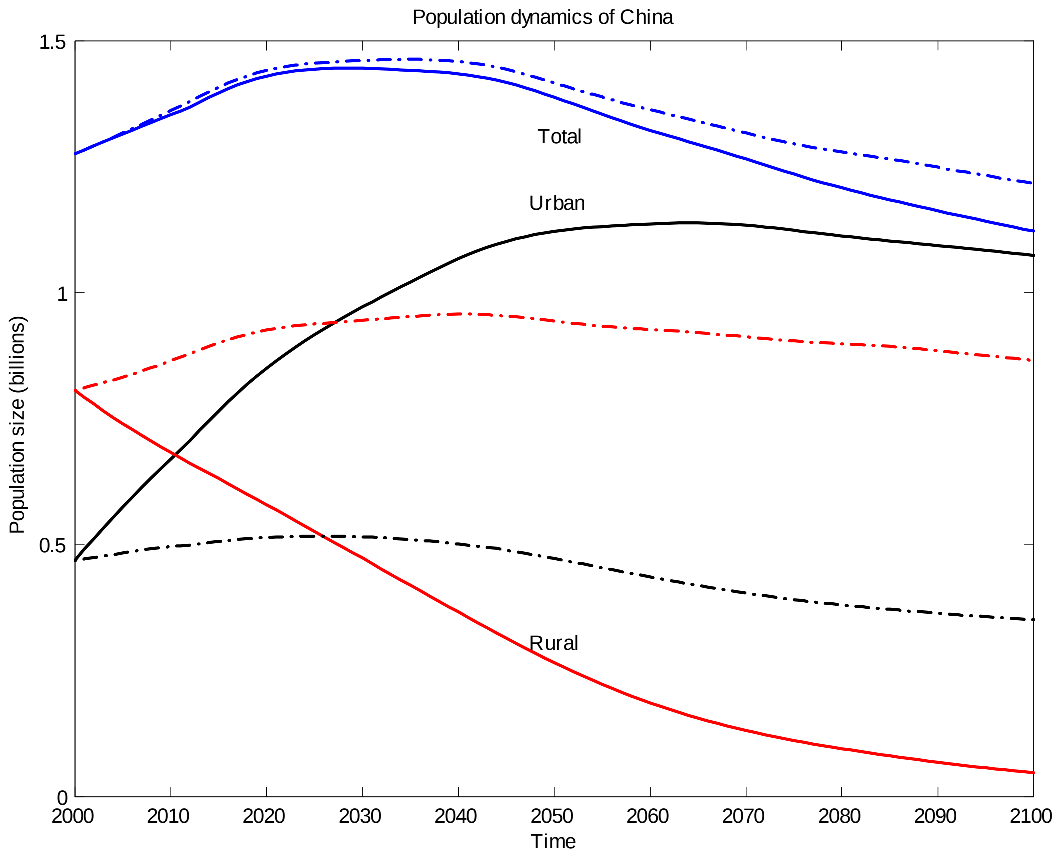

Figure 1: The figure shows the projected population dynamics for 2000-2100 (solid lines) broken down by rural and urban population. The dashed lines show the corresponding natural population dynamics, i.e., the counterfactual projection where there is no urban-rural migration.

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.

A key feature of this demographic transition is an ageing population. Figure 2 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).[10] We also plot, for contrast, the old-age dependency ratio in the no migration counterfactual (dotted 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.

[10] In 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.

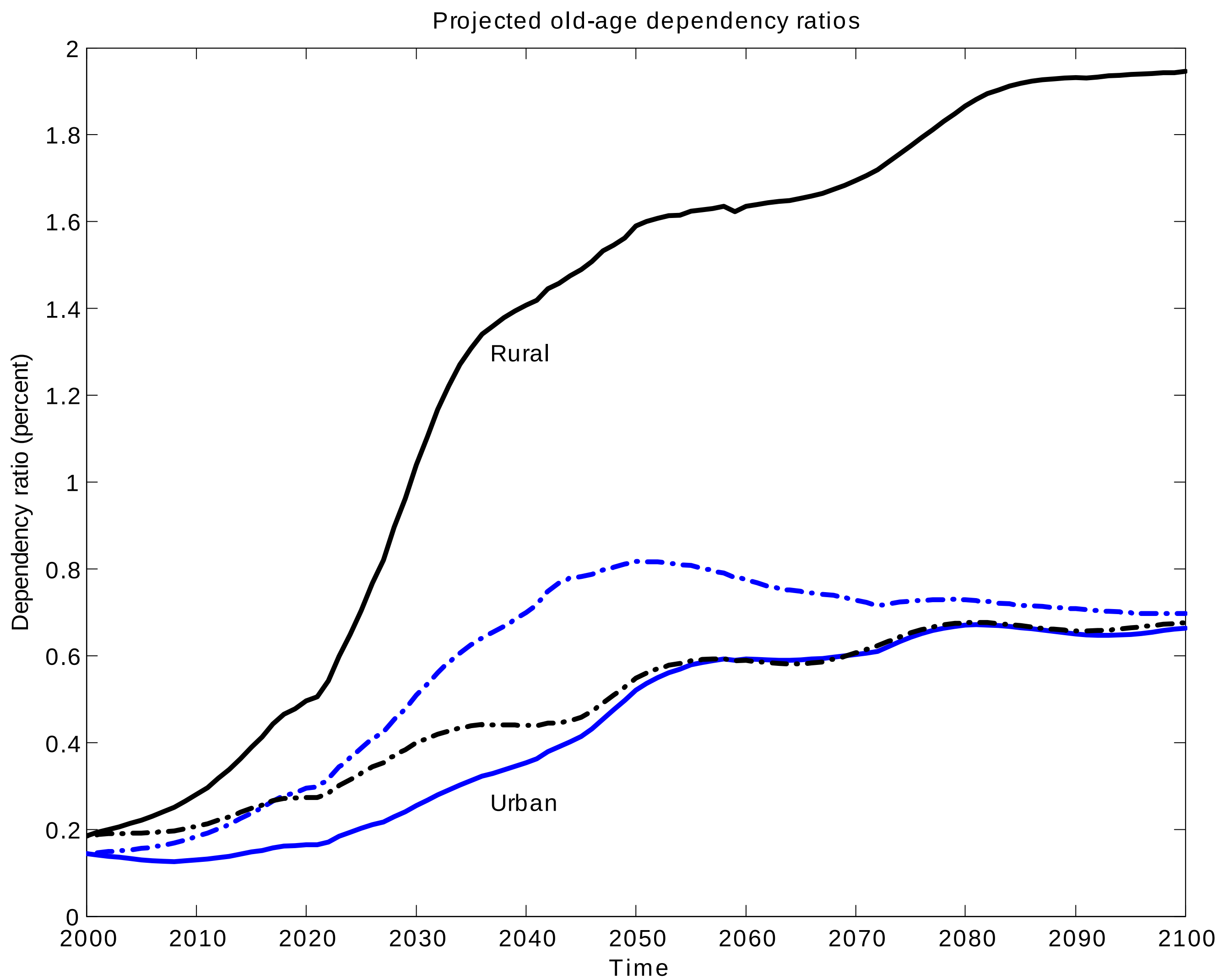

Figure 2: The Figure shows the projected old-age dependency ratios for 2000-2100 (solid lines) for rural (black lines) and urban (blue lines) population. The dashed lines show the corresponding old-age dependency ratios conditional on the natural population dynamics, i.e., in the no migration counterfactual.

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 5 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,

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

where c is consumption and $1-h$ is leisure. 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:

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

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:

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

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 pay-as-you-go 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 introduce equilibrating measures, such as a reduction in the replacement rate and a stricter enforcement of social security contributions. The 1997 system has two tiers. The first tier is a standard transfer-based system with resource-pooling at the provincial level. The second tier 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." De facto, given the low capitalization of the system, it can be regarded as a transfer-based system which permits, as do e.g. 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 pillars of the system.

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

$$b_{t,t+j} = q_{t+J_W} \cdot \left(0.6 \cdot \bar{y}_{t+J_W} + 0.4 \cdot \bar{y}_{t+j}\right), \tag{2}$$

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.[11] The tax and the benefit rule do not guarantee that the system is financially viable. In fact, we will

[11] The 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 partcipated in the system. For instance, in the annual National

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.[12] In the baseline model, we assume a constant coverage rate of 60%. We also consider an alternative scenario in which the coverage rate grows over time. One issue concerns the coverage rate of migrant workers. Since we do not have precise infomation in this regard, 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;[13] 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 up to period T>2012, in the sense that the current replacement rate ($q_t = 60\%$) apply for those who retire before period T. After T, 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 fully funded 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 *fully funded reform* scenario, the pension system is abolished in 2012. However, the government does not default its outstanding liabilities: those who are already retired recieve 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

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%. [XXX ADD MORE INSTITUTIONAL DETAILS]

[12] The coverage rate is equal to the number of employees participated in the pension system divided by the number of urban employees. Both numbers are obtained from China Statistiscal 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.

[13] In 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.

security tax. 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 – similarly to an ongoing scheme started recently by the government on a limited scale.[14] The *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 for Chinese urban workers by Song and Yang (2011). 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 future dynamics of wages per effective unit of labor is shown in Figure 3. In this projection, wages per effective unit of labor grow at approximately 6.3% between 2000 and 2011, 5.9% between 2011 and 2030, and 2.7% between 2030 and 2050. In the long run, wage is assumed to grow at 2% per year, in line with wage growth in the United States over the last century. In section 5, we show that the assumed wage rate dynamics of Figure 3 is the equilibrium outcome of a calibrated version of the model of Song, Storesletten and Zilibotti (2011).

There has been substantial human capital accumulation in China over the last 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 14 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. Since younger cohorts have more years of education, wage growth across cohorts will exceed that shown in Figure 3. However, the education level for an individual remains constant over his/her worklife, so Figure 3 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 unaccessible 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

[14] The 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 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%.

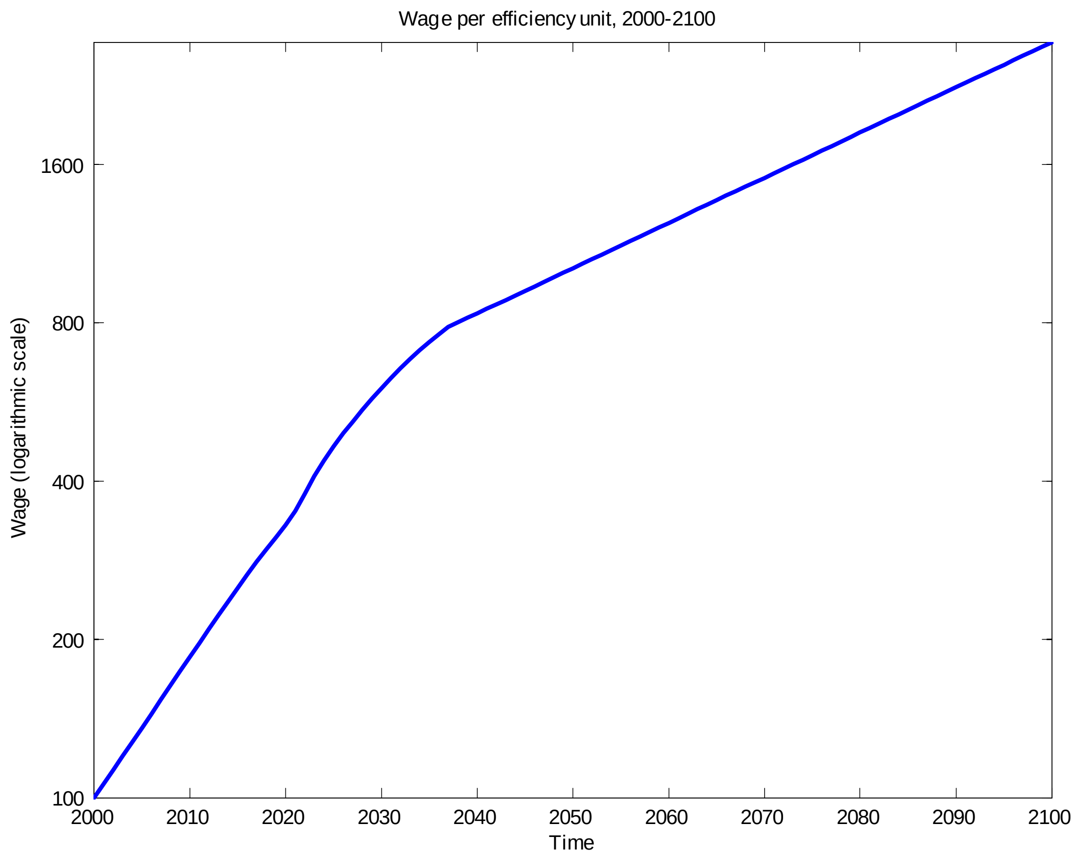

Figure 3: The figure shows the projected real wage per efficiency unit in urban areas, normalized to 100 in 2000. The process is the endogenous outcome of the general equilibrium model of section 5.

saving for Chinese households. For example, in 2002 more than 68% of households' financial assets 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.[15]

Building on Song et al. (2011), the model of Section 5 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.[16]

Consider, finally, preference parameters: The discount factor is set to $\beta = 1.012$ to capture the large private savings in China. This is close, if slightly higher, to 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 4.3). In Section 5 we document that with $\beta = 1.012$ the model economy matches China's average aggregate saving rate during 2000-2010. The alternative low β is chosen so as to imply a zero foreign surplus in the long run for China.

We assume logarithmic preferences over consumption. In the baseline calibration, we assume an inelastic labor supply with no consumption-leisure trade off. However, both the social security tax and the provision of pension in old age can distort labor supply. To account for the possibility that such distortion could affect our results, we consider in the robustness section a version of the model with Cobb-Douglas preferences over consumption and leisure: $u(c,h) = \log c + \Omega \log(1-h)$. The implied intertemporal elasticity of consumption and aggregate labor supply elasticity are consistent with standard assumptions in the business-cycle literature (cf. Cooley, 1995). We set the parameter $\Omega = 1.67$ so that agents work 9/24 of their time endowment (according to CHIP 2002, urban workers work on average 45 hours per week, or 9 hours per day with a five-day week). As we will see below, the labor supply margin turns out to be quantitatively unimportant for our main question.

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

[15] The balance sheet of the Chinese government consists mainly of three items: foreign govenment 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.

[16] Assuming 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).

wage growth of 6.2%, excluding human capital growth. The distribution of wealth in 2000 is then obtained endogenously. The initial government wealth in 2000 is set to 60% 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 5.

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.[17]

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 30.2% as of year 2012.

Figure 4 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 2052. The government builds up a government trust fund amounting to 256% 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 2052.[18]

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

[18] Note that in panel c the government net wealth (i.e., minus the debt) is falling sharply between 200 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.

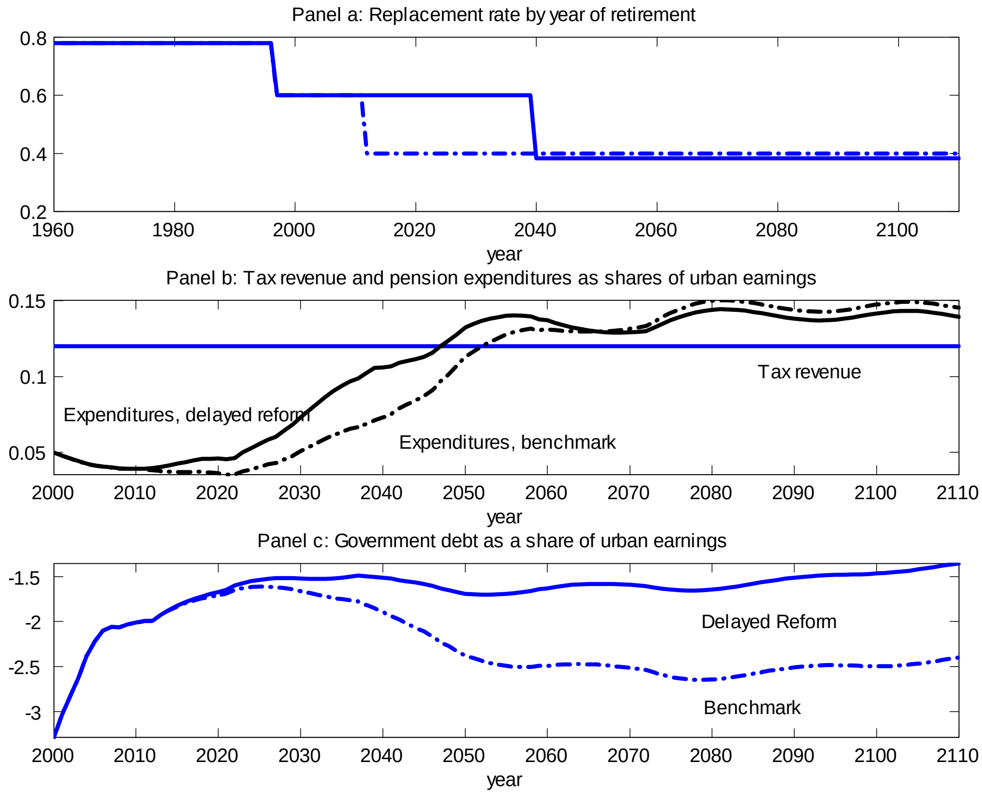

Figure 4: Panel (a) shows the replacement rate q_t for the benchmark reform (dashed line) and for the case when the reform is delayed until 2040. Panel (b) shows tax revenue (blue) and expenditures (green), expressed as a share of aggregate urban labor income (benchmark reform is dashed and the delay-until-2040 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.

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 (2006)), or whether it could be deferred. In addition, we consider two more radical alternative reforms: a move to a fully funded, pure contribution-based system, and a move in the opposite direction to a pure pay-as-you-go 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.[19] 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:

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

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

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

where superscripts $BENCH$ stand for the allocation in the benchmark reform and stars stand for the allocation in the alternative reform.[20]

The planner experiences a welfare gain (loss) from the alternative allocation whenever $\omega > 1$ ($\omega < 1$). 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, $\phi = R/\left(1+g\right)$, 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.*

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, at which a reform similar to the benchmark

[19] Note 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.

[20] Note 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.

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 4 describes the positive effects of delaying the reform until 2040. Panel (a) shows that the post-reform replacement rate falls now to 38.3%, which is only 1.7 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 2065. Moreover, the system starts running a deficit already in 2047. As a result, the government accumulates a smaller trust fund during the years in which the dependancy 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 areexogenous in this section, will be derived as the endogenous outcome of a calibrated general equilibrium model with credit market imperfections in section 5.

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 2090. Consequently, a sizeable adjustment is required in 2100: the replacement rate must fall to 29.4% 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 decline 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 14.6% gains of the generation retiring in 2039. On the contrary, all generations retiring after 2039 lose, though their welfare losses are quantitatively small, being less than 0.8% 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 2060. The losses accruing to the future generations are now significantly larger. All agents retiring after 2100 suffer a loss equivalent to 5% of their lifetime consumption.

Figure 8 shows the welfare gains/lossed of delaying the reform until year T, according to the

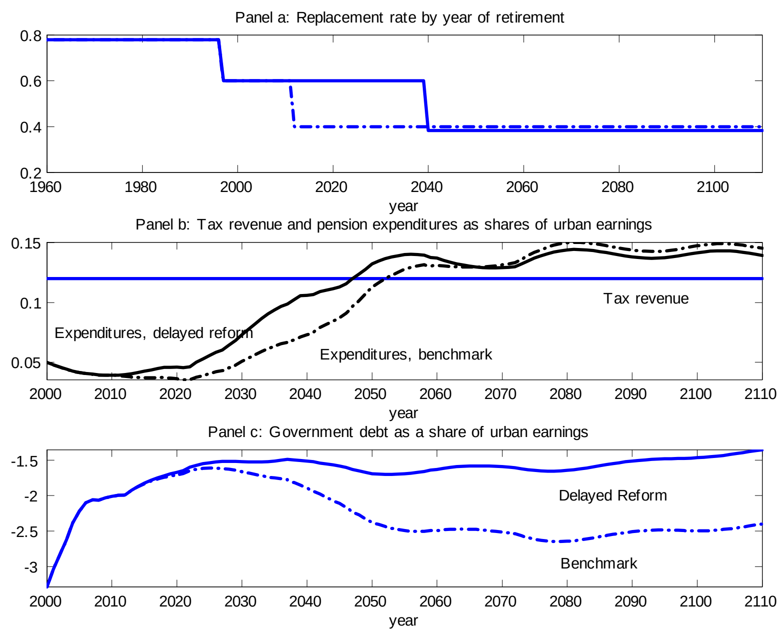

Figure 5: Panel (a) shows the replacement rate q_t for the benchmark reform (dashed line) and for the case when the reform is delayed until 2040. Panel (b) shows tax revenue (blue) and expenditures (green), expressed as a share of aggregate urban labor income (benchmark reform is dashed and the delay-until-2040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

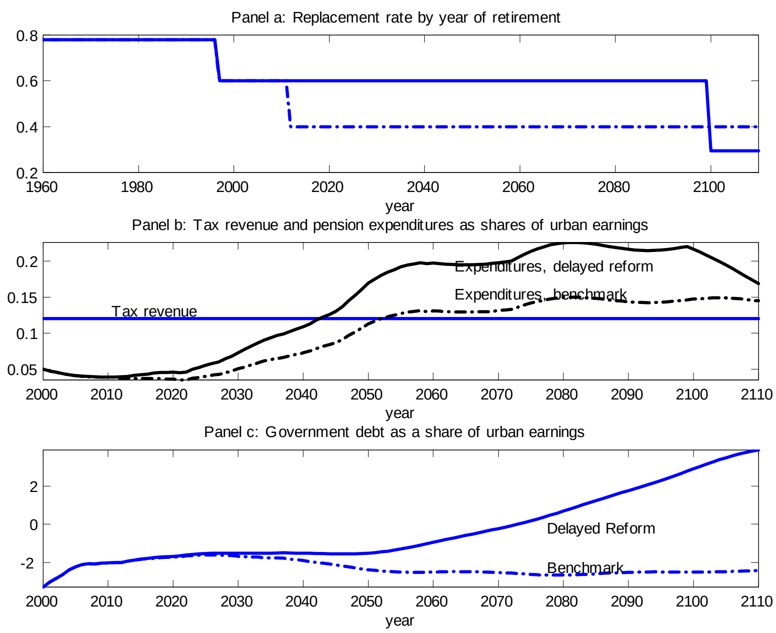

Figure 6: Panel (a) shows the replacement rate q_t for the case when the reform is delayed until 2040 (solid line) and the benchmark reform (dashed line). Panel (b) shows tax revenue (blue) and expenditures (green), expressed as a share of aggregate urban labor income (benchmark reform is dashed and 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.

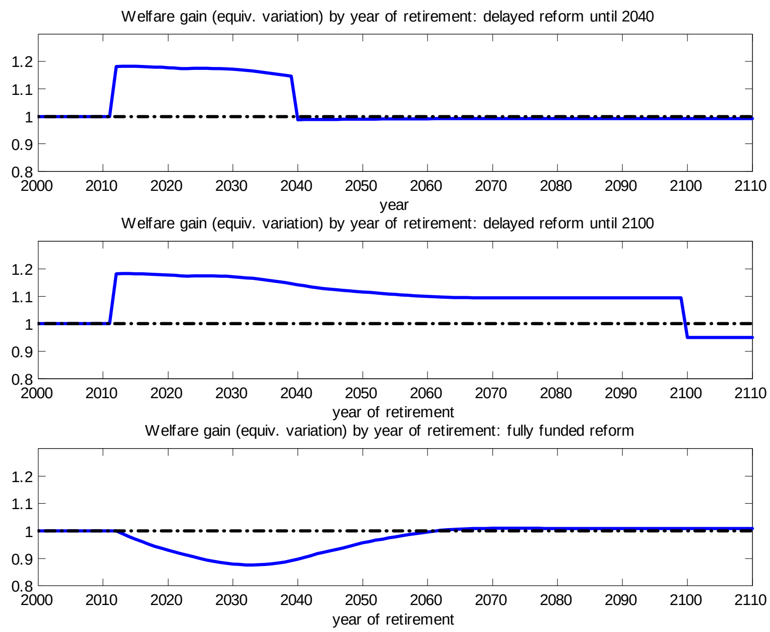

Figure 7: The figure shows the welfare gain for each cohort, indexed by their year of retirement, of implementing an alternative reform instead of the benchmark reform. Panel (a) is the welfare gain of delaying the reform until 2040, Panel (b) the gain of delaying the reform until 2100, and Panel (c) the gain of moving to a fully funded system in 2012. The welfare gain is the equivalent variation gain, i.e., what share of remaining lifetime consumption an individual is willing to forego to get the alternative reform instead of the benchmark reform.

utiltarian social welfare function. The figure displays two curves: In the upper curve, we have the 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 $\omega = 1.052$ high-discount planner, i.e., the delayed reform is equivalent to a permanent 5.2% 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 6.3 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, $\omega = 1.008$.

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 2459. 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, as the next alternative reform, switching to a fully funded 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 fully funded system is equivalent to no pension system. However, switching to a fully funded 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, 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.[21] 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 200% of total labor earnings in 2011. A permanent 0.6% 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

[21] In 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.

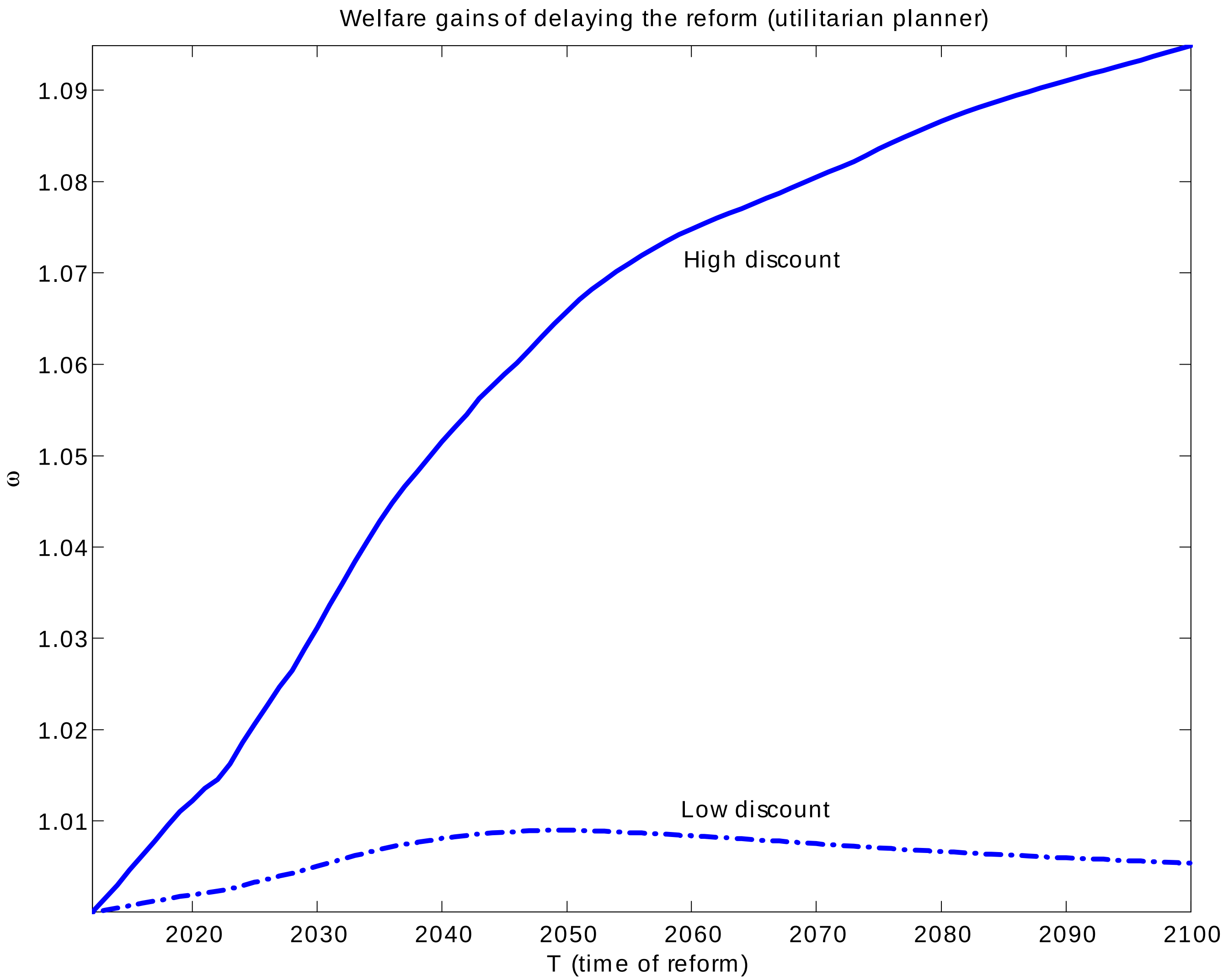

Figure 8: The figure shows the consumption equivalent gain/loss accruing to a high-discount planner (solid lines) and to a low-discount planner (dashed lines) of delaying the reform until time T relative to the benchmark reform. When $\omega = 1$, the planner is indifferent between the delayed reform and the benchark reform. When $\omega > 1$ ($\omega < 1$) the planner strictly prefers the alternatuve (benchmark) reform.

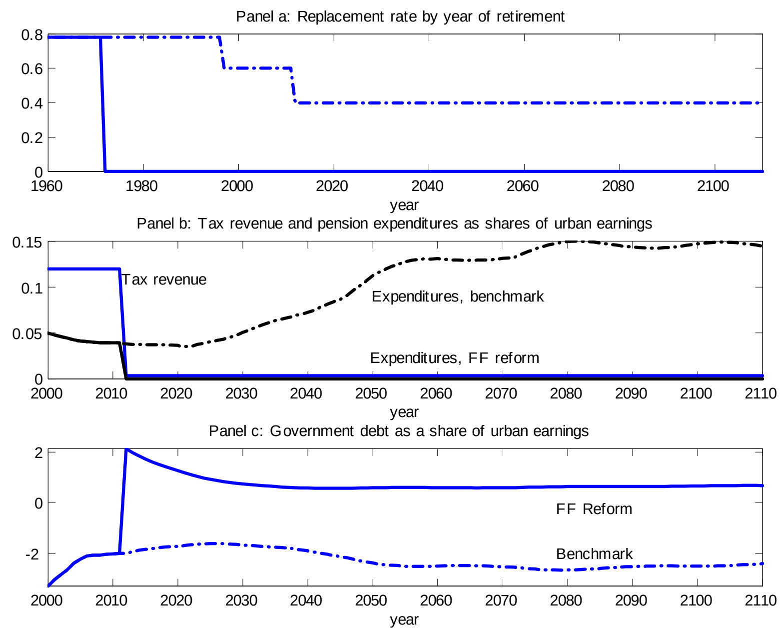

Figure 9: The figure shows outcomes for the fully funded reform (solid lines) versus the benchmark reform (dashed lines). Panel (a) shows the replacement rate, Panel (b) shows taxes (blue) and pension expenditures (green) expressed as a share of aggregate urban labor income, and Panel (c) the government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

labor earnings around 2035. Agents born after 2035 live in a low-tax society with no intergenerational transfers.

Panel (c) of Figure 7 shows the welfare effects of the fully funded reform relative to the benchmark. The welfare effects are now opposite to those of the delayed reforms. The cohorts retiring between 2012 and 2061 are harmed by the fully funded 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 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 12% for those retiring in 2030-35.[22] 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 2061. The fact that generations retiring sufficiently far in time gain is guaranteed by the assumption that the economy is dynamically efficient. However, the long-run gains are very modest, just about 0.2% in consumption equivalent terms. Both the high- and the low-discount planner strictly prefer the benchmark over the fully funded reform. The consumption equivalent discounted loss of a fully reform is 4.3% for the high-discount planner and 0.8% for the low-discount planner.

We view these results as interesting since a fully funded reform has been advocated in the policy debate to be the best response to adverse demographic dynamics. For example, Feldstein (1996) 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 (he assumes a 7% wage growth forever).

4.2.3 Pay as you go reform

We now analyze the effect of moving to a pure pay-as-you-go (PAYGO) system. In particular, we let the contribution rate remain 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:[23]

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

Figure **??** shows the outcome of this reform. Panel (a) reports the average replacement rate by year. Note that this figure is different from the corresponding panel shown in the previous experiments where the replacement rate was cohort specific and was reported by the year of retirement of each

[22] Averaging across cohorts retiring between 2012 and 2039 yields a 9% loss.

[23] Note 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 $\delta = 0.54\%$.

cohort). The PAYGO reform implies a higher replacement ratio than the benchmark reform until 2050. In fact, until 2040XXX the replacement ratio is higher than even the average 60% promised by the current rules. 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 2040. 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. This can be seen clearly in panels (b) and (c). Welfare gains are very pronounced for all cohorts retiring before 2040, 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 exceeding that suffered in the benchmark reform.

Due to the strong redistribution in favor of early generations, the utilitarian welfare is significantly higher than in the benchmark reform, under both a high- and low-discount planner. The consumption equivalent gains relative to the benchmark reform are, respectively, 12.9% and 1.7% for urban workers. These gains are larger than under all alternative reforms (delayed reform and fully funded 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 eight years, i.e., retirement age would have to increase from 58 to 66 years. 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 24), suggesting that the government implicitly anticipates a respource transfer from urban to rural areas. The modified consolidated government budget constraint then becomes: XXX MUST CORRECT THIS FORMULA

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

$$A_0 + \sum_{t=0}^{\infty} R^{-t} \left(\sum_{j=0}^{J_W} \zeta_j \left[\tau_t N_{t-j,t}\ w_t\ h_{t-j,t} + \tau_t^r N_{t-j,t}^r\ w_t^r\ h_{t-j,t}^r \right] - \sum_{j=J_W+1}^{J} \left[N_{t-j,t} b_{t-j,t} + N_{t-j,t}^r b_{t-j,t}^r \right] \right) \geq 0, \tag{4}$$

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 of new old-age programs that the Chinese government is currently introducing for rural areas.[24] 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 0.65% 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 more than 19%. When weighting rural and urban households by their respective population shares one obtains an aggregate welfare gain of 7.2% relative to the benchmark reform.[25]

The second experiment (*drastic reform*) consists of turning the Chinese pension system universal, pooling all Chinese workers and retirees – in both rural and urban areas – into a system with common

[24] This 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 ($XXX) per month (XXX REFERENCE). 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 XXX SOURCE). 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 (XXX REFERENCE) states that the rule of the new system shuld 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%.

[25] A high-discount planner who only cares for urban households participating in the pension system would incur a welfare loss of less than 0.7% from expanding the pension system to rural inhabitants. A high-discount planner who only cares for rural households would incur a welfare gain of 16.6%. When weighting rural and urban households by their respective population shares one obtains an aggregate welfare gain of 2.5% relative to the benchmark reform.

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 $q_t = 38.6\%$ 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 0.75% 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 (+35% if evaluated by the high-discount planner). Urban workers not participating in the system would also gain substantially (+16% if evaluated by the high-discount planner). The average effect (assessed from the standpoint of the high-discount planner weighting equally all inhabitants) is 16.8%.

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.

4.3 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*.

4.3.1 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% starting from 2012. In this case, the benchmark reform implies a replacement rate of 41.7%. 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 10 plots the welfare gains/losses of generations retiring betwen 2000 and 2110 in the case of a delay of the reform till 2040 (dashed line) and 2100 (continuous line). The top-right panel of Figure 10 yields the welfare gains/losses in the case of a fully funded reform. 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. The top left panel of Figure 12 yields the consumption equivalent gain accruing to the (high- and low-discount) planners as a function of the year in which the (delayed) reform is implemented.

Delaying the reform until 2040 (2100) yields a replacement rate of 40% (31.8%). 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 10% and 15%. 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 3.5%, which is significantly lower than the 5.2% gain found in the baseline economy. In the case of the low-discount planner, the gain is a mere third of that in the baseline economy. Thus, between one and two thirds 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 7%, compared with 9.5% 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 fully funded alternative reform harms earlier cohorts, whereas it benefits all cohorts retiring after 2050. 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 fully funded reform suffers a loss of 7% in the low wage growth economy, compared to a 12.2% loss in the baseline economy. Accordingly, the high-discount planner suffers a smaller welfare loss (2.2%) than in the baseline economy (4.3%). Thus, about half 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 fully funded reform over any of the alternatives. She would also prefer no delay to any of the delayed reforms.

Finally, the large welfare gains from the PAYGO alternative reform by and large vanish. While the high-discount planner woukd still prefer the PAYGO reform to the benchmark reform, the consumption equivalent gain would be a mere 3.9% relative to 12.9% in the high growth scenario. Perhaps more interesting, the low discount planner who has no built-in preference for earlier generations would now prefer the benchmark reform to the PAYGO reform. Thus, the welfare ranking order of the low discount planner is: fully funded 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 fully funded 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.

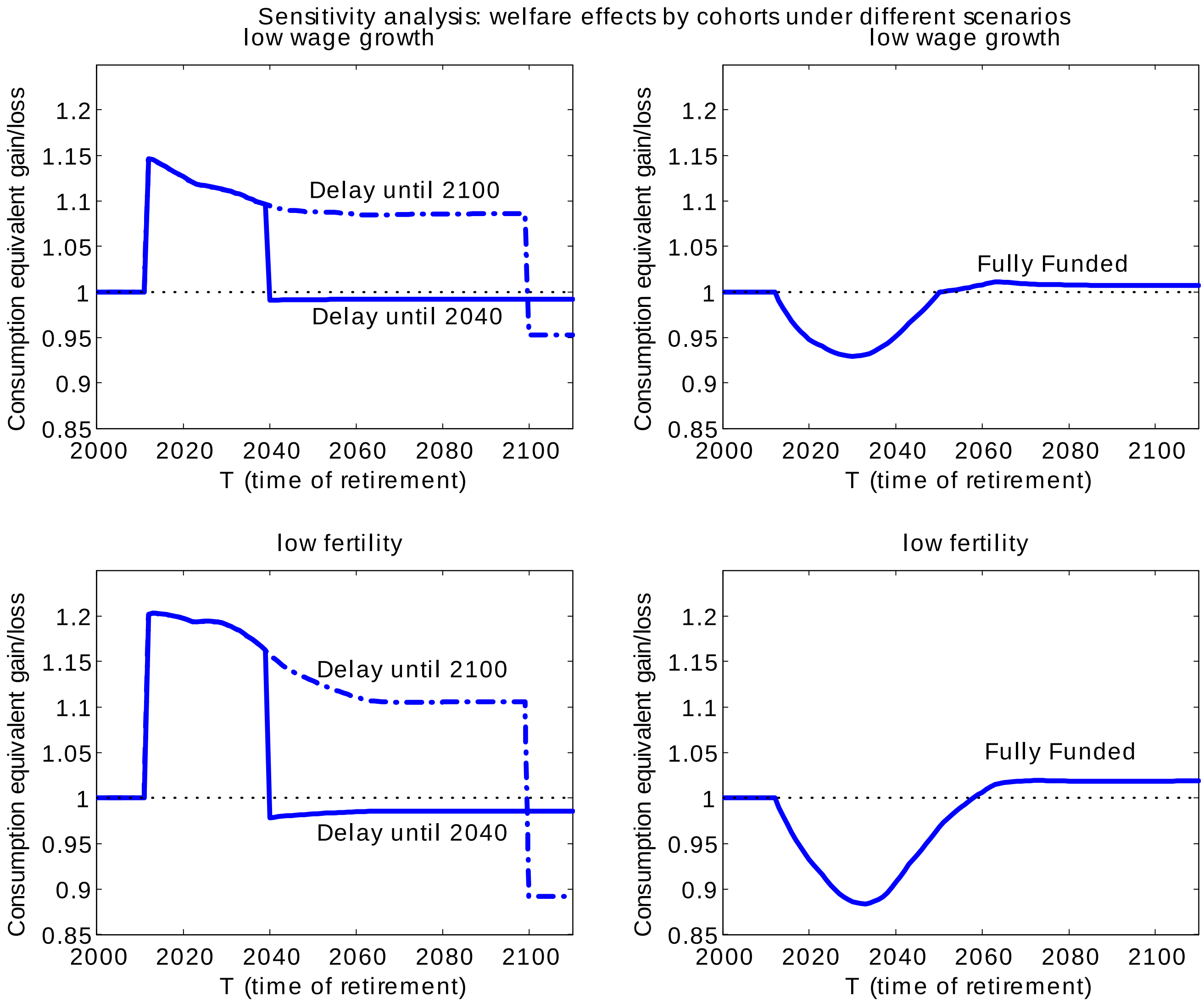

Figure 10: The figure shows consumption equivalent gains/losses accruing to different cohorts in two alternative scenarios. The top panels refer to the low wage growth scenario of section 4.3.1. The bottom panels refer to the low fertility scenario of section 4.3.2. The left-hand panels show the consumption equivalent gains/losses associated with delaying the reform until 2040 (solid lines) and 2100 (dashed lines), respectively. The right-hand panels show the consumption equivalent gains/losses associated with a fully funded reform.

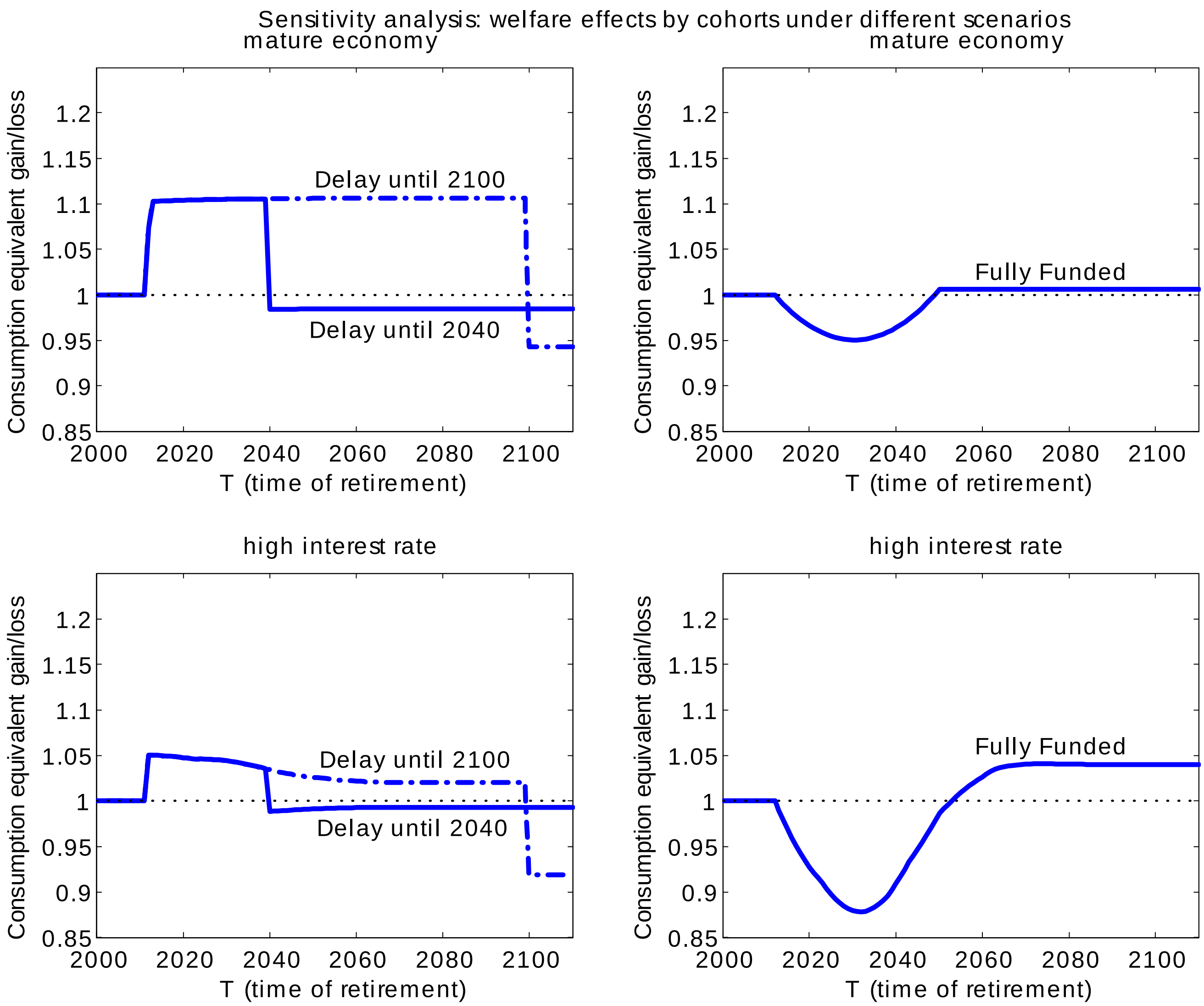

Figure 11: The figure shows consumption equivalent gains/losses accruing to different cohorts in two alternative scenarios. The top panels refer to the mature economy scenario of section ??. The bottom panels refer to the high interest rate scenario of section 4.3.3. The left-hand panels show the consumption equivalent gains/losses associated with delaying the reform until 2040 (solid lines) and 2100 (dashed lines), respectively. The right-hand panels show the consumption equivalent gains/losses associated with a fully funded reform.

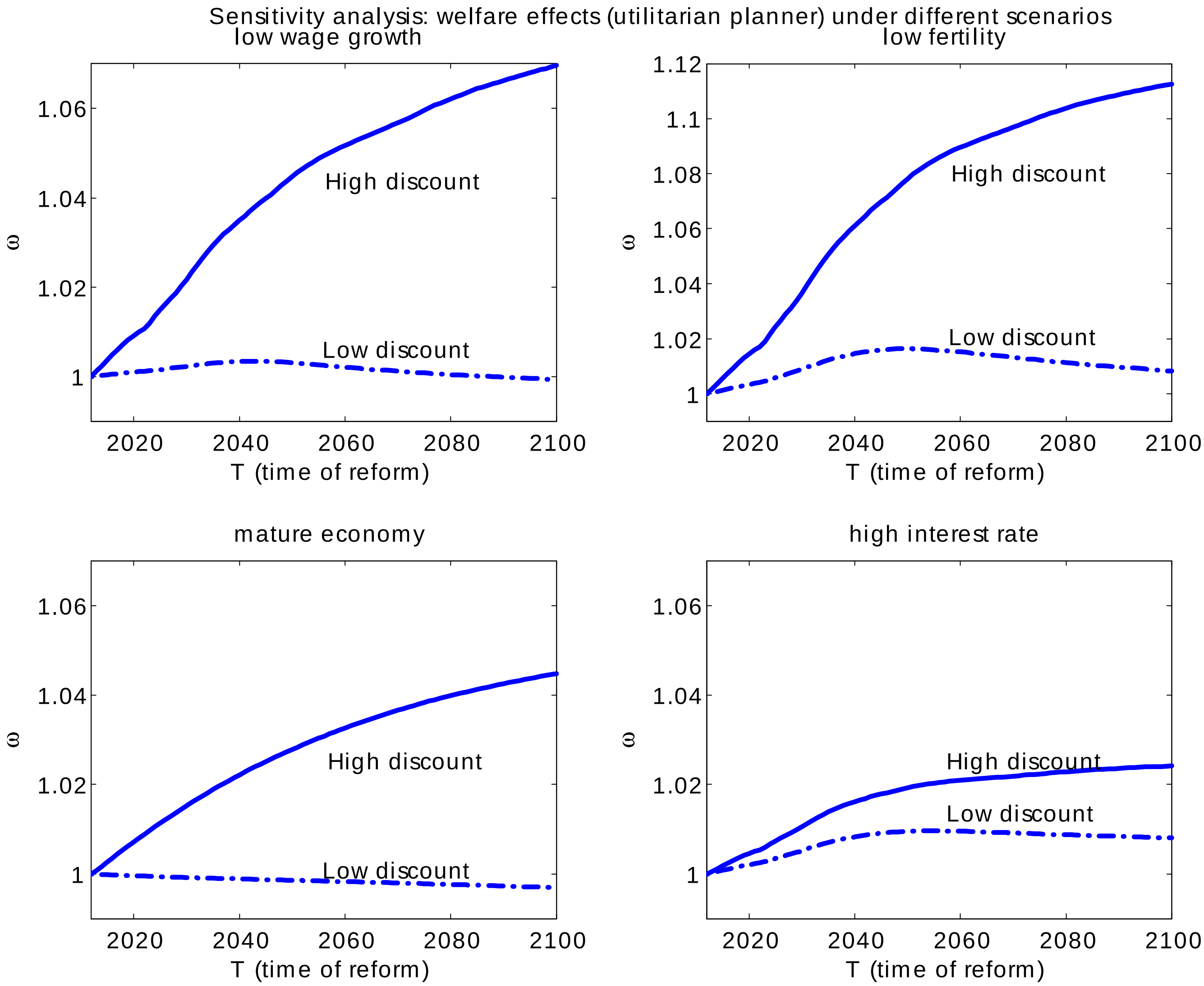

Figure 12: The figure shows the consumption equivalent gain/loss accruing to a high-discount planner (solid lines) and to a low-discount planner (dashed lines) of delaying the reform until time T relative to the benchmark reform. When $\omega = 1$, the planner is indifferent between the delayed reform and the benchark reform. When $\omega > 1$ ($\omega < 1$) the planner strictly prefers the alternatuve (benchmark) reform. Panel a refers to the low wage growth scenario of section 4.3.1. Panel b refers to the low fertility scenario of section 4.3.2. Panel c refers to the mature economy scenario of section **??**.

4.3.2 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 Yi (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 bottom panels of Figure 10 plot the welfare gains/losses of generations retiring betwen 2000 and 2110 in the case of a delayed and fully funded reform, respectively. The top right of Figure 12 yields the consumption equivalent gain accruing to the planners for delayed reforms.

Under this low-fertility scenario, the benchmark reform requires an even more draconian adjustment. The replacement rate must be set equal to 38% as of 2012. Delaying the reform is now substantially more costly. A reform in 2040 requires a replacement rate of 35%, whereas a reform in 2100 requires a replacement rate of 15.5%. 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 17% and 20%). 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 6.1% 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 11% if the reform is delayed until 2100. However, the welfare loss for the future generations is also large, equal to ca. 11%. 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 1.5% (0.8%). In all cases the gains are larger than in the baseline model. The losses associated with the fully funded reform are about the same as in the baseline model. Interestingly, the PAYGO reform yields larger gains than in the benchmark reform (14.9% with the high-discount and 4.6% with the low-discount planner, respectively).

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.

4.3.3 High interest rate

In this section, we consider a scenario in which the interest rate is equal to 6%. There are two main differences. First, delaying the reform yields a much smaller gain for the transitional generations, see the bottom left panel of Figure 11. Second, the fully funded reform entails larger gains for the future generations, see the bottom right panel of Figure 11. The most interesting comparison in this case is between PAYGO and fully funded.

XXX TO BE COMPLETED

4.4 Elastic labour supply

XXX TO BE WRITTEN: differences are very small

5 A dynamic general equilibrium model

Up to now we have taken the wages and the rate of return exogenous when analyzing pension reforms in China. As we demonstrated in section 4.3 the welfare effects depend significantly on the wage sequence. In this section, we construct a dynamic general equilibrium model that delivers the wage sequence assumed in the baseline model in section in Section 3 as an equilibrium outcome. Consequently, the allocation of consumption and savings of workers and retirees and the fiscal situation of th government will be exactly as analyzed above.

The model is closely related to Song et al. (2011), augmented with the demographic model and the pension system of Section 3. The key feature is that, due to asymmetric financial market imperfections, firms with different rates of return co-exist in a competitive economy for a prolonged time.

5.1 The production sector

The production sector consists of two types of firms: (i) *financially integrated* (F) firms, which are standard neoclassical firms; and (ii) *entrepreneurial* (E) firms which are owned by (old) entrepreneurs who are residual claimants on the profits and delegate the management of their firms to specialized agents, called the *managers*. The key assumption is that, due to contractual imperfections, only F firms have an unconstrained access to credit markets. E firms can run more productive technologies (see Song et al. 2011 for microfoundations of this assumption). However, they are subject to credit constrains that limits their size and growth. Due to such constraints, the less productive F firms may survive in equilibrium, although their employment and output share decline over time.

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

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

where Y is output and K and N denote capital and labor, respectively. 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 an 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 $K_F = AN_F \left(\alpha / \left(r^l + \delta\right)\right)^{-\frac{1}{1-\alpha}}$, where δ is the depreciation rate. R^l is determined by an arbitrage condition as will be shown below. As the capital-labor ratio is pinned down by r^l, so is the equilibrium wage:

$$w_t \geq (1-\alpha)\left(\frac{\alpha}{r^l+\delta}\right)^{\frac{\alpha}{1-\alpha}} A_t. \tag{5}$$

Strict inequality holds when $N_F = 0$, i.e., when F firms are inactive.

E firms are subject to credit constraints originating from informational imperfection, as in Song et al. (2011). In particular, a minimum 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, and by σ the maximum share of the E firm capital that can be financed externally. Then, the credit constraint yields

$$K_{Et} \leq (1+\sigma)\,\Omega_{E,t}. \tag{6}$$

Three regimes are possible: (i) the credit constraint (6) is binding and F firms are active (hence, the wage is pinned down by (5) holding with equality); (ii) the credit constraint (6) is binding and F firms are inactive; (iii) the credit constraint (6) is not binding and F firms are inactive. In regimes (ii) and (iii), (5) holds with strict inequality.

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

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

implying that K_{Et} is determined by past savings and investment decisions of entrepreneurs, and is a state variable. Moral hazard also plagues the relationships between old entrepreneurs and young managers, and an incentive constraint requires that managers be paid in each period a share of the firms' revenue, denoted by ψ.[26] Profit maximization yields, then, the following optimal labor hiring decision:

$$\begin{aligned} N_{Et} &= \arg\max_{\tilde{N}_t}\left\{(1-\psi)\,(K_{Et})^{\alpha}\left(\chi A_t \tilde{N}_t\right)^{1-\alpha} - w_t \tilde{N}_t\right\} \\ &= ((1-\psi)\,\chi)^{\frac{1}{\alpha}}\left(\frac{r^l+\delta}{\alpha}\right)^{\frac{1}{1-\alpha}} \frac{K_{Et}}{\chi A_t}. \end{aligned} \tag{8}$$

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, (5) and (8). We assume that $(1-\psi)^{\frac{1}{\alpha}}\,\chi^{\frac{1-\alpha}{\alpha}} > 1$ ensuring that the return to capital is higher in E firms. Note that in regime (i) the rate of return to capital is a linear function of r_t^l in both E and

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

F firms. The equilibrium of regime (i) is closed by the condition that employment in the F sector is determined residually, namely,

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

Consider, next, regime (ii), such that only the E sector is active, and thus $N_{E,t} = N_t$, and the borrowing constraint is binding, hence, (7) holds. In this case, the rate of return to capital and labor equal the respectively marginal products. Then:

$$w_t = (1-\alpha)(1-\psi)(\chi A_t)^{1-\alpha} (K_{E,t}/N_t)^{\alpha},$$

and the gross rate of return on entrepreneurial wealth is given by

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

In regime (ii), the stock of capital continues to be 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 entrepreneurs can earn on alternative investment opportunities (e.g., bonds). Namely,

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

and $K_{E,t}$ ceases to be a state variable, being determined by r_t^l. More formally, $K_{Et} = \chi A \left(\alpha / \left(r_t^l + \delta\right)\right)^{1/(1-\alpha)} N_t$, and $w_t = (1-\alpha)\left(\alpha / \left(r_t^l + \delta\right)\right)^{\alpha/(1-\alpha)} \chi A_t$.

The law of motion of entrepreneurial wealth is determined by the optimal managers' and entrepreneurs' saving decisions which are discussed below.

5.2 Banks

Competitive financial intermediaries (*banks*) with an access to perfect international financial markets collect savings from workers and invest them into 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. 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.

5.3 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" and acquire through their managerial experience skills enabling them to become "entrepreneurs" when, at age J_E, they turn old. Total managerial compensations equal $M_t = \psi Y_E$. 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 – but can become entrepreneur. Namely, they can invest their wealth (or part of it) into setting up their own business, and hire managers and workers. Thereafter, they earn no labor earnings but become the residual claimants of the firm's profits. We assume that entrepreneurs are not in the pension system. Their lifetime budget constraint equals, then, is given by:

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

5.4 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 productivity (A_0) and the population distribution. The key economic force is the savings motive of managers and entrepreneurs. If the economy starts in regime (i), then all managerial savings are invested in the entrepreneurial business as soon as a manager becomes an old entrepreneur. As long as this drives capital accumulation at a sufficiently high rate in the entrepreneurial sector, the employment share of E firms will grow and that of F firms will decline over time. The key parameters are the time discount rate, the rate of return in the E firms and the world interest rate. In particular:

- a high β imply 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 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;

- 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.

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.

5.5 Calibration

This section [including figures reported] is very preliminary.

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 set exogenously: $\alpha = 0.5$ so that the capital share of output is 0.5 (Bai et al. 2006), and $\delta = 0.1$ yields a 10% annual depreciation rate of capital.

The remaining parameters are calibrated internally, so that the model matches a set of empirical moments. We set the parameters ψ and χ so as to match two moments, assuming that China is initially in the first phase of the transition: (1) match a K/Y ratio in E-firms of 50% of the corresponding ratio in F-firms. This ratio is what Song et al (2011) document for manufacturing industries, after controlling for industry type. (2) match a rate of return on capital which is 9% larger in E-firms than in F-firms during the first phase of the transition.[27] The implied parameter values are $\psi = 0.267$ and $\chi = 2.729$. This implies that TFP of an E-firm is 1.65 times larger than TFP of an E-firm.[28]

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%. With an average DPE employment share of 15% during the 1998-2000 period, 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 beleive 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 $\sigma = 0.89$, so that entrepreneurs can borrow 89 cents for each dollar in equity. This value for σ implies that the growth in DPE employment share is in line with the private employment growth between 2000 and 2008 in urban areas. We set the secular exogenous productivity growth, i.e., the growth in A_t, so that the model generates an aggregate GDPpc growth of XXX% for China during 2000-2009. The resulting growth rate in A_t is 2.41% larger than the associated world growth rate from

[27] Song et al. (2011) document that in manufacturing, DPEs have on average a ratio of profits per unit of book-value capital 9% larger 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).

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

1992 to 2011. 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 XXX[PRELIMINARY]. This occurs in year 2025.

The initial conditions are set as follows. The total entrepreneur wealth in 2000 is set equivalent to 9% of GDP so that the 2000 DPE employment is 20%. The distribution of that entrepreneur wealth is obtained by assuming that in 1992 all entrepreneurs are endowned 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 80% of GDP in 2000 so as to generate a net foreign surplus equal to 20% of GDP in 2000.

5.6 Simulated output trajectories

The calibrated model yields growth forecast that we view as plausible. Figure 13 shows the evolution of productivity and output per capita forecasted by our model. The average growth rate of GDP per worker remains very high until 2020, when it remains stable above 9% per year (see upper panel). Although productivity growth is forecasted to slow down thereafter, due to the end of the urban transition, it is expected to remain above 5% for the following decade. Productivity convergences continue till 2040 and then dies off. Note that the growth of GDPpc is lower than that of GDPpw after 2012, due to the increase in the dependency ratio.

Under the conservative assumption that the (exogenous) rate of technological convergence will decline in the coming years, this implies that China is expected to grow at an average 6% between 2012 and 2040. The contribution of human capital is XXX% per year. In this scenario, the GDP per capita of China will be about 80% of that of the US by 2045, remaining stable therafter. Assuming an average GDP per capita growth in advanced economies of 2% per year, China is set to surpass the United States in terms of total GDP in 2015 and to become more than twice as large as the US level in the long run.

The wage sequence that was assumed in section 3 is now endogenous. Wages are forecasted to grow at an average 6.2% until 2030, and to slow down thereafter.

5.7 Sensitivity

This section is very preliminary.

In this section, we perform sensitivity analysis.

5.7.1 High savings and foreign surplus

Although the growth forecasts appear to be sensible, the calibrated economy generates a very large amount of savings. For instance, by 2070 the economy has a wealth-GDP ratio equal to XXX%. This implies a massive foreign surplus, which reaches XXX. To understand the reason, it is useful to recall

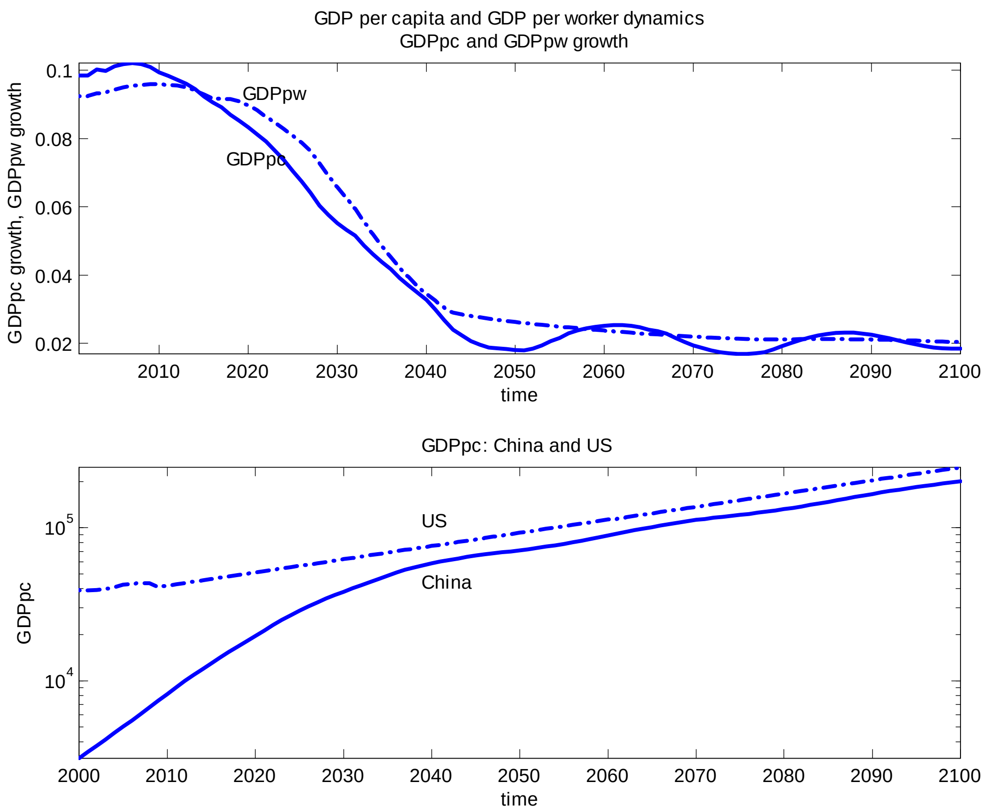

Figure 13: The upper panel reports the forecasted GDPpc and GDPpw annual growth, according to the calibrated model of section 5. The lower panel reports the forecasted GDPpc of China and the US (logarithmic scale). GDPpc is assumed to grow at an annual 2%rate in the US after 2011.

that the model is calibrated to match the growth performance in the first decade of the century. In that period, China has had a high growth rate (that we expect to continue in the future as explained above), and yet a very high saving rate (XXX add some figures here). In our model, such a high saving rate can only be matched by choosing a high discount factor ($\beta = 1.012$). Should this discount factor remain so high, China would have an even high saving rate in future. In addition, entrepreneurial firms will grow out of the financial constraint by XXX, and part of the entrepreneurial savings will also be invested in foreign assets.

In our model, a high β is a stand-in for a number of institutional features that are not explicitly considered and that may imply a high propensity to save over and beyond pure preferences. For instance, they may partly be due to high precautionary savings for uninsurable idiosyncratic shocks. One might argue that it is implausible that the Chinese economy will have so high saving rates as forecasted by our calibrated model.

It is important to note that in our model the long term macroeconomic performance do not hinge on the domestic propensity to savings. Domestic capital accumulation and wages are determined in the long run by the world interest rate. Thus, β only determines the extent of the foreign position.[29] Yet, one might worry that the quantitative results of the welfare analysis may be sensitive to the assumption about β.

To address this concern, we considered an alternative scenario where all cohorts born after 2012 have a lower β. In particular, we target the level of β so as to ensure that China has a net foreign position of zero in the long run. We view this as a conservative assumption. The results are shown in the Appendix [XXX not included in this version]. The analysis of the alternative pension arrangements yields essentially the same results as in the high-β economy. Thus, the calibration of β of the workers is unimportant.

5.7.2 Financial development

The model borrows from Song et al. (2011) the assuption that E firms are financially constrained. It is important to note that the importance of the financial constraint to which entreppreneurial firms are subject declines over time. In particular, as the economy enters stage (iii) the share of external financing declines and the financial constraint is no longer binding. In stage (iv), firms are no longer externally financed and entrepreneurs hold positions in foreign assets. The calibrated economy enters stage (iii) in XXX

However, in the baseline calibration, the parameter that regulates the financial constraint is assumed to be constant over time. It may be interesting to know what would happen if this constraints were relaxed earlier on due to financial development. An increase in σ would clearly accelerate the short-run growth rate, causing a faster convergence to the end of the transiton. Wage growth would accelerate earlier but so would its posterior decline.

[29] Note, though, that the propensity to saving of the entrepreneurs is important in determining the speed of the transition.

To see the effects of financial development, we consider a stark experiment in which all financial frictions are removed in 2012. This includes both the borrowing constraint on E firms and the iceberg intermediation cost. While this has a large effect on growth path (the GDP per worker more than doubles at the time of the reform, there are no major differences in the welfare analysis. Both the gains from delaying the reform and the losses associated with a fully funded reform are smaller.

[INCOMPLETE..]

6 Conclusions

To be written

References

Abel, Andrew, N. Gregory Mankiw, Lawrence H. Summers, and Richard J. Zeckhauser, "Assessing Dynamic Efficiency: Theory and Evidence", *Review of Economic Studies*, Vol. 56, No. 1 (Jan., 1989), pp. 1-19.

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.

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.

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) pp. 173-188.

Hu, Ying, 2003, Quantitative Analysis of the Population Transfers from Rural Area to Urban Area (in Chinese), Statistical Research, 2003(7), 20-24.

Hurd, Michael D. "Mortality Risk and Bequests." *Econometrica* 57 (July 1989): 779-813.

Johnson, Gale D., 2002, Provincial Migration in China in the 1990s, *China Economic Review*, 14, 22-31.

Lavely, 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.

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, "Critical Assumptions in the Stern Review on Climate Change," *Science*, 13 July 2007, Vol. 317. no. 5835, pp. 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.

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, 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.