NBER WORKING PAPER SERIES

RISK SHARING, INEQUALITY AND FERTILITY

Roozbeh Hosseini
Larry E. Jones
Ali Shourideh

Working Paper 15111
http://www.nber.org/papers/w15111

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2009

We are indebted to Alice Schoonbroodt for all of her helpful comments and suggestions at various stages of this project. We would also like to thank V.V.Chari, Chris Phelan and Ariel Zetlin-Jones for comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Risk Sharing, Inequality and Fertility
Roozbeh Hosseini, Larry E. Jones, and Ali Shourideh
NBER Working Paper No. 15111
June 2009
JEL No. C61,D30,D63,D64,H21,H23,H43

ABSTRACT

We use an extended Barro-Becker model of endogenous fertility, in which parents are heterogeneous in their labor productivity, to study the efficient degree of consumption inequality in the long run. In our environment a utilitarian planner allows for consumption inequality even when labor productivity is public information. We show that adding private information does not alter this result. We also show that the informationally constrained optimal insurance contract has a resetting property - whenever a family line experiences the highest shock, the continuation utility of each child is reset to a (high) level that is independent of history. This implies that there is a non-trivial, stationary distribution over continuation utilities and there is no mass at misery. The novelty of our approach is that the no-immiseration result is achieved without requiring that the objectives of the planner and the private agents disagree. Because there is no discrepancy between planner and private agents' objectives, the policy implications for implementation of the efficient allocation differ from previous results in the literature. Two examples of these are: 1) estate taxes are positive and 2) there are positive taxes on family size.

Roozbeh Hosseini
Department of Economics
W. P. Carey School of Business
Arizona State University
P.O. Box 873806
Tempe, AZ 85287-3806
roozbeh.hosseini@asu.edu

Larry E. Jones
Department of Economics
University of Minnesota
4-101 Hanson Hall
1925 Fourth Street South
Minneapolis, MN 55455
and NBER
lej@umn.edu

Ali Shourideh
Department of Economics
University of Minnesota
4-101 Hanson Hall
1925 Fourth Street South
Minneapolis, MN 55455
shour004@umn.edu

1 Introduction

A key question in normative public finance is the extent to which it is socially efficient to insure agents against shocks to their circumstances. The basic trade-off is one between providing incentives for productive agents to work hard – thereby making the pie big – versus promising to transfer resources from more productive agents to less to insure them against the possibility of being poor. This is the problem first analyzed in Mirrlees (1971) where he characterized the solution to a problem of this form in a static setting. More recently, a series of authors (e.g., Green (1987), Thomas and Worrall (1990) and Atkeson and Lucas Jr. (1992)) have extended this analysis to cover dynamic settings – agents are more productive some times and less others. A common result from this literature is that the socially efficient level of insurance (ex ante and under commitment) involves an asymmetry between how good and bad shocks are treated. In particular, when an agent is hit with a bad shock, the decrease in what he can expect in the future is more than the corresponding increase after a good shock – there is a negative drift in expected future utility. This feature of the optimal contract in dynamic settings has become known as 'immiseration.'

Two recent papers, Phelan (2006) and Farhi and Werning (2007) have given a novel interpretation to the immiseration result. This is to interpret different periods in the model as different generations. In this interpretation, current period agents care about the future because of dynastic altruism a la Barro (1974). Thus, parents care about the level of consumption of their children, their grandchildren, etc. Under this interpretation, social insurance is comprised of two conceptually different components. These are: 1) Insurance against the uncertainty coming from current generation productivity shocks, 2) Insurance against the uncertainty coming from the shock of what family you are born into – what future utility was promised to your parents (e.g., through the intergenerational transmission of wealth).

In this case, optimally provided social insurance against the bad luck of lower than average productivity will not necessarily lead to an outcome in which more and more of the wealth of society is concentrated among an ever shrinking population – i.e., immiseration will not necessarily occur. Indeed, whether or not this occurs is completely determined by whether the planner discounts the utility of subsequent generations in the same way that parents do, or if, to the contrary, the planner puts additional extra weight on future generations over and above that given by altruistic parents.

When this is the case, Phelan (2006) and Farhi and Werning (2007) show that a feature of the optimal insurance scheme is that there is a stationary distribution of consumption, etc. This stationary distribution is what determines the long run level of inequality in an efficient allocation. It depends on the extent to which society weighs the utilities of future generations beyond the minimum that occurs due to dynastic altruism.

In an intergenerational setting with dynastic altruism such as what is studied by these authors, a natural question arises: To what extent are these results altered when the size of generations – i.e., fertility – is itself endogenous as in the model of Barro and Becker (1989) and Becker and Barro (1988)?

The study of this question is the focus of this paper.

We show that the explicit inclusion of the fertility dimension into the model alters the qualitative character of the optimal allocations in two important ways. First, we show that even when the planner does *not* put extra weight on future generations, there is a stationary distribution in per capita variables. That is, per child consumption, labor supply and expected future utility all follow a stationary probability law even when social and private discount factors agree – there is no immiseration in per capita terms. In addition to this, since fertility is explicitly included, the model has implications about the properties of fertility in an ex ante socially efficient arrangement (for example, fertility also follows a stationary law of motion). Because of this, the model has implications about the best way to design policies relating to fertility choice (e.g., child care deductions, tax credits for children, education subsidies, etc.) Incidentally, this socially efficient allocation also has features observed in the data, such as a negative income-fertility relationship, intra-family insurance provision and inter-vivos transfers. It therefore seems like the inclusion of fertility is a very natural way of getting away from immiseration.

To develop these intuitions, we analyze a model in which labor productivity is private information and varying over time – here interpreted as different generations. Thus, an individual can be born with high or low productivity and only he knows this. He cares about his own consumption, the size of his family (a la Barro and Becker (1989)) and the utility level that each of his children will enjoy. Each of his children is also subject to a labor productivity shock assumed to be i.i.d. Thus, this is a model of dynamic private information with i.i.d. shocks in which different periods are interpreted as different generations. The difference between this and the previous literature is that family size is endogenous and directly affects the utility of the parent.

From a mechanical point of view, the inclusion of fertility gives the planner an extra instrument to use to induce current agents to truthfully reveal their productivities. That is, the planner can use both family size and continuation utility of future generations.

In the normal case (i.e., without fertility choice), in order to induce truth telling today, the planner (optimally) chooses to 'spread' out continuation utilities so as to be able to offer insurance in current consumption. The incentives for the planner to do this are present after every possible history. That is, the continuation utility increases after a high shock and decreases after a low shock. Hence, continuation utilities are pushed ever outward. Under the usual Inada conditions, this outward pressure is asymmetric and has a negative bias – it is cheaper to provide incentives in the future when continuation utilities are lower. Because of this continuation utilities are pushed to their lower bound – inequality becomes greater and greater over time.

In contrast, when fertility is endogenous, this optimal degree of spreading in continuation utility for the parent can be thought of as being provided through two distinct sources – spreading of per child continuation utility and spreading in family sizes through differential fertility. In general both of these instruments are used to provide incentives, but, the way that they are used is different. We find that there is a natural limit to the amount of spreading that is done through per child continuation utilities. There are two manifestations of this. First, there is an upper bound on continuation utilities that is never exceeded. So, even if a family has a very long series of good shocks, the utility of the children does not continue to grow but stays at a fixed, high level, w_0. Second, this same level of continuation utility is used as a reward for the children of currently highly productive workers no matter what their past history was (i.e., the previous continuation utility). That is, continuation utility gets 'reset' to w_0 subsequent to every realization of a high shock. Thus, there is a limited amount of inequality in per capita consumption, labor supply, etc., in the long run.

This reasoning concerning the limits of long run inequality in per capita variables differs from what is found in Farhi and Werning (2007) in two ways. The first of these is the basic reason for the breakdown in the immiseration result. In Farhi and Werning (2007), it is because of a difference between social and private discounting – society puts more weight on future generations than parents do. Here, immiseration breaks down even when social and private discounting are the same. The second difference concerns the movements of per capita consumption over time. The version of the Inverse Euler Equation that holds in the Farhi and Werning (2007) world when society is more patient than individuals implies that consumption has a mean reversion property. In our model, as discussed above, the resetting property implies that consumption (and continuation utility, etc.) revert to the 'top' of the distribution each time a high shock is realized.

The intuition behind the resetting property is the product of two complementary effects. The first of these concerns the assignment of continuation utility when there is no private information problem. Because of the form of the Barro and Becker (1989) dynastic utility function, continuation utility, from the perspective of the parent is a homothetic function of family size and per child continuation utility. Because of this, the relative breakdown between these two is scale independent. Thus, the increase in total family continuation utility in response to an increase in wealth (i.e., parent continuation utility) is equal to the increase in total family size – per child continuation utility is independent of wealth. It follows that, under full information, the continuation utility of a child depends on the productivity of his parent, but not on the continuation utility of the parent – i.e., continuation utility of children is reset in a way that is independent of the family history of shocks. When productivity levels are private information this argument continues to hold, but only for those types whose first order conditions are undistorted due to the fact that no one wants to pretend to be them. Since this always holds for the highest type, it follows that we have resetting at any time the highest possible shock occurs.

In contrast to this, we find that there is no upper bound on the amount of spreading that occurs through the choice of family size. What we find is that along any subset of the family tree, the population dies out if the discount factor is the inverse of the interest rate. However, this does not necessarily imply that population shrinks, since this property holds even when mean population is growing. Rather, some strands of the dynasty tree die out and others expand. Those that are growing are exactly those sub-populations that have the best 'luck.'

It's also true that the driving force responsible for getting rid of the immiseration result here is substantially different from previous work. From a mechanical point of view, the allocations that are considered by these authors can equivalently be thought of as putting lower bounds on the continuation utility levels of future generations in a problem where the social and private discount factors are the same. As such, they are closely linked to the approach followed in Atkeson and Lucas Jr. (1995) and Sleet and Yeltekin (2006). Here, the reason is different. It is because of the inclusion and optimal exploitation by the planner of a new choice variable, family size. Key properties of this new variable are that it is observable by the planner and that the way it enters utility is to affect total continuation utility of the parent.

There are several other interesting differences between the two approaches. For example, one of the key ways that Phelan (2006) and Farhi and Werning (2007), differ from earlier work is that in socially efficient scheme, the Inverse Euler Equation need not hold. Indeed, in those papers, the inter-temporal wedge is always negative in contrast to earlier papers. This can be interpreted (in some implementations) as requiring a negative estate tax. This has been interpreted as meaning that, in order to overcome immiseration, a negative inter-temporal wedge was necessary. This does not hold for us however, since a version of the Inverse Euler Equation holds. This implies that there will always be a positive 'wedge' in the FOC determining savings and hence, estate taxes are always implicitly positive.

An interesting new feature that emerges is the dependence of taxes on family size. What we find is that for everyone other than the highest type, there is a positive tax wedge on the fertility-consumption margin – fertility is discouraged to better provide incentives for truthful revelation.

Another difference between the two approaches concerns the degree of intervention required to realize the optimal allocation. When social and private discounting agree (as is true here), the socially efficient allocation can be implemented through a one time redistribution and a strong legal system to enforce private bequest contracts. These bequest contracts strongly resemble intergenerational transfers observed in the data: poorer children tend to receive transfers from luckier family members. In contrast, in Phelan (2006) and Farhi and Werning (2007), there is a persistent difference between the preferences of the planner and the agent, and because of this, persistent intervention is required.

1.1 Related Literature

Our paper is related to the large literature on dynamic contracting including Green (1987), Thomas and Worrall (1990), Atkeson and Lucas Jr. (1992) (and many others). These papers established the basic way of characterizing the optimal allocation in endowment economies where there is private information. They also show that, in the long run, inequality increases without bound, i.e. the immiseration result. Phelan (1998) shows that this result is robust to many variations in the assumptions of the model. Moreover, Khan and Ravikumar (2001) establish numerically that in a production economy, the same result holds and although the economy grows, the detrended distribution of consumption has a negative trend. We contribute to this literature by extending the model to allow for endogenous choice of fertility. We employ the methods developed in the aforementioned papers to analyze this problem.

As mentioned before, a number of earlier papers have developed models with private information where there is no immiseration, including Atkeson and Lucas Jr. (1995), Phelan (2006), Sleet and Yeltekin (2006), and Farhi and Werning (2007). One feature that is shared by these models is that they are mathematically equivalent to a problem with a lower bound on continuation utility. Our paper differs from these in that the basic mechanism that drives the result is different. Because of this, there are also different implications about evolution of consumption, etc. In addition, our paper differs from the earlier work on on dynamic optimal taxation as in Golosov, Kocherlakota, and Tsyvinski (2003), Albanesi and Sleet (2006), and Golosov and Tsyvinski (2006) by characterizing to the design of optimal taxes when fertility is endogenous choice variable.

Finally, our paper has some novel implications about fertility per se. First, the socially efficient allocation is characterized by a negative income-fertility relationship—independently on specific assumptions on curvature in utility (see Jones et al. (2008) for a recent summary). Given the realism of our results on implementation of this socially efficient allocation, this suggests that intergenerational income risk and intra-generational risk sharing may be important factors to explain the observed negative income-fertility relationship. What our model has in common with other theories that deliver this relationship with ability heterogeneity is that child costs are in terms of leisure (time). Second, very few papers have analyzed ability heterogeneity and intergenerational transmission of wealth in dynastic models with fertility choice. Our paper is most related to Alvarez (1999) who analyzes intergenerational income risk but assumes that it is uninsurable (i.e. all children receive the same realization and there is no insurance markets across dynasties) and does therefore not address the question of risk sharing. For the most part of his paper, he rules out time costs and therefore finds that fertility is increasing in both, ability and wealth, while bequests are independent on both. Hence, this version of the model generates no intergenerational persistence of wealth, beyond that coming from intergenerational correlations in ability shocks.[1] Our model generates such persistence even with i.i.d. shocks.

In section 3 the benchmark model is laid out as well as its properties. Section 4 contains the general model with private information. In section 5 we study the main implications of the private information model about long-run inequality. Section 5 is devoted to discussing implementations of efficient allocations in our environment. Finally, in section 6, we analyze numerical examples.

2 A Two-Period Example

In this section we present the key ideas in the paper in the context of a two period example. We start by characterizing the ex ante efficient allocation under full information and then go on to add private information.

Consider a two-period economy populated with a continuum of parents with mass 1 who live for one period. Each parent receives a random productivity θ in the set $\Theta = \{\theta_L, \theta_H\}$ in which $\theta_H > \theta_L$. [2] At date 1, each parent's productivity, θ, is realized, they consume, work and decide about the number of children. The cost of having a child is in terms of leisure. Every child requires b units of leisure to raise. The coefficient b can be thought of as market

[1]When analyzing how crucial his assumptions are he finds that when per child costs are allowed to depend on wages (i.e. a time cost), fertility is decreasing in ability and increasing in wealth, as long as utility is more curved than log (i.e. number and utility of children are substitutes).

[2]We are normalizing population in date-1 to 1, which we will relax later.

value of maternity leave for women. The child lives for one period and consumes out of the savings done by their parents. The parent's utility function is the following:

$$u(c_1) + h(1 - l - bn) + \beta n^\eta u(c_2)$$

where l is hours worked, n is the number of children and c_t is consumption per person in period t. From this, it can be seen that the parent has an altruistic utility function where the degree of altruism is determined by β. A worker of productivity $\theta \in \Theta$ who works for l hours has effective labor supply of θl. For now, we assume that everything is observable to the planner.

In what follows we denote the aggregate consumption of all children by $C_2 = n_2 c_2$. We will use the following, technical conditions:

Assumption 1 *1. The functions u and h are strictly increasing, strictly concave and continuously differentiable;*

2. The function $n^\eta u(C/n)$ is strictly quasi-concave in (C, n).

Suppose each parent is promised an ex ante utility W_0 at date zero. We study the problem of a planner that has access to a saving technology at rate R and wants to allocate resources efficiently such that the ex ante utility to each parent is at least W_0.

$$\begin{aligned} \min_{c_1(\theta), n(\theta), l(\theta), c_2(\theta)} \quad & \sum_{\theta \in \Theta} \pi(\theta) \left(c_1(\theta) + \frac{1}{R} n(\theta) c_2(\theta) \right) - \sum_{\theta \in \Theta} \pi(\theta) \theta l(\theta) & (1) \\ \text{s.t.} \quad & \sum_{\theta \in \Theta} \left[\pi(\theta) \left(u(c_1(\theta)) + h\left(1 - l(\theta) - bn(\theta)\right) + \beta n(\theta)^\eta u(c_2(\theta)) \right) \right] \geq W_0 \end{aligned}$$

Taking first order conditions, we can write the optimality conditions as follows:

$$c_1(W_0, \theta) = c_1(W_0, \theta') \qquad \forall \theta, \theta' \in \Theta \tag{2}$$

$$\theta u'(c_1(W_0, \theta)) = h'\left(1 - l(W_0, \theta) - bn(W_0, \theta)\right) \tag{3}$$

$$\beta R n(W_0, \theta)^{\eta - 1} u'(c_2(W_0, \theta)) = u'(c_1(W_0, \theta)) \tag{4}$$

$$\begin{aligned} \beta \eta n(W_0, \theta)^{\eta - 1} u(c_2(W_0, \theta)) = {} & bh'\left(1 - l(W_0, \theta) - bn(W_0, \theta)\right) \\ & + \beta n(W_0, \theta)^{\eta - 1} u'(c_2(W_0, \theta)) c_2(W_0, \theta) \end{aligned} \tag{5}$$

The interpretation of equation (2) is standard – it is the full risk sharing condition that is common in Mirrlees problems without informational frictions and separable preferences.

Equation (3) is also standard. It represents the static trade off between leisure and consumption for different types. Using this along with equation (2), it follows directly that higher productivity types get less leisure. Equation (4) is the Euler equation, augmented by the fact that population is growing. Thus, the rate of return on capital, R, should be corrected by the rate of population growth, $n^{\eta-1}$. Equation (5) is new here. It equates the marginal value of having an extra child to the marginal cost of having an extra child. The two terms on the right hand side reflect the two components of costs that come from an increase in fertility. First is the value of the extra leisure that is used to raise the additional child. Second, is the reduction in per capita consumption that will occur in the second period if fertility is increased but nothing else is changed. The left hand side is the utility value of the increase in fertility.

Replacing (3) and (4) in (5), gives the following equation:

$$\begin{aligned} \beta\eta n(W_0,\theta)^{\eta-1}u(c_2(W_0,\theta)) &= \beta n(W_0,\theta)^{\eta-1}u'(c_2(W_0,\theta))c_2(W_0,\theta) && (6)\\ &\quad +b\theta\beta R n(W_0,\theta)^{\eta-1}u'(c_2(W_0,\theta)) && (7)\end{aligned}$$

which simplifies to

$$\eta u(c_2(W_0,\theta)) = u'(c_2(W_0,\theta))c_2(W_0,\theta) + b\theta R u'(c_2(W_0,\theta)) \tag{8}$$

As it can be seen from (8), the consumption of each child, c_2, only depends on θ and does not depend on promised utility to the parent, W_0. This is in sharp contrast to the special case in which fertility is not a choice. In those environments risk sharing implies that consumption of each child is independent of parent's shock θ and depends directly on parents promised utility.

To gain insight about this result we rewrite problem (1) slightly differently. Let $m = 1-l-bn$ be parents leisure. Consider the problem of social planner who minimizes the cost of allocating parents time between leisure and parenting as well as consumption to parents ad children:

$$\begin{aligned} \min_{c_1(\theta),n(\theta),m(\theta),C_2(\theta)} \quad & \sum_{\theta\in\Theta}\pi(\theta)\left(c_1(\theta)+\theta m(\theta)\right)+\sum_{\theta\in\Theta}\pi(\theta)\left(b\theta n(\theta)+\frac{1}{R}C_2(\theta)\right) && (9)\\ \text{s.t.} \quad & \sum_{\theta\in\Theta}\pi(\theta)\left(u(c_1(\theta))+h\left(m(\theta)\right)\right)+\sum_{\theta\in\Theta}\pi(\theta)\beta n(\theta)^{\eta}u\left(\frac{C_2(\theta)}{n(\theta)}\right)\geq W_0 \end{aligned}$$

This is the same as problem (9) except that we substituted out hours, $l(\theta) = 1-m(\theta)-bn(\theta)$ and per child consumption, $c_2 = \frac{C_2}{n}$, and rearranged terms in the objective function. The first term in the objective function is planner's expenditure on parents' consumption and leisure (denominated in parents' consumption). The second term is the total expenditure on children: their total consumption and time spent parenting (again, denominated in parents' consumption).

We can separate this problem in two stages. In stage one planner decides how much consumption and leisure to give to parents and also how much utility each parent should get from having children. In the second stage planner decides how many children each parent should have and total consumption of all children for each parent. We solve this problem backward. Suppose the planner has decided in the first stage that a type θ parent will enjoy utility W from having children. Then in the second stage the planner solves the following problem for the parent who has shock θ:

$$\hat{V}(W,\theta) = \min_{C_2,n} \quad b\theta n + \frac{1}{R}C_2 \qquad (10)$$
$$\text{s.t.} \quad n^{\eta} u\left(\frac{C_2}{n}\right) = W$$

Denote the solution to this problem as $C_2(W,\theta)$ and $n(W,\theta)$. $\hat{V}(W,\theta)$ is the cost of delivering utility W from having children to a parent who has type θ. In the first stage the planner solves the following problem

$$\min_{c_1(\theta),m(\theta),W(\theta)} \quad \sum_{\theta\in\Theta} \pi(\theta)\left(c_1(\theta) + \theta m(\theta) + \hat{V}(W(\theta),\theta)\right) \qquad (11)$$
$$\text{s.t.} \quad \sum_{\theta\in\Theta} \pi(\theta)\left(u(c_1(\theta)) + h\left(m(\theta)\right) + \beta W(\theta)\right) \geq W_0$$

Denote the solution to this problem as $c_1(W_0,\theta), m(W_0,\theta)$ and $W(W_0,\theta)$. Given cost function $\hat{V}(W,\theta)$ the first stage problem is standard, except for the fact that utility to the parents from having children, $W(W_0,\theta)$, explicitly depends on θ. The reason is that cost of having children is parents' time and therefore this cost is different for parents who have different productivity shocks. This makes the planner's cost function, $\hat{V}(W,\theta)$, explicitly depend on θ.[3]

Our result that consumption of each child, c_2, does not depend on parents' promised utility can be understood by looking at the second stage problem. First note that the way c_2 varies with promised utility depends on how C_2 (consumption of all children) and n (number of

[3]This is not the case when fertility is not a choice variable or cost of having children is only in terms of consumption good. In those environments parent's future promised utility (utility form having children) is equalized across different parents. Because cost of delivering those promised utilities are the same across parents who experience different shocks.

children) vary with promised utility. For example, in the case in which fertility is not a choice and number of children are constant, both C_2 and c_2 are Monotone increasing in W. We argue that in our formulation of altruism (which we borrow from Barro and Becker (1989) and Becker and Barro (1988)) the ratio of $\frac{C_2}{n}$ is independent of W. To see this note that objective of the planner in the second stage problem is homogeneous of degree one and constraint set is homogeneous of degree η. It can be easily seen that the solution to this problem is homogeneous of degree $\frac{1}{\eta}$ in W. In other words, the solution has the form $C_2(W,\theta) = \tilde{C}_2(\theta)W^{1/\eta}$ and $n(W,\theta) = \tilde{n}(\theta)W^{1/\eta}$. Therefore, the ratio $\frac{C_2(W,\theta)}{n(W,\theta)}$ does not depend on W and only depends on technology and preference parameters. Since this feature of the solution will be a recurrent theme throughout the paper it is useful to give it a name. We will call this the 'resetting' property – i.e., per capita utility is reset to a level that is independent of state variables.

There are four qualitative properties of the solution to this problem that will have close analogies in the results that will come later. We summarize them here as a Proposition for future reference.

Proposition 1 *Under Assumption 1:*

1. *$c_1(W_0, \theta)$ is independent of θ, and strictly increasing in W_0,*
2. *$l(W_0, \theta)$ is strictly increasing in θ, and strictly decreasing in W_0,*
3. *$n(W_0, \theta)$ is strictly decreasing in θ, and strictly increasing in W_0,*
4. *$c_2(W_0, \theta)$ is strictly increasing in θ, but independent of W_0.*

Proof. See Appendix A.1. ■

This Proposition shows one of the features of the model; when the cost of children is in terms of parental leisure, a utilitarian planner would allow for consumption inequality. As we know in the standard models with constant population, planners set marginal utilities equal. Here the same property holds. However, the 'correct' marginal utility is the per capita marginal – in period 2 it is $\beta n(\theta)^{\eta-1}u'(c_2(\theta))$. Therefore, since more productive households have fewer children, full risk sharing implies a degree of inequality in consumption among children. In this simple two period example, the inequality that results depends only the parent's income rather than that of the child. A high productivity parent will have a lower number of children and as a result of risk-sharing, her children will have a higher per capita consumption.

Before we proceed further, it is helpful to present an asymptotic property of the model when

W_0 become very small and contrast it to a model with no fertility choice. For the purpose of this exercise assume that both $u(\cdot)$ and $h(\cdot)$ are unbounded below. Then it is straight forward to show that $c_1(W_{0,}\theta), m(W_{0,}\theta) \to 0$ and $W(W_{0,}\theta) \to -\infty$ as $W_0 \to -\infty$. Now consider the second stage problem (problem (10)). If fertility is not a choice (i.e., it is constant), then we have $\lim_{W\to-\infty} C_2(W,\theta) = 0$ and $\lim_{W\to-\infty} u\left(\frac{C_2(W,\theta)}{n(W,\theta)}\right) = -\infty$. This follows directly from constraint (11).

However, when fertility is a choice variable, following the discussion above we know that $C_2(W,\theta)$ and $n(W,\theta)$ are homogeneous of degree $\frac{1}{\eta}$. Therefore,

$$\lim_{W\to-\infty} C_2(W,\theta) = \lim_{W\to-\infty} n(W,\theta) = 0$$

But we also know that $\frac{C_2(W,\theta)}{n(W,\theta)}$ is independent of W and therefore $\lim_{W\to-\infty} u\left(\frac{C_2(W,\theta)}{n(W,\theta)}\right) \neq -\infty$.

Immiseration corresponds to the situation where $\frac{C_2}{n} \to 0$ as $W \to -\infty$. The homothetic formulation of utility usually used in Barro/Becker models of fertility choice gives an extreme version where this does not occur – $\frac{C_2}{n}$ is independent of W. Thus, although the no immiseration result is particularly stark in the Barro/Becker case, we suspect that it holds in much more generality.

2.1 Private information

In this section we use our simple example to show that when productivity of the parent is private information a version of the resetting property still holds. More precisely, we will show that the consumption of each child of a parent which has the high productivity shock is independent of promised utility to parents (W_0).

Consider again the original planner's problem (9) and assume that productivity shock θ is private information. Then, the appropriate planner's problem must also include incentive compatibility constraints. In what follows we make the assumption that only downwards incentive constraints are binding – a parent with shock θ_H have incentive to pretend to be of a type θ_L with.[4] The new planner's problem is:

$$\min_{c_1(\theta),n(\theta),l(\theta),c_2(\theta)} \quad \sum_{\theta\in\Theta=\{\theta_H,\theta_L\}} \pi(\theta)\left(c_1(\theta)+\frac{1}{R}n(\theta)c_2(\theta)\right)-\sum_{\theta\in\Theta}\pi(\theta)\theta l(\theta) \tag{12}$$

$$\text{s.t.} \quad \sum_{\theta\in\Theta=\{\theta_H,\theta_L\}} \left[\pi(\theta)\left(u(c_1(\theta))+h\left(1-l(\theta)-bn(\theta)\right)+\beta n(\theta)^\eta u(c_2(\theta))\right)\right]\geq W_0$$

$$u(c_1(\theta_H))+h\left(1-l(\theta_H)-bn(\theta_H)\right)+\beta n(\theta_H)^\eta u(c_2(\theta_H))\geq$$
$$u(c_1(\theta_L))+h\left(1-\frac{\theta_L}{\theta_H}l(\theta_L)-bn(\theta_L)\right)+\beta n(\theta_L)^\eta u(c_2(\theta_L))$$

Note that in this problem no one pretends to be the highest type θ_I. Therefore, the allocations of this type are undistorted. To see this suppose λ and μ are the Lagrange multipliers on the promise keeping constraint and incentive compatibility constraint respectively. Then, the first order conditions are:

$$\begin{aligned}
&\pi(\theta_H)=(\lambda\pi(\theta_H)+\mu)u'(c_1(W_0,\theta_H))\\
&\pi(\theta_H)\theta_H=(\lambda\pi(\theta_H)+\mu)h'(1-l(W_0,\theta_H)-bn(W_0,\theta_H))\\
&\frac{1}{R}\pi(\theta_H)n(W_0,\theta_H)=(\lambda\pi(\theta_H)+\mu)\beta n(W_0,\theta_H)^\eta u'(c_2(W_0,\theta_H))\\
&\frac{1}{R}\pi(\theta_H)c_2(W_0,\theta_H)=(\lambda\pi(\theta_H)+\mu)\left[-bh'\left(1-l(W_0,\theta_H)-bn(W_0,\theta_H)\right)\right.\\
&\qquad\qquad\left.+\beta\eta n(W_0,\theta_H)^{\eta-1}u(c_2(W_0,\theta_H))\right]
\end{aligned}$$

Combining these equations we get:

$$\begin{aligned}
&\theta_H u'(c_1(W_0,\theta_H))='\left(1-l(W_0,\theta_H)-bn(W_0,\theta_H)\right)\\
&\beta Rn(W_0,\theta_H)^{\eta-1}u'(c_2(W_0,\theta_H))=u'(c_1(W_0,\theta_H))\\
&\beta\eta n(W_0,\theta_H)^{\eta-1}u(c_2(W_0,\theta_H))=bh'\left(1-l(W_0,\theta_H)-bn(W_0,\theta_H)\right)\\
&\qquad\qquad+\beta n(W_0,\theta_H)^{\eta-1}u'(c_2(W_0,\theta_H))c_2(W_0,\theta_H)
\end{aligned}$$

Note that these equations are identical to equations (2)-(5). By combining them we get the following equation

$$\eta u(c_2(W_0,\theta_H))=u'(c_2(W_0,\theta_H))c_2(W_0,\theta_H)+b\theta_H Ru'(c_2(W_0,\theta_H))$$

[4] We can show that at the full information efficient allocations the downward constraints are binding and upward constraints are slack. We also verify the slackness of upward constraints in out numerical example (in infinite horizon environment). In general we cannot prove that only downward constraints are binding because the preferences do not exhibit single-crossing property.

which is equation (8) evaluated at $\theta = \theta_H$. Therefore, as above, $c_2(W_0, \theta_H)$ is independent of W_0. Moreover, the value of $c_2(W_0, \theta_H)$ is the same under full information and private information (nothing in above equation depends on the incentive constraint).

There are two key features in the model that derive this result. First, is the homotheticity property emphasized above. The second is the fact that upward incentive constraints are not binding. Because of this, the allocation of the highest productivity type is undistorted.

Note that it is immediate that the asymptotic properties that we discussed for the full information allocation in section 2 holds for the private information allocation of the highest type.

3 The Benchmark Model – Full Information

In this section, we lay out the benchmark model with no private information and discuss its properties. Time is discrete from 0 to ∞. At date 0, the economy is populated by N_{-1} agents. Each agent lives for one period. Agents, when alive, draw labor productivity shocks, can consume and have children. The cost of raising a child is in terms of time where we assume that it takes b units of time to raise each child.

Here, the only source of uncertainty is the idiosyncratic risk of the productivity shock, θ_t that an agent in period t receives. We assume that θ_t is an i.i.d. stochastic process and takes on values in the set Θ and has distribution $\pi(\theta)$. There is also a production function $F(K, L) = RK + L$ as in the example above where K is aggregate capital and L is the aggregate effective labor hours. The initial level of capital is given by K_0.

Define $N_t(\theta^{t-1})$ as the current population of a cohort whose ancestors received history of shocks of θ^{t-1}. We can define a *feasible* allocation $\{c_t(\theta^t), l_t(\theta^t), n_t(\theta^t), K_t\}_{t=0}^{\infty}$ as:

$$\sum_{\theta^t} \pi(\theta^t) N(\theta^{t-1}) c_t(\theta^t) + K_{t+1} \leq \sum_{\theta^t} \pi(\theta^t) N(\theta^{t-1}) \theta_t l(\theta^t) + RK_t$$

where

$$N_t(\theta^{t-1}) = N_{t-1}(\theta^{t-2}) n_{t-1}(\theta^{t-1}) \quad , \quad N_0(\theta^{-1}) = N_{-1}$$

Since the production function is linear, we can suppress capital and write feasibility as the following constraint:

$$\sum_{t=0}^{\infty}\frac{1}{R^t}\sum_{\theta^t\in\Theta^{t+1}}\pi(\theta^t)N_{-1}\left[\prod_{s=0}^{t-1}n_s(\theta^s)\right]\left(c_t(\theta^t)-\theta_t l_t(\theta^t)\right)\leq K_0 \tag{13}$$

The utility of an agent in period t is the following:

$$U_t = u(c_t) + h(1-l_t-bn_t) + \beta n_t^{\eta} U_{t+1} \tag{14}$$

Therefore a planner that maximizes the expected utility of the generation in period-0 faces the following problem:

$$\begin{aligned}\max\quad & \sum_{t=0}^{\infty}\beta^t\sum_{\theta^t}\pi(\theta^t)\left[\prod_{s=0}^{t-1}n_s(\theta^s)\right]^{\eta}\left(u(c_t(\theta^t))+h(1-l_t(\theta)-bn_t(\theta^t))\right) && \text{(P1)}\\ \text{s.t.}\quad & (13),\end{aligned}$$

where the objective in (P1) is derived by iterating on (14)[5].

As it is mentioned in Alvarez (1999), the above problem by itself is not concave in n_t, c_t. This is due to the fact that we have the product of n_t and c_t in the budget constraint and as a result budget set is not convex. Alvarez (1999) proposes a novel way of resolving this issue by defining new variables as product of per capita allocations with population of the associated cohort. We will use his approach to make the problem concave. Hence, we define the following variables:

$$\begin{aligned}C_t(\theta^t) &= c_t(\theta^t)N_t(\theta^{t-1})\\ L_t(\theta^t) &= l_t(\theta^t)N_t(\theta^{t-1})\end{aligned}$$

The problem in terms of the new variables is the following:[6]

$$\begin{aligned}\max\quad & \sum_{t=0}^{\infty}\beta^t\sum_{\theta^t}\pi(\theta^t)N_t^{\eta}(\theta^{t-1})\left(u\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right)+h\left(1-\frac{L_t(\theta^t)}{N_t(\theta^{t-1})}-b\frac{N_{t+1}(\theta^t)}{N_t(\theta^{t-1})}\right)\right) && \text{(P1')}\\ \text{s.t.}\quad & \sum_{t=0}^{\infty}\frac{1}{R^t}\sum_{\theta^t\in\Theta^{t+1}}\pi(\theta^t)\left(C_t(\theta^t)-\theta L_t(\theta^t)\right)\leq K_0\\ & N_0(\theta^{-1}) = N_{-1} : \text{given}.\end{aligned}$$

[5]We need to define $\prod_{s=0}^{-1} n_s^{\eta}(\theta^s) = 1$.

[6]Notice that the product term in the objective function as well as in 13 is by definition $N_t(\theta^{t-1})$.

We make the following assumptions to ensure that the above problem is concave.

Assumption 2 *1. The function* $N^\eta u\left(\frac{C}{N}\right)$ *is strictly concave and strictly increasing in* (C, N).

2. The function $N^\eta h\left(\frac{M}{N}\right)$ *is strictly concave and strictly increasing in* (M, N).

3. $\eta < 1$.

From the strict concavity of (P1') it follows that there is a unique solution.

It will be useful for what follows to make one further transformation of this problem. Since the budget set in (P1') is convex and the objective is concave by assumption 2, we can follow the literature and transform this problem into a recursive cost minimization problem in which the extra state is continuation utility, W. See the Appendix for details. Thus, the solution to above problem is the same as the solution to the following recursive problem:

$$\begin{aligned} V(N,W) = \min \quad & \sum_{\theta\in\Theta} \pi(\theta)\left(C(\theta) - \theta L(\theta) + \frac{1}{R}V(N'(\theta), W'(\theta))\right) \qquad (P3) \\ \text{s.t.} \quad & \sum_{\theta\in\Theta} \pi(\theta)\left(N^\eta\left(u\left(\frac{C(\theta)}{N}\right) + h\left(1 - \frac{L(\theta)}{N} - b\frac{N'(\theta)}{N}\right)\right) + \beta W'(\theta)\right) \geq W \end{aligned}$$

Although, technically, we have N as one of the state variables for our problem, it can be shown that (because of homotheticity properties) we can rewrite the problem in terms of per capita variables and eliminate N as a state variable. To do so, we define $w = W/N^\eta$ and $\tilde{V}(N,w) = V(N, N^\eta w)/N$. Using these definitions (P3) becomes:[7]

$$\begin{aligned} N\tilde{V}(N, N^\eta w) = \min \quad & \sum_{\theta\in\Theta} \pi(\theta) N\left(c(\theta) - \theta l(\theta) + \frac{1}{R}\frac{N'(\theta)}{N}\tilde{V}(N'(\theta), N'(\theta)^\eta w'(\theta))\right) \\ \text{s.t.} \quad & \sum_{\theta\in\Theta} \pi(\theta)\left(N^\eta\left(u\left(c(\theta)\right) + h\left(1 - l(\theta) - bn(\theta)\right)\right) + \beta N'(\theta)^\eta w'(\theta)\right) \geq N^\eta U \end{aligned}$$

where we have used $W'(\theta) = N'(\theta)^\eta w'(\theta)$. Define $n(\theta) = \frac{N'(\theta)}{N}$. Using this definition, we can eliminate N from the objective function as well as the constraint. It is then obvious that $\tilde{V}$ is independent of N and $\tilde{V}(N,w) = v(w)$. Hence, $v(w)$ satisfies the following functional

[7] Small letters represent per cohort member variables.

equation:

$$v(w) = \min_{c(\theta),l(\theta),n(\theta),w'(\theta)} \sum_{\theta\in\Theta} \pi(\theta) \left(c(\theta) - \theta l(\theta) + \frac{1}{R} n(\theta) v(w'(\theta)) \right) \qquad (P4)$$
$$\text{s.t.} \quad \sum_{\theta\in\Theta} \pi(\theta) \left(u(c(\theta)) + h(1 - l(\theta) - bn(\theta)) + \beta n(\theta)^\eta w'(\theta) \right) \geq w$$

3.1 Properties of the Optimal Allocation

In this section, we discuss the properties of the optimal allocation from problem (P3).

First we have a standard result from dynamic programming:

Theorem 1 *Under assumption 2, $V(N, W)$ is strictly convex. Moreover if $u(\cdot)$ and $h(\cdot)$ are C^1, then $V(\cdot, \cdot)$ is differentiable.*

Proof. See theorem 4.8 and 4.11 in **?**. ■

To do any further characterization of the solution to (P3), we will want to use the FOC's from this problem. This requires that the solution be interior however. The usual approach to guarantee interiority is to use Inada conditions. We use a version of these here to guarantee that c, $1 - l - bn$, and n are interior. The version that we use is stronger than usual and necessary. We do this because we will need this stronger version below when we consider an environment with private information.

Assumption 3 *Assume that both u and h are bounded above by 0, and unbounded below. Note that this implies that $\eta < 0$ is required for the overall concavity of utility and hence an Inada condition on n is automatically satisfied.*

Under this assumption, it follows that consumption, leisure and fertility are all strictly positive. This is not enough to guarantee that the solution is interior however, since hours worked might be zero. Indeed, there is no way to guarantee that $l > 0$ in this model. This is because of the way hours spent raising children enter the problem. Because of this feature of the model, it might be true that the marginal value of leisure exceeds the marginal product of an hour of work even when $l = 0$. The usual way of handling this problem by assuming that $h'(1) = 0$ will not work in this case since we know that $n > 0$. Hence, the marginal value of leisure at zero work will always be positive, even if $h'(1) = 0$. Because of this, when continuation utility is sufficiently high, it is always optimal for work to be zero.

In addition to this, in some cases, there are types that never work. This will be true when it is more efficient for a type to produce goods through the indirect method of having children and having their children work than through the direct method of working themselves. This will hold for a worker with productivity θ, if $\theta < \frac{E(\theta)}{bR}$. That is, $l(w,\theta) = 0$ for all w if $\theta < \frac{E(\theta)}{bR}$. For this reason, we will rule this situation out by making the following assumption:

Assumption 4 *Assume that, for all i, $\theta_i > \frac{E(\theta)}{bR}$.*

This assumption does not guarantee that $l(w,\theta) > 0$ for all w, but it can be shown that for low enough continuation utility $l > 0$. In what follows, we will simply assume that $l > 0$ for most of the paper. We will return to this issue below when we show that a stationary distribution exists.

Using the above derivation of (P4), we also have the following corollary. (See Appendix A.2 for the proof.):

Corollary 2 *Assume that $V(\cdot,\cdot)$ is twice differentiable. Then $v(w)$ is convex and strictly increasing and $\eta w v'(w) - v(w)$ is increasing.*

The optimality conditions of the problem are similar to those in section (2).

$$v'(w) = \frac{1}{u'(c(\theta))}, \qquad \forall \theta \in \Theta \tag{15}$$

$$\theta u'(c(\theta)) = h'(1 - l(\theta) - bn(\theta)) \text{ if } l > 0 \tag{16}$$

$$\beta R n(\theta)^{\eta-1} = u'(c(\theta)) v'(w'(\theta)) \tag{17}$$

$$\beta \eta n(\theta)^{\eta-1} w'(\theta) = \frac{1}{R} v(w'(\theta)) u'(c(\theta)) + bh'(1 - l(\theta) - bn(\theta)) \tag{18}$$

The interpretation of the above equations are as before. First, equation (15) comes from the Envelope condition and implies full intra-family risk-sharing. Second, equation (16) represents the trade-off between leisure and consumption. Equation (17) is the standard Euler equation and it represents the trade-off between consumption today and consumption tomorrow. Finally, equation (18) equates the marginal benefit of having an extra child - LHS - to the marginal cost of having an extra child - RHS - which consists of immediate utility cost in terms of leisure as well as the cost of having fewer per child resources available tomorrow in terms of utility today.
Replacing equations (16) and (17) in (18) gives us the following important equation:

$$\eta v'(w'(\theta)) w'(\theta) - v(w'(\theta)) = bR\theta \tag{19}$$

Equation (19) together with corollary 2 implies that $w'(\theta)$ is increasing in θ. Moreover, it implies that $w'(\theta)$ is independent of w. That is, there is no history dependence in continuation utility, when all the choices - labor supply, fertility, consumption - are undistorted. This will play an important role in the rest of the paper. The intuition for this result is similar to that given in the two period case – it follows from the homotheticity of the cost function, $V(N, W)$.

The properties of the optimal allocation for this problem are summarized in the following proposition - the claims hold only if $l(w, \theta)$ is interior:

Proposition 2 *Under assumption 2, the policy functions from (P4) satisfy:*

1. *$c(w, \theta)$ is independent of θ, and strictly increasing in w,*

2. *$l(w, \theta)$ is st. increasing in θ and strictly decreasing in w,*

3. *$n(w, \theta)$ is st. decreasing in θ and strictly increasing in w,*

4. *$w'(w, \theta)$ is strictly increasing in θ, but independent of w.*

Proof. See Appendix A.3. ■

Proposition 1 describes the properties of long-run inequality in this model. In fact, we know that starting from any ex-ante distribution of continuation utilities, if we apply the above policy functions, the next period distribution of continuation utilities is independent of the initial distribution. Moreover, the long-run distribution of w takes on values in the solutions of equation (19) and hence, is a simple transformation of the probability density of θ. From equation (15), the long run distribution for c can be directly obtained from that of w. As can be seen it is not a degenerate distribution as would usually be the case with separable utility between consumption and leisure. Even though this is true, it follows from equation (15) that c does not depend on the current shock (θ_t), but depends on the history of previous shocks (θ^{t-1}) as summarized in w_t. Moreover, because of the resetting property (the fact that $w'(w, \theta)$ is independent of w), it follows that w_t depends only θ_{t-1}. Hence, it follows that c_t depends only on θ_{t-1}. Thus, although our model shares a common feature with more standard models with non-separable preferences between consumption and leisure, this arises for different reasons. Here, it arises from the non-separability between family size and per child continuation utility.

4 Adding Private Information

In this section, we will extend the model in section 3. The model is exactly the same as before except that each agent's productivity is private information. The planner, can observe the output for each agent but not hours worked nor productivity. Using the revelation principle, we will only focus on direct mechanisms in which each agent is asked to reveal his true type. As is typical in problems like these, it can be shown that the full information optimal allocation does not satisfy incentive compatibility. Although the argument is more complex than in the usual case, we show (in Appendix A.5) that under the full information allocation, a higher productivity type would want to pretend to be a lower productivity type.

In addition to this, in Mirrleesian environments with private information where a single crossing property holds, one can show only downward incentive constraints bind. Here, however, the single crossing property does not hold due the lack of separability between fertility and leisure. Because of this, we do not currently have a proof that the only incentive constraints that ever bind are the downward ones. In keeping with what others have done (e.g., Phelan (1998) and Golosov and Tsyvinski (2007)), we assume that agents can only report a level of productivity that is less than or equal to their true type.[8] In the appendix, we give a sufficient condition for this to be true. Under this assumption, we can restrict reporting strategies, σ, to satisfy $\sigma_t(\theta^t) \leq \theta_t$. (Here, for every history θ^t, $\sigma_t(\theta^t)$ is agent's report of its productivity in period t and $\sigma^t(\theta^t)$ is the history of the reports.) Moreover, by the later restriction on reports, we have $\sigma_t(\theta^t) \leq \theta_t$. Call the set of restricted reports Σ. Then, an allocation is said to be *incentive compatible* if

$$\sum_{t,\theta^t} \beta^t \pi(\theta^t) N_t(\theta^{t-1})^\eta \left[u\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right) + h\left(1 - \frac{L_t(\theta^t)}{N_t(\theta^{t-1})} - b\frac{N_{t+1}(\theta^t)}{N_t(\theta^{t-1})}\right)\right] \geq \tag{20}$$

$$\sum_{t,\theta^t} \beta^t \pi(\theta^t) N_t(\sigma^{t-1}(\theta^{t-1}))^\eta \left[u\left(\frac{C_t(\sigma^t(\theta^t))}{N_t(\sigma_{t-1}(\theta^{t-1}))}\right) + h\left(1 - \frac{\sigma_t(\theta^t) L_t(\sigma^t(\theta^t))}{\theta_t N_t(\sigma^{t-1}(\theta^{t-1}))} - b\frac{N_{t+1}(\sigma^t(\theta^t))}{N_t(\sigma^{t-1}(\theta^{t-1}))}\right)\right]$$

$$\forall \ \sigma \in \Sigma$$

Hence, the planning problem becomes the following:

$$\max \quad \sum_{t,\theta^t} \beta^t \pi(\theta^t) N_t(\theta^{t-1})^\eta \left[u\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right) + h\left(1 - \frac{L_t(\theta^t)}{N_t(\theta^{t-1})} - b\frac{N_{t+1}(\theta^t)}{N_t(\theta^{t-1})}\right)\right]$$

$$\text{s.t} \quad \sum_{t,\theta^t}^{\infty} \frac{1}{R^t} \pi(\theta^t) \left[C_t(\theta^t) - \theta_t L_t(\theta^t)\right] \leq K_0$$

$$\sum_{t,\theta^t} \beta^t \pi(\theta^t) N_t(\theta^{t-1})^\eta \left[u\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right) + h\left(1 - \frac{L_t(\theta^t)}{N_t(\theta^{t-1})} - b\frac{N_{t+1}(\theta^t)}{N_t(\theta^{t-1})}\right)\right] \geq$$

$$\sum_{t=0}^{\infty} \beta^t \pi(\theta^t) N_t(\sigma_{t-1}(\theta^{t-1}))^\eta \left[u(\frac{C_t(\sigma_t(\theta^t))}{N_t(\sigma_{t-1}(\theta^{t-1}))}) + h(1 - \frac{Y_t(\sigma_t(\theta^t))}{\theta_t N_t(\sigma_{t-1}(\theta^{t-1}))} - b\frac{N_{t+1}(\sigma_t(\theta^t))}{N_t(\sigma_{t-1}(\theta^{t-1}))})\right]$$

$$\forall \, \sigma \in \Sigma$$

[8] In numerically calculated examples, this assumption is redundant.

Using the same techniques as in section 3, we can show that the above problem is equivalent to the following functional equation:

$$V(N,W) = \min_{C(\theta),L(\theta),N'(\theta)} \sum_{\theta\in\Theta} \pi(\theta)\left[C(\theta) - \theta L(\theta) + \frac{1}{R}V(N'(\theta),W'(\theta))\right] \tag{P5}$$

$$\text{s.t} \quad \sum_{\theta\in\Theta} \pi(\theta)\left[N^\eta\left(u\left(\frac{C(\theta)}{N}\right) + h\left(1 - \frac{L(\theta)}{N} - b\frac{N'(\theta)}{N}\right)\right) + \beta W'(\theta)\right] \geq W$$

$$N^\eta\left(u\left(\frac{C(\theta)}{N}\right) + h\left(1 - \frac{L(\theta)}{N} - b\frac{N'(\theta)}{N}\right)\right) + \beta W'(\theta) \geq$$

$$N^\eta\left(u\left(\frac{C(\hat{\theta})}{N}\right) + h\left(1 - \frac{\hat{\theta}L(\hat{\theta})}{\theta N} - b\frac{N'(\hat{\theta})}{N}\right)\right) + \beta W'(\hat{\theta}) \tag{21}$$

$$\forall \theta > \hat{\theta}.$$

As we can see, the problem is homogeneous in N and therefore as before, if we define $v(N,w) = \frac{V(N,N^\eta w)}{N}$, $v(\cdot,\cdot)$ will not depend on N and satisfies the following functional equation:

$$v(w) = \min_{c(\theta),l(\theta),n(\theta)} \sum_{\theta\in\Theta} \pi(\theta)\left(c(\theta) - \theta l(\theta) + \frac{1}{R}n(\theta)v(w'(\theta))\right) \tag{P5'}$$

$$\text{s.t.} \quad \sum_{\theta\in\Theta} \pi(\theta)\left(u(c(\theta)) + h(1 - l(\theta) - bn(\theta)) + \beta n(\theta)^\eta w'(\theta)\right) \geq w$$

$$u(c(\theta)) + h(1 - l(\theta) - bn(\theta)) + \beta n(\theta)^\eta w'(\theta) \geq$$

$$u(c(\hat{\theta})) + h(1 - \frac{\hat{\theta}l(\hat{\theta})}{\theta} - bn(\hat{\theta})) + \beta n(\hat{\theta})^\eta w'(\hat{\theta}),\ \forall\ \theta > \hat{\theta}$$

To ensure that the minimization in problem (P5) has a unique solution, we need the function $V(N,W)$ to be convex. A sufficient condition for convexity of the value function is that the constraint set correspondence should be convex with respect to the state variables. In this problem, due to perfect substitutability of fertility and labor supply, our constraint set correspondence is not convex. Therefore, we cannot show that the value function is convex using the standard methods. One way to resolve this issue is by allowing for randomization. Allowing for randomization, makes all the constraints linear in the probability distributions and therefore the constraint correspondence would be convex. This is the method used in Phelan and Townsend (1991) and Doepke and Townsend (2006). However, optimal plan may not necessarily involve lotteries since convexity of constraint correspondence is only sufficient for convexity of the value function.[9]

[9] In our numerical example the value function is convex even without the use of lotteries. In the appendix B, we characterize the case where utility from leisure is linear and there we can show that the constraint

Therefore, we will make the following assumption:

Assumption 5 *$V(N, W)$ is convex with respect to (N, W).*

4.1 Inverse Euler Equation

An important feature of dynamic Mirrleesian models with private information is the Inverse Euler Equation. Golosov et al. (2003) extend the original result of Rogerson (1985) and show that in a dynamic Mirrleesian model with private information, when utility is separable in consumption and leisure, the Inverse Euler equation holds when processes for productivity come from a general class. Here we will show that a version of the Inverse Euler equation holds. To do so, consider problem (P5). Suppose the multiplier on promise keeping is λ and multiplier on (24) is $\mu(\theta, \hat{\theta})$. Then the first order condition with respect to $W'(\theta)$ is the following:

$$\pi(\theta)\frac{1}{R}V_W(N'(\theta), W'(\theta)) + \lambda\pi(\theta)\beta + \beta\sum_{\theta>\hat{\theta}}\mu(\theta, \hat{\theta}) - \beta\sum_{\theta<\hat{\theta}}\mu(\hat{\theta}, \theta) = 0.$$

Define $\mu(\theta, \hat{\theta}) = 0$, if $\hat{\theta} \geq \theta$. Summing the above equations over all θ's, we have

$$\frac{1}{R}\sum_{\theta}\pi(\theta)V_W(N'(\theta), W'(\theta)) + \beta\lambda + \beta\sum_{\theta}\sum_{\hat{\theta}}\mu(\theta, \hat{\theta}) - \beta\sum_{\theta}\sum_{\hat{\theta}}\mu(\hat{\theta}, \theta) = 0.$$

Moreover, from Envelope Condition:

$$V_W(N, W) = -\lambda.$$

Therefore, we have

$$\sum_{\theta}\pi(\theta)V_W(N'(\theta), W'(\theta)) = \beta R V_W(N, W).$$

Now consider the first order condition with respect to $C(\theta)$:

$$\pi(\theta) + \lambda\pi(\beta)N^{\eta-1}u'\left(\frac{C(\theta)}{N}\right) + N^{\eta-1}u'\left(\frac{C(\theta)}{N}\right)\sum_{\hat{\theta}}\mu(\theta, \hat{\theta}) - N^{\eta-1}u'\left(\frac{C(\theta)}{N}\right)\sum_{\hat{\theta}}\mu(\hat{\theta}, \theta) = 0.$$

Thus,

$$V_W(N'(\theta), W'(\theta)) = \frac{\beta R}{N^{\eta-1}u'\left(\frac{C(\theta)}{N}\right)}.$$

correspondence is convex.

Therefore, we know that

$$V_W(N_{t+1}(\theta^t), W_t'(\theta^t)) = \frac{\beta R}{N_t(\theta^{t-1})^{\eta-1} u'\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right)}.$$

Hence, we can derive the Inverse Euler Equation:

$$E\left[\frac{1}{N_{t+1}(\theta^t)^{\eta-1} u'\left(\frac{C_{t+1}(\theta^{t+1})}{N_{t+1}(\theta^t)}\right)}|\theta^t\right] = \frac{\beta R}{N_t(\theta^{t-1})^{\eta-1} u'\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right)}. \tag{22}$$

An intuition for this equation is worth mentioning. Consider decreasing per capita consumption of an agent with history θ^t and saving that unit. There will be R units available the next day that can be distributed among the descendants. We increase consumption of agents of type θ^{t+1} by $\epsilon(\theta_{t+1})$ such that:

$$\begin{aligned} n_t(\theta^t) \sum_\theta \pi(\theta)\epsilon(\theta) = R \\ u'(c_{t+1}(\theta^t, \theta))\epsilon(\theta) = u'(c_{t+1}(\theta^t, \theta'))\epsilon(\theta') = \Delta. \end{aligned}$$

The first is the resource constraint implied by redistributing the available resources. The second one makes sure that the incentives are aligned. In fact it implies that the change in the utility of all types are the same and there is no incentive to lie. The above equations imply that

$$n_t(\theta^t) \sum_{\theta_{t+1}} \pi(\theta_{t+1}) \frac{\Delta}{u'(c_{t+1}(\theta^t, \theta_{t+1}))} = R$$

Since the change in utility from this perturbation must be zero, we must have $\beta\Delta = u'(c_t(\theta^t))$. Replacing in the above equation leads to equation (22). We summarize this as a Proposition:

Proposition 3 *If the optimal allocation is interior, the solution satisfies a version of the Inverse Euler Equation:*

$$E\left[\frac{1}{N_{t+1}(\theta^t)^{\eta-1} u'\left(\frac{C_{t+1}(\theta^{t+1})}{N_{t+1}(\theta^t)}\right)}|\theta^t\right] = \frac{\beta R}{N_t(\theta^{t-1})^{\eta-1} u'\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right)}.$$

4.2 The Resetting Property and Long Run Properties of Income and Fertility

We have shown in section 3 that per capita promised value is history independent. Here we will show that a similar property holds when we add private information to the model, but only for types for which the presence of private information does not introduce any distortions – i.e., for any type such that no other type is tempted to pretend to be. In this case, the resetting property is satisfied. This can be derived from the first order conditions of the the recursive formulation. Suppose that $\Theta = \{\theta_1 < \cdots < \theta_n\}$. Taking first order condition with respect to $N'(\theta_n)$ and $L(\theta)$, respectively, gives us the following equations:

$$\pi(\theta_n)\frac{1}{R}V_N(N'(\theta_n), W'(\theta_n)) = b\left(\lambda\pi(\theta_n) + \sum_{\hat{\theta}<\theta_n}\mu(\theta_n, \hat{\theta})\right)N^{\eta-1}h'\left(1 - \frac{L(\theta_n)}{N} - \frac{N'(\theta_n)}{N}\right)$$

$$\pi(\theta_n)\theta_n + \left(\lambda\pi(\theta_n) + \sum_{\hat{\theta}<\theta_n}\mu(\theta_n, \hat{\theta})\right)N^{\eta-1}h'\left(1 - \frac{L(\theta_n)}{N} - \frac{N'(\theta_n)}{N}\right) = 0$$

Therefore,

$$V_N(N'(\theta_n), W'(\theta_n)) = -bR\theta_n$$

Using the fact that $V(N, W) = Nv(N^{-\eta}W)$, we will have

$$v(w'(\theta)) - \eta w'(\theta)v'(w'(\theta)) = -bR\theta_n.$$

We can see that $w'(w, \theta_n)$ is independent of promised continuation utility. That is, $w'(w, \theta_n) = w'(\hat{w}, \theta_n)$ for all w, $\hat{w}$. Denote by w_0 this level of promised continuation utility – $w_0 = w'(w, \theta_n)$.

The resetting property means that once a parent receives a high productivity shock, the per capita allocation for her descendants is independent of the parents level of wealth – an extreme version of social mobility holds.

Because of this, it follows that there is no immiseration in this model, under very mild assumptions, in the sense that per capita utility does not converge to its lower bound. To see this, first consider the situation if $n(w, \theta)$ is independent of (w, θ). In this case, from any initial position, the fraction of the population that will be assigned to w_0 next period is at least $\pi(\theta_n)$. This by itself implies that there is not a.s. convergence to the lower bound of continuation utilities. When $n(w, \theta)$ is not constant, the argument involves more steps. Assume that n is bounded above and below – $0 < a \leq n(w, \theta) \leq a'$. Then, the fraction of descendants being assigned to w_0 next period is at least $\frac{\pi(\theta_n)a}{(1-\pi(\theta_n))a'}$. Again then, we see that there will not be a.s. immiseration.

We summarize this discussion in a Proposition.

Proposition 4 *Continuation utility has a 'resetting' property, $w'(w, \theta_n) = w_0$ for all w. Assume that $n(w, \theta)$ is bounded above and below – $0 < a \leq n(w, \theta) \leq a'$ – then, if $\pi(\theta_n) > 0$, it follows that continuation utilities do not converge a.s. to the lower bound – no immiseration occurs.*

Intuitively, the resetting property here mirrors the argument given above in the full information case (Proposition 1, part 4). That is, since no 'type' wants to pretend to have $\theta = \theta_n$, the first order conditions for the allocation when $\theta = \theta_n$ are the same with or without private information. Since we saw in the full information case that w' is independent of w, we have the same property holding here.

Next, we turn to a discussion of the limiting properties of fertility and population.

We have two tools at our disposal for this, our version of the Inverse Euler Equation and the proposition above.

One version of the Inverse Euler Equation can be written as:

$$\sum_{\theta} \pi(\theta) V_W(N'(\theta), W'(\theta)) = \beta R V_W(N, W).$$

Using the homotheticity properties of V, we can rewrite this as:

$$\sum_{\theta} \pi(\theta) \left(N'(\theta)\right)^{1-\eta} v'(w'(\theta)) = \beta R N^{1-\eta} v'(w).$$

If $\beta R = 1$, we see that $X_t = N_t^{1-\eta} v'(w_t)$ is a non-negative martingale. Thus, the martingale convergence theorem implies that there exists a non-negative random variable with finite mean, X_∞, such that $X_t \to X_\infty$ *a.s.*

As is standard in this literature, to provide incentives for truthful revelation of types, we must have 'spreading' in $\left(N'(\theta)\right)^{1-\eta} v'(w'(\theta))$ (details in the Appendix). Thus, it follows that $X_\infty = 0$ *a.s.* But, as we have seen above, it is not possible for $v'(w_t)$ to converge to zero since it is equal to $v'(w_0)$ at least $\frac{\pi(\theta_n)a}{(1-\pi(\theta_n))a'}$ percent of the time. Thus, it follows that $\left(N'(\theta^t)\right)^{1-\eta} \to 0$ *a.s.*

Intuitively, the planner is relying heavily on overall dynasty size to provide incentives and less on continuation utilities. This is something that sets this model apart from the more standard approach with exogenous fertility.

Finally, the fact that $\left(N'(\theta^t)\right)^{1-\eta} \to 0$ *a.s.* does not mean that fertility converges to zero almost surely, rather, it means that it is less than replacement (i.e., $n < 1$). This is most easily seen in the context of a discussion of the stationary distribution which we take up in the next section.

Summarizing:

Proposition 5 *Assume that* $\beta R = 1$*, then,* $(N'(\theta^t)) \to 0$ *a.s.*

5 Stationary Distributions

In this section we discuss the existence of stationary distributions for the endogenous variables of the model. There are two issues here. First, is there a stationary distribution for continuation utilities and is it non-trivial (i.e., can we rule out immiseration)? Second, because the size of population is endogenous here and could be growing (or shrinking), we must also show that the growth rate of population is also stationary. We deal with this problem in general here, and then study a special case with only two shocks in detail below.

The policy functions for fertility and future continuation utilities in problem (P5') above are $n(\theta, w), w'(\theta, w)$. Consider a measure of continuation utilities over $\mathbb{R}$, Ψ. Then, applying the policy functions to the measure Ψ, gives rise to a new measure over continuation utilities, $T\Psi$:

$$T(\Psi)(A) = \int_w \sum_\theta \pi(\theta)\mathbf{1}_{\{(\theta,w);w'(\theta,w)\in A\}}(w,\theta)n(\theta,w)d\Psi(w) \tag{23}$$
$$\forall A : \text{Borel Set in } \mathbb{R}$$

For a given measure of promised value today, Ψ, $T(\Psi)(A)$ is the measure of agents with continuation utility in the set A tomorrow. The overall population growth generated by Ψ is given by

$$\gamma_\Psi = \frac{\int_{\mathbb{R}} \sum_\theta \pi(\theta)n(\theta,w)d\Psi(w)}{\Psi(\mathbb{R})} = \frac{T(\psi)(\mathbb{R})}{\Psi(\mathbb{R})}$$

Now, suppose Ψ is a *probability* measure over continuation utilities. Ψ is said to be a *stationary distribution* if:

$$T(\Psi) = \gamma_\Psi \Psi$$

This is equivalent to having a constant distribution of per capita continuation utility along a Balanced Growth Path in which population grows at rate γ_Ψ.

To show that a stationary distribution exists, the first step will be to show that continuation utility can be confined to a compact set. The key step in this argument, and one that differentiates the endogenous fertility model from the usual case, is to show that continuation utilities are bounded below. That is, even as promised utility, w, gets lower and lower, continuation utility, $w'(w, \theta)$, is bounded away from $-\infty$.

To this end, we show that as $w \to -\infty$, the optimal allocation converges to $c = 0$, $l = 1$, $n = 0$. The interesting thing about this allocation is that no incentive constraints are binding and hence, the optimal allocation has properties similar to those in the full information case – there is resetting of continuation utility for all types when w is low enough. Formally:

Proposition 6 *Suppose that v is continuously differentiable*[10]*. Then there exists a $\underline{w}_i \in \mathbb{R}$, such that*

$$\lim_{w \to -\infty} w'(w, \theta_i) = \underline{w}_i$$

See Appendix A.7 for the proof.

The fact that, with utility unbounded below, the incentive problem is less severe for low values of promised utility also holds in models with exogenous fertility. Loosely speaking, as w gets smaller, the allocations look more and more similar to full information allocations, whether fertility is endogenous or exogenous. What makes an endogenous fertility model different from an exogenous one is the properties of full information allocations. We know that with endogenous fertility, in a frictionless model continuation utility is bounded below (by the shock-specific resetting values), while when fertility is exogenous, continuation utility gets arbitrarily small. This immediately implies that there is no per-capita immiseration with endogenous fertility.

From this, it follows immediately that continuation utility can be confined to a compact set.

Corollary 3 *Suppose that v is continuously differentiable. Then for all $\hat{w} < 0$, $w'(w, \theta)$ is bounded below on $(-\infty, \hat{w}]$.*

Since utility is unbounded below and η is negative, $n(w, \theta_i)$ must be positive. But, we need more than this for the existence of a stationary distribution. We need to show that the mapping $\Psi \to \frac{T(\Psi)}{\gamma_\Psi}$ is continuous. For this, we need that $n(w, \theta)$ is continuous in w and that γ_Ψ is bounded away from 0. Finally, we need continuation utilities to lie in a compact set (i.e., we need to bound w' away from 0 as well) so that the relevant function maps a compact set into itself. This can be shown to be true in certain special cases (see the Appendix), but we have not yet shown it in complete generality.

Thus, to proceed, we first make the following assumption:

[10] The assumption about differentiability of the value function is an ad-hoc one. It is perhaps worth mentioning that since v is weakly increasing, by a theorem from Lebesgue, the points of non-differentiability is of measure zero. So we know that v is almost everywhere differentiable. The classical proof of differentiability of the value function is due to Benveniste and Schienkman. Their proof requires convexity of the value function, which we could not show.

Assumption 6 *There exists some $\bar{w} < \sup_c u(c)$ such that for any $w \in (-\infty, \bar{w}]$, $w'(w,\theta)$ belongs to $(-\infty, \bar{w}]$ for all $\theta \in \Theta$.*

Now, we are ready to prove our main result about existence of stationary distribution. Let $\mathcal{M}([\underline{w}, \overline{w}])$ be the set of regular probability measures on $[\underline{w}, \bar{w}]$.

Theorem 4 *If the solution to the functional equation implies a unique policy function $n(w,\theta)$, for all (w,θ), then there exists a measure $\Psi^* \in \mathcal{M}([\underline{w}, \bar{w}])$ such that $T(\Psi^*) = E_{\Psi^*}(n) \cdot \Psi^*$.*

Proof. Since $[\underline{w}, \bar{w}]$ is compact in $\mathbb{R}$, by Riesz Representation Theorem (Dunford and Schwartz (1958), IV.6.3), the space of regular measures is isomorphic to the space $C^*([\underline{w}, \bar{w}])$, the dual of the space of bounded continuous functions over $[\underline{w}, \bar{w}]$. Moreover, by Banach-Alaoglu Theorem (Rudin (1991), Theorem 3.15), the set $\{\Psi \in C^*([\underline{w}, \bar{w}]); ||\Psi|| \leq k\}$ is a compact set in the weak-* topology for any $k > 0$. Equivalently the set of regular measures, Ψ, with $||\Psi|| \leq 1$, is compact. Since non-negativity and full measure on $[\underline{w}, \bar{w}]$ are closed restrictions, we must have that the set

$$\{\Psi : \Psi \text{ a regular measure on } [\underline{w}, \bar{w}], \Psi([\underline{w}, \bar{w}]) = 1, \Psi \geq 0\}$$

is compact in weak-* topology.

By definition,

$$T(\Psi)(A) = \int_{[\underline{w}, \bar{w}]} \sum_{i=1}^{n} \pi_i \mathbf{1}\left\{w'(w,\theta_i) \in A\right\} n(w,\theta_i) d\Psi(w).$$

The assumption that the policy function is unique implies that it is continuous by the Theorem of the Maximum. It also follows from this that n is bounded away from 0 on $[\underline{w}, \bar{w}]$ (since otherwise utility would be $-\infty$). From this, it follows that T is continuous in Ψ. Moreover,

$$E_{\Psi}(n) = \int_{[\underline{w}, \bar{w}]} \sum_{i=1}^{n} \pi_i n(w,\theta_i) d\Psi(w) \geq \underline{n} > 0.$$

is a continuous function of Ψ and is bounded away from zero.

Therefore, the function

$$\hat{T}(\Psi) = \frac{T(\Psi)}{E_{\Psi}(n)} : \mathcal{M}([\underline{w}, \bar{w}]) \to \mathcal{M}([\underline{w}, \bar{w}])$$

is continuous. Therefore, by Schauder-Tychonoff Theorem (Dunford and Schwartz (1958), V.10.5), $\hat{T}$ has a fixed point $\Psi^* \in \mathcal{M}([\underline{w}, \bar{w}])$. ∎

This theorem immediately implies that there is a stationary distribution for per capita consumption, labor supply and fertility. Moreover, since promised utility is fluctuating in a bounded set, per capita consumption has the same property. This is in contrast to the models with exogenous fertility where a shrinking fraction of the population will have an ever growing fraction of aggregate consumption.

The resetting property at the top, has also important implications about intergenerational social mobility. In fact, it makes sure that any smart parent will have children with a high level of wealth - as proxied by continuation utility. Finally, there is a lower bound on how much of this mobility occurs:

Remark 5 *Suppose that $\bar{w} = w_0$. Choose $A > 0$ so that $\frac{n(w,\theta_n)}{n(w,\theta)} \geq A$ for all w and θ. Suppose that $l(w, \theta_n) > 0$, for all $w \in [\underline{w}, w_0]$, then for any $\Psi \in \mathcal{M}([\underline{w}, \bar{w}])$, we have:*

$$\hat{T}(\Psi)(\{w_0\}) \geq \frac{\pi_n A}{1 - \pi_n + \pi_n A}.$$

See Appendix A.10 for proof.

5.1 Special Cases

In this section, we examine what happens when there are only two possible productivity shocks, $\theta_L < \theta_H$. We study two special cases given this assumption. In the first, we allow for a general utility function for leisure and give sufficient conditions for the stationary distribution to be unique. In the second, we assume that utility of leisure is linear and give a global stability result.

5.1.1 A Uniqueness Result

We begin by allowing for general utility functions for leisure and assume that resetting property at the top holds for every $w < w_0$. For this case, using a direct constructive proof, we show that the model implies a long-run distribution for per capita consumption and characterize its properties. What is convenient about this example is that we can show that if resetting holds at all w and assumption (6) with $\overline{w} = w_0$, the stationary distribution is unique.

We know that $w_0 = w'(w, \theta_H)$ is independent of w for all $w \leq w_0$. Hence, we can define the following set of promised values:

$$\mathcal{W} = \{w_n | w_{n+1} = w'(w_n, \theta_L), \forall n \geq 0\}$$

By Corollary 3, there is a lower bound $\underline{w}$ such that $\underline{w} \leq w \leq w_0$, for all $w \in \mathcal{W}$.

Assumption 7 *Assume that $w_j \neq w_i$ if $j \neq i$.*

Consider a distribution over $\mathcal{W}$, $\Psi = (\psi_0, \psi_1, \cdots)$ with $\sum_{i=0}^{\infty} \psi_i = 1$. For Ψ to be a stationary distribution, there must exist a γ such that the following conditions hold:

$$\gamma\psi_0 = \pi_H \sum_{i=0}^{\infty} n(w_i, \theta_H)\psi_i \tag{24}$$

$$\gamma\psi_j = \pi_L n(w_{j-1}, \theta_L)\psi_{j-1}, \quad j \geq 1 \tag{25}$$

Iterating on equation (25) implies the following:

$$\psi_m = \left(\frac{\pi_L}{\gamma}\right)^m n(w_{m-1}, \theta_L) n(w_{m-2}, \theta_L) \cdots n(w_0, \theta_L)\psi_0$$

Replacing in (24) implies the following equation:

$$\gamma = \pi_H \left(n(w_0, \theta_H) + \sum_{m=1}^{\infty} \left(\frac{\pi_L}{\gamma}\right)^m n(w_{m-1}, \theta_L) n(w_{m-2}, \theta_L) \cdots n(w_0, \theta_L) n(w_m, \theta_H) \right) \tag{26}$$

Given that in the original problem, we must have $n(w, \theta) \leq 1/b$. This means that the right hand side of the above equation is lower than $\sum_{m=0}^{\infty} (\pi_L/(b\gamma))^m$. Therefore, if we let $\gamma \to \infty$, the right hand side converges to a finite number, $\pi_H n(w_0, \theta_H)$. Notice that the left hand side is strictly increasing and the right hand side is strictly decreasing in γ. Moreover, at $\gamma = 0$ RHS is higher than LHS and at $\gamma = \infty$, RHS is lower than LHS. Because of this, if we knew that RHS was continuous, this would be sufficient to say that there is a γ satisfying equation (26) and that it is unique. To handle this last technical detail, we proceed as follows – Define γ_K as follows:

$$\gamma_K = \pi_H \left(n(w_0, \theta_H) + \sum_{m=1}^{K} \left(\frac{\pi_L}{\gamma_K}\right)^m n(w_{m-1}, \theta_L) n(w_{m-2}, \theta_L) \cdots n(w_0, \theta_L) n(w_m, \theta_H) \right).$$

By definition, $\gamma_K < \gamma_{K+1}$. We know that $n(w,\theta) < \frac{1}{b}$, therefore

$$\gamma_K < \pi_H \left(\frac{1}{b} + \sum_{m=1}^{K} \left(\frac{\pi_L}{\gamma_K} \right)^m \left(\frac{1}{b} \right)^{m+1} \right).$$

Suppose that $\frac{\pi_L}{b\gamma_K} < 1$ or $\frac{\pi_L}{b} < \gamma_K$. Then, the above inequality implies that

$$b\gamma_K < \pi_H \frac{1}{1 - \frac{\pi_L}{b\gamma_K}} \Rightarrow \gamma_K < \frac{\pi_H + \pi_L}{b} = \frac{1}{b}.$$

This shows that γ_K is a bounded increasing sequence. Hence, there exists γ^* such that $\gamma_K \to \gamma^*$ with $\gamma^* > \gamma_K$. It needs to be shown that at γ^*, RHS of (26) exists. Suppose not and that the sum is infinity. Define $F_K(\gamma)$ to be the RHS of (26) up to K-th term. $F_K(\gamma)$ is a continuous and decreasing function. Therefore, $\gamma_K = F_K(\gamma_K) > F_K(\gamma^*)$. Moreover, $F_K(\gamma^*) \to F(\gamma^*)$ and hence $F(\gamma^*) \leq \gamma^*$. This means that RHS of (26) cannot be infinity and (26) is satisfied at γ^*.

Now by Corollary 6 in the Appendix, we know that

$$\exists\, A > 0 \quad ; \qquad n(w,\theta_H) \geq A n(w,\theta_L) \quad \forall w \in [\underline{w}, \overline{w}]. \tag{27}$$

Therefore, using (26), we will have

$$\gamma = \pi_H \left(n(w_0,\theta_H) + \sum_{m=1}^{\infty} \left(\frac{\pi_L}{\gamma} \right)^m n(w_{m-1},\theta_L) n(w_{m-2},\theta_L) \cdots n(w_0,\theta_L) n(w_m,\theta_H) \right) \geq$$
$$\pi_H A \sum_{m=0}^{\infty} \left(\frac{\pi_L}{\gamma} \right)^m n(w_m,\theta_L) n(w_{m-1},\theta_L) \cdots n(w_0,\theta_L)$$
$$\Rightarrow \sum_{m=0}^{\infty} \left(\frac{\pi_L}{\gamma} \right)^{m+1} n(w_m,\theta_L) n(w_{m-1},\theta_L) \cdots n(w_0,\theta_L) \leq \frac{\pi_L}{A\pi_H}.$$

Now define, ψ_0 as

$$\psi_0 = \frac{1}{1 + \sum_{m=0}^{\infty} \left(\frac{\pi_L}{\gamma} \right)^{m+1} n(w_m,\theta_L) n(w_{m-1},\theta_L) \cdots n(w_0,\theta_L)}$$

By the above inequality, we know that ψ_0 exists and it is greater than zero. Moreover, we can automatically define ψ_i's using (25). Hence, the definition of γ, being the solution to (26) together with the definition of ψ_0, makes sure that Ψ satisfies (24)-(25) and hence, it is a stationary distribution. As it appears in the proof, in some sense, bounded relative fertility[11] together with the resetting property at the top are the key elements of having a long-run stationary distribution. First, every time any one receives a high shock, her promised value is reset. Secondly, relatively, there are enough children being born by high types so that we get stationarity. Moreover, the above proof shows that when the set of w's is restricted to $\mathcal{W}$, the stationary distribution is unique.

5.1.2 Linear Utility of Leisure

In this section we restrict attention to the special case where utility is linear in leisure – $h(m) = m$. Here, things simplify considerably. In the appendix, we show that both $n(w, \theta)$ and $w'(w, \theta)$ are independent of w. Thus, in the two shock case, there are only two relevant values for continuation utility, $w_0 = w'(w, \theta_H)$ and $w_1 = w'(w, \theta_L)$. Correspondingly, let $n_i = n(w, \theta_i)$.

Given this property for the policy function w', for any Ψ, $T(\Psi)$ has mass concentrated on the set $\mathcal{W} = \{w_0, w_1\}$. Because of this, for the purpose of characterizing the stationary distribution, we can summarize T by the two by two matrix:

$$\begin{bmatrix} \pi_H n_H & \pi_L n_L \\ \pi_H n_H & \pi_L n_L \end{bmatrix}.$$

As can be seen from this, the population growth rate is given by $\gamma^* = \pi_H n_H + \pi_L n_L$. Hence, $\hat{T}$ as defined above is given by:

$$\begin{bmatrix} \frac{1}{\gamma^*} & \pi_H n_H & \pi_L n_L \\ & \pi_H n_H & \pi_L n_L \end{bmatrix}.$$

Thus, for any initial distribution Ψ^0, $\hat{T}(\Psi^0) = \Psi^*$ where $\Psi^*(w_0) = \frac{\pi_H n_H}{\gamma^*}$ and $\Psi^*(w_1) = \frac{\pi_L n_L}{\gamma^*}$ and population grows at rate γ^*.

5.1.3 Stability

The example with linear utility of leisure is useful because it shows that in some cases, global stability can be guaranteed. It also shows a difficulty with showing this property in general. The first step to prove global stability in a Markov chain is to prove that there is a unique invariant distribution for a irreducible and acyclical Markov chain. Since our transition function $\hat{T}$ is not Markov, we cannot use the standard method of proving global stability. In fact, the stationary distribution might not be unique. Intuitively, there may be

[11] A high type's number of kids relative to the low type's

more than one population growth rate (γ^*) that is consistent with stationarity even though T is both acyclic and irreducible. If this is true, there may be two (or more) pairs, (Ψ_1^*, γ_1^*) and (Ψ_2^*, γ_2^*) that are stationary – $\hat{T}_1^* = \frac{T}{\gamma_1^*}$ and $\hat{T}_2^* = \frac{T}{\gamma_2^*}$ – where both $\hat{T}_1^*$ and $\hat{T}_2^*$ are acyclic and irriducible.[12] We suspect that some progress can be made on this problem in some cases. For example, in the two shock example given above, we showed (in section 5.1.1) that there is a unique stationary distribution, but we have not yet shown that it is globally stable. A key assumption in that argument is that there are no cycles in the sequence $\{w_n\}$. Indeed, if this assumption does not hold, the stationary distribution together with the population growth rate is not necessarily unique.

6 Implementation

In this section, we focus on implementing the efficient allocations described above through decentralized decision making with taxes. We break this discussion into two components. In the first, we specialize to a two period example so as to explicitly characterize how tax implementations are used to alter private fertility choices. In the second, we discuss the 'wedges' that appear in agents first order conditions – i.e., how do these differ from the full information efficient allocation.

6.1 A Two Period Example

To highlight that feature of the model that is new here – fertility choice – we restrict attention to a two period example. We assume that there is a one time shock, realized in the first period. For simplicity, we will assume that consumption of children is fixed at c_2.

The constrained efficient allocation $c_1^*(\theta), l^*(\theta), n^*(\theta)$ solves the following problem

[12]Mathematically, stationary distributions correspond to pairs of eigen vectors and eigen values for the linear operator T, (Ψ, γ). When T is the Markov transition matrix for an irreducible and acyclic chain, there is one positive eigen value (which has value 1), and the associated eigen vector is the the unique invariant distribution. For us, even when T is irreducible and acyclic, there may be two positive eigen values, each corresponding to stationary distributions. These eigen values then correspond to the associated population growth rates.

$$\begin{aligned}
\max \quad & \sum_{i=H,L} \pi_i \left[u(c_i) + h(1 - l_i - bn_i) + \beta n_i^\eta u(c_2)\right] \\
\text{s.t.} \quad & \sum_{i=H,L} \pi_i \left[c_i + \frac{1}{R} n_i c_2\right] \leq \sum_{i=H,L} \pi_i \theta_i l_i + RK_0 \\
& u(c_H) + h(1 - l_H - bn_H) + \beta n_H^\eta u(c_2) \geq \\
& \quad u(c_L) + h(1 - \frac{\theta_L l_L}{\theta_H} - bn_L) + \beta n_L^\eta u(c_2).
\end{aligned}$$

Now suppose that we want to implement the above allocation with a tax in first period of the form $T(y, n)$. Then the consumer's problem is the following:

$$\begin{aligned}
\max \quad & u(c_1) + h(1 - l - bn) + \beta n^\eta u(c_2) \\
\text{s.t.} \quad & c_1 + k_1 \leq Rk_0 + \theta l - T(\theta l, n) \\
& nc_2 \leq Rk_1
\end{aligned}$$

It can be shown that if T is differentiable and if y is interior for both types $T_n(\theta_H l_H^*, n_H^*) = 0, T_y(\theta_H l_H^*, n_H^*) = 0$ – there are no (marginal) distortions on the decisions of the agent with the high shock. Thus, what we need to do is to characterize the types of distortions that are used to get the low type to choose the correct allocation.

It is well known that, even in the simple case in which there is no fertiliy choice, the constrained efficient allocation cannot be implemented by a continously differentiable tax function. (This is also true in our ennvironment.) However, there exists continous and piecewise differentiable tax functions which implement the constrained efficient allocation. Next, we construct the analog of this for our environment.

Let $\bar{u}_L$ (resp. $\bar{u}_H$) be the level of utility received at the socially efficient allocation by the low (resp. high) type, and define two versions of the tax function:

$$\begin{aligned}
\bar{u}_L &= u(y - T_L(y, n) - \frac{1}{R} nc_2) + h(1 - \frac{y}{\theta_L} - bn) + \beta n^\eta u(c_2), \\
\bar{u}_H &= u(y - T_H(y, n) - \frac{1}{R} nc_2) + h(1 - \frac{y}{\theta_H} - bn) + \beta n^\eta u(c_2).
\end{aligned}$$

T_L, is designed to make sure that the low type always gets utility $\bar{u}_L$ if they satisfy their budget constraint with equality while T_H, is defined similarly.[13]

[13] The assumption for fixed c_2 is added in order to simplify the above definitions. We can modify the above

We will build the overall tax code, $T(y,n)$, by using T_L as the effective tax code for the low type and T_H as the one for the high type. Given this, it follows that the distortions, at the margin, faced by the two types are described by the derivatives of T_L (T_H) with respect to y and n.

Proposition 7 *If the allocation is interior,*

1. *The tax function*
$$T(y,n) = \max\{T_L(y,n), T_H(y,n)\}$$
implements the efficient allocation.

2. *There are no distortions in the decisions of the high type –* $\frac{\partial T}{\partial y}(y_H^*, n_H^*) = \frac{\partial T_H}{\partial y}(y_H^*, n_H^*) = 0$ *and* $\frac{\partial T}{\partial n}(y_H^*, n_H^*) = \frac{\partial T_H}{\partial n}(y_H^*, n_H^*) = 0$.

3. *At the choice of the low type,* (y_L^*, n_L^*)*,* T *is differentiable from below with* $\frac{\partial T}{\partial y^-}(y_L^*, n_L^*) = \frac{\partial T_L}{\partial y}(y_L^*, n_L^*) > 0$ *and* $\frac{\partial T}{\partial n^-}(y_L^*, n_L^*) = \frac{\partial T_L}{\partial n}(y_L^*, n_L^*) > 0$.

Proof. See Appendix. ■

The new finding here is that the planner chooses to tax the low type at the margin for having more children – $\frac{\partial T_L}{\partial n}(y_L^*, n_L^*) > 0$. In the Mirrlees model without fertility choice, for incentive reasons, the planner wants to make sure that the low type consumes more leisure (relative to consumption) than he would in a full information world – this makes it easier to get the high type to truthfully admit his type. This is accomplished by having a positive marginal labor tax rate for the low type. Here, there is an additional incentive effect that must be taken care of. This is for the planner to make sure that the low type doesn't use too much of his time free from work raising children. This would also make it more appealing to the high type to lie. To offset this here, the planner also charges a positive tax rate on children for the low type. These two effects taken together ensure that the low type has low consumption and fertility and high leisure thereby separating from the high type.

6.2 Distortionary Wedges

In this section we discuss some of the properties of the socially efficient allocation characterized in the previous sections and how it differs from the full information allocation. We

definition to allow for variable c_2. In that formulation, given (y,n), T_θ is defined as how much present value should be taken away from each type so that he is indifferent between deviating to (y,n) accompanied by optimal saving and no deviation.

focus on the distortions that are present in the leisure-consumption margin, the intertemporal margin and the fertility-leisure margin.

1. The Consumption-Leisure margin: For a type (w, θ), we have:

$$u'(c(w,\theta)) = \frac{h'\left(1 - l(w,\theta) - bn(w,\theta)\right)}{(1-\tau_l(w,\theta))\theta}$$

   where $\tau_l(w, \theta)$ – the 'wedge' in the consumption-leisure margin – is related to the multipliers on the incentive constraints. Because of our assumption that only downward incentive constraints can bind, it can be shown that $\tau_l(w, \theta) > 0$. In this sense, the distortion to the consumption-leisure margin that we find here is the same as that in a standard model without fertility choice. Intuitively, the planner taxes labor income of the lower types to make it easier to separate them from higher types – i.e., to relax incentive constraints.

2. The Intertemporal margin: From the Inverse Euler Equation, we have:

$$E\left[\frac{1}{N_{t+1}(\theta^t)^{\eta-1}u'\left(\frac{C_{t+1}(\theta^{t+1})}{N_{t+1}(\theta^t)}\right)}|\theta^t\right] = \frac{\beta R}{N_t(\theta^{t-1})^{\eta-1}u'\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right)}.$$

   Using Jensen's Inequality, we have:

$$\frac{1}{E\left[N_{t+1}(\theta^t)^{\eta-1}u'\left(\frac{C_{t+1}(\theta^{t+1})}{N_{t+1}(\theta^t)}\right)|\theta^t\right]} < \frac{\beta R}{N_t(\theta^{t-1})^{\eta-1}u'\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right)},$$

   or,

$$N_t(\theta^{t-1})^{\eta-1}u'\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right) < \beta R E\left[N_{t+1}(\theta^t)^{\eta-1}u'\left(\frac{C_{t+1}(\theta^{t+1})}{N_{t+1}(\theta^t)}\right)|\theta^t\right].$$

   This can be rewritten as:

$$N_t(\theta^{t-1})^{\eta-1}u'\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right) = \beta(1-\tau_k(\theta^t))RE\left[N_{t+1}(\theta^t)^{\eta-1}u'\left(\frac{C_{t+1}(\theta^{t+1})}{N_{t+1}(\theta^t)}\right)|\theta^t\right]$$

   where $(1 - \tau_k(\theta^t))$ is the wedge in the intertemporal Euler Equation. [14] When there is

[14] This 'wedge' has multiple implementations using taxes. In some of these the tax on capital income is state contingent from saver's perspective, i.e., $\tau_k(\theta^t, \theta_{t+1})$, while in others it is not, i.e., $\tau_k(\theta^t)$. See Kocherlakota (2005) for a discussion.

no private information, it follows, as usual, that $\tau_k(\theta^t) = 0$. With private information, it follows, from above, that $\tau_k(\theta^t) > 0$. Intuitively, the planner dissuades agents from saving at their normal level to reduce wealth, and hence, incentive problems, in the future.

This is the same as what is found in Golosov et al. (2003). It has a slightly different interpretation here since a period corresponds to a generation. Because of this, it should be interpreted as a distortion in the decision to leave bequests (as in Farhi and Werning (2007)). Thus, we find that the 'tax' on aggregate bequests (B) is positive here. To be added $-\frac{\partial^2 T}{\partial B \partial n} >< 0$?

3. The Fertility-Consumption margin: When there is no private information, the trade-off between fertility and leisure is captured in equation 13:

$$\eta\beta\left(n(w,\theta)\right)^{\eta-1} w' = \frac{1}{R} v(w'(w,\theta)) u'(c(w,\theta)) + bh'\left(1 - l(w,\theta) - bn(w,\theta)\right).$$

When information is private, it can be shown that:

$$\eta\beta\left(n(w,\theta)\right)^{\eta-1} w' > \frac{1}{R} v(w'(w,\theta)) u'(c(w,\theta)) + bh'\left(1 - l(w,\theta) - bn(w,\theta)\right).$$

Thus, this can be rewritten as:

$$\eta\beta\left(n(w,\theta)\right)^{\eta-1} w' = (1+\tau_n(w,\theta)) \frac{1}{R} v(w'(w,\theta)) u'(c(w,\theta)) + bh'\left(1 - l(w,\theta) - bn(w,\theta)\right)$$

where $1 + \tau_n(w,\theta)$ is tax rate on the total increase in future expenditures that comes with increasing dynasty size by one (i.e., on children, their children, their children's children, etc.).

Hence, generalizing what we saw in the two period example, in general, the planner uses a positive tax on fertility.

7 Numerical Examples

In this section we solve an example numerically to illustrate some of the results presented in previous sections. We also explore some properties of optimal social contract that we have not established formally.

Individuals have CRRA preferences over consumption and leisure

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma}, \text{ and } h(m) = \phi\frac{m^{1-\sigma}}{1-\gamma}$$

in which $m = 1 - l - bn$ is leisure, l is hours worked and n is number of kids for each parent.

For this example we asssume the following values for parameters: $\beta = 0.3$, $R = 4$, $\sigma = 1.5$, $\phi = 0.5$, $b = 0.41$ and $\eta = -2$. We assume two levels of productivity shocks $\{\theta_L, \theta_H\} = \{2, 6\}$. Shocks are i.i.d across generations and dynasties and the probability of the high shock is $\pi_H = 0.1$.

7.1 Endogenous fertility: private information vs. full information

Figure 1 shows the allocation in the economy in which fertility choice is endogenous. We have presented the allocation rule for both the full information and private information economies. The graphs highlight a few points.

1. If curren promised utility to the parent is high enough, it is efficient to have the low ability type work zero hours. This holds both under full information and private information. Furthermore, if current promised utility to the parent is even higher, it is efficient to have both types work zero hours. At this level of promised utility (and higher levels) the utility is delivered in a first best fashion.

2. For all levels of current promised utility such that hours are positive, the fertility allocation is monotone increasing in current promised utility (full information and private information). For levels of current promised utility such that hours are zero, fertility decreases with current promised utility.

3. For all levels of current promised utility such that hours are positive, promised utility to children under full information is independent of current promised utility of the parent (this is formally established in previous sections).

4. For all levels of current promised utility such that hours are positive, promised utility to the high skilled parents' children under private information is independent of current promised utility of the parent (this is formally established in previous sections).

5. For all levels of current promised utility such that hours are positive, promised utility to the low skilled parents' children under private information information is monotone decreasing in current promised utility of the parent.

6. For all levels of current promised utility that hours is zero, promised utility to the kids is monotone increasing in current promised utility of the parent.

It is important to note here that incentives are provided both by the level of promised utility to the children and the number of children. In other words the future utility that is promised to a parent is $n(w,\theta)^{\eta} w'(w,\theta)$. This promised utility is always monotone increasing in the current utility promised to the parent. Figure 2 illustrates this. Also note that, under full information, the future utility promised to the high skilled parents is always lower than the one promised to low skilled parents (for full info the blue line lies below the red line). Also, it appears that under full information, future promised utility to the parents have higher variance.

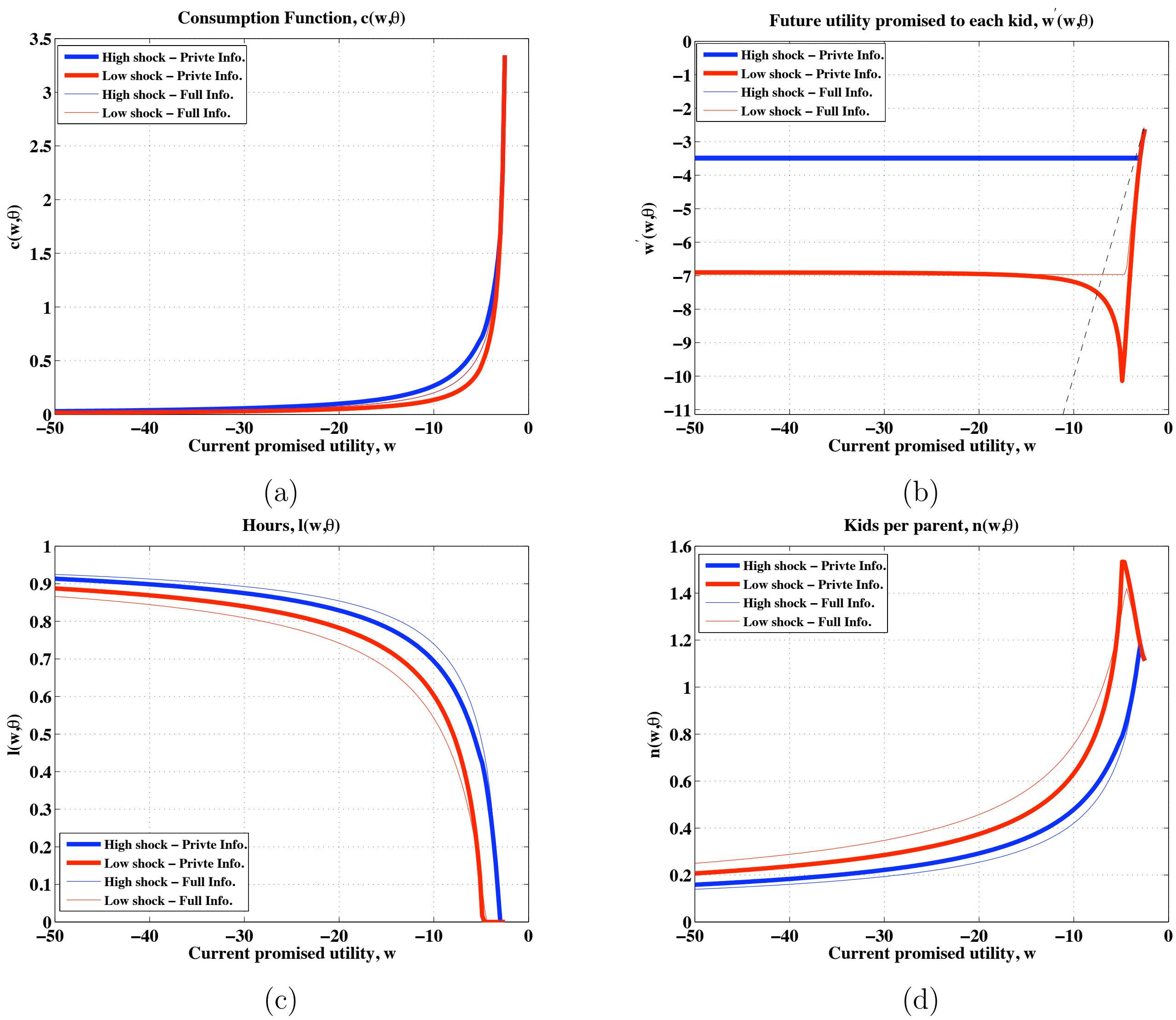


Figure 1: Optimal consumption, hours, fertility and promised utility allocations. Allocations in private information economy are plotted in thicker line. The blue indicates high shock and red indicates low shock. The dashed line in panel (b) is the 45 degree line.

We can use the procedure outlined in section 5 to compute the growth rate of population, γ and the stationary distribution of promised utility, Ψ. For full the information economy the stationary distribution has support $\mathcal{W}^{FI} = \{-9.1607, -4.6591, -3.1255\}$ and the frequency distribution is given by $\psi^{FI} = \{0.8998, 0.058, 0.0421\}$. The growth rate of population is $\gamma^{FI} = 1.0624$.

For the private information economy the stationary distribution has support $\mathcal{W}^{PI} = \{-11.5428, -9.5892, -9.5837, -9.5836, -9.5833, -9.4970, -4.6938, -3.0826\}$ and the frequency distribution is $\psi^{PI} = \{0.1495, 0.0927, 0.0660, 0.3581, 0.0782, 0.1089, 0.0858, 0.0607\}$. The growth rate of population is $\gamma^{PI} = 1.0355$. Figure 3 shows the stationary distributions.

7.2 Endogenous fertility vs. exogenous fertility

In this section we compare the efficient allocation under private information in the benchmark model (with fertility choice) to the allocation that comes out of a standard Mirrleesian environment. The functional forms and parameters are the same as previous section. We compare the case where $\eta = 0$ (exogenous fertility) to the case where $\eta = -2$.

Figure 4 shows the consumption, hours and promised utility (to the kids) allocations.

Figure 5 plots the future promised utility to the parents for both environments, i.e. $n(w, \theta)^{\eta} w'(w, \theta)$. Note that in the standard dynamic Mirrleesian environment, $\eta = 0$ and future promised utility to the parent is the same as utility promised to the each kid.

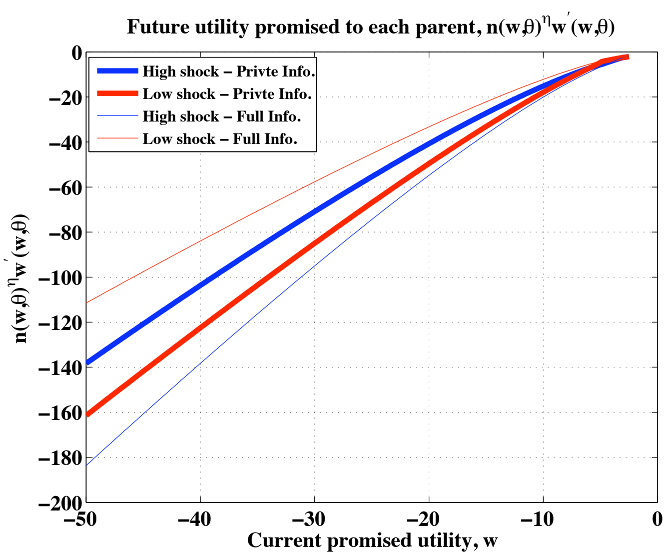
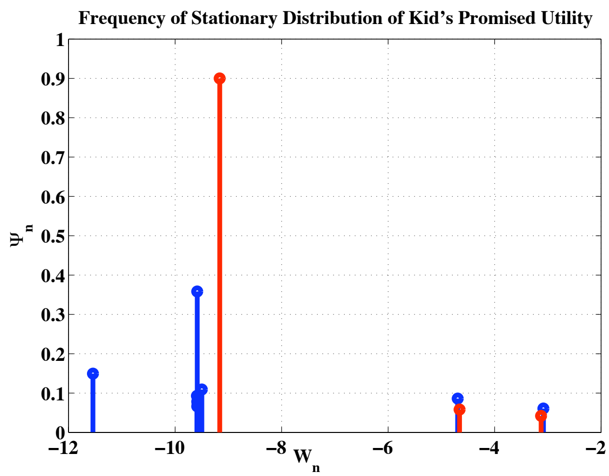
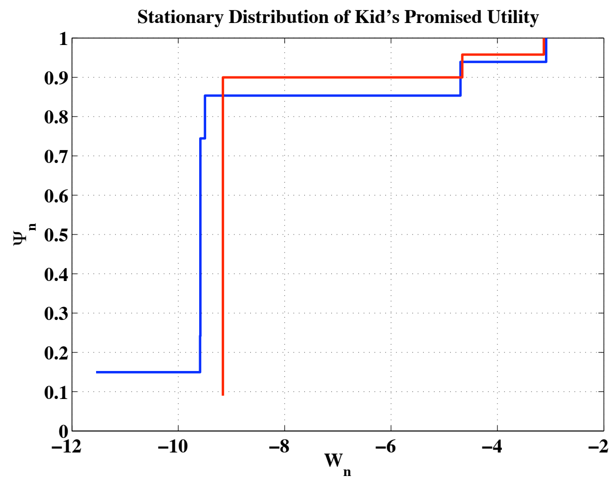


Figure 2: Future promised utility to the parents. Allocations in private information economy are plotted in thicker line. The blue indicates high shock and red indicates low shock.



(a)



(b)

Figure 3: Stationary distribution. Panel (a) shows the frequency and panel (b) shows the CDF. Stationary distribution in private information economy is plotted in blue. The red graph is the stationary distribution under full information.

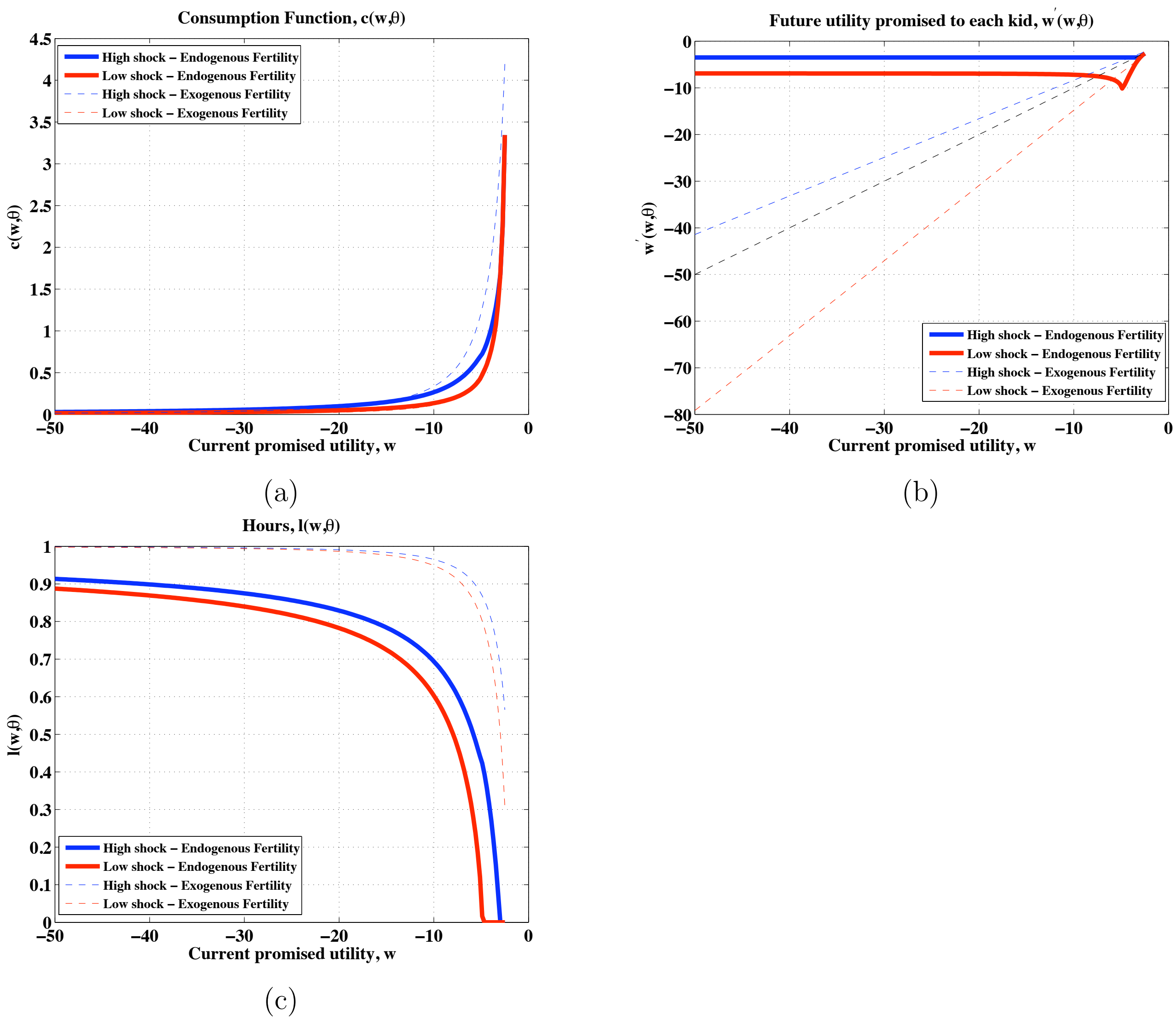


Figure 4: Optimal consumption, hours, fertility and promised utility (to the kids) allocations. Allocations in benchmark economy are plotted in thicker line. The blue indicates high shock and red indicates low shock. The dashed line in panel (b) is the 45 degree line.

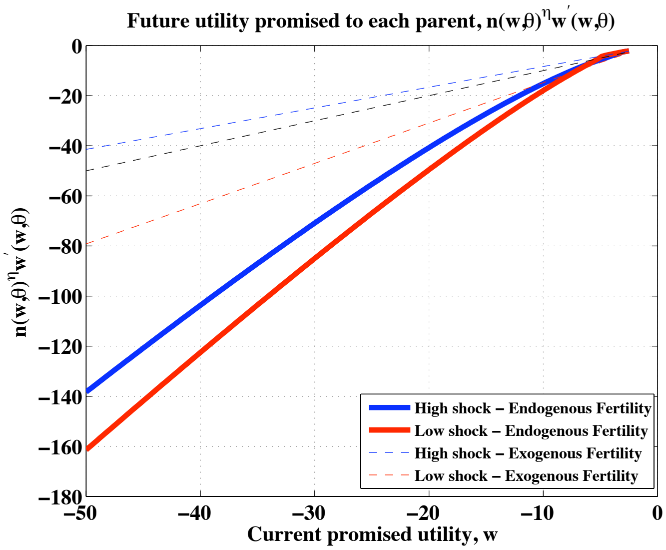


Figure 5: Future promised utility to the parents. Allocations in benchmark economy are plotted in thicker line. The blue indicates high shock and red indicates low shock. . The dashed line is the 45 degree line (it is relevant only for the $\eta = 0$ case, since otherwise the future utility and current utility promised to the parent are not in the same space).

Appendix

A Proofs

A.1 Proof of Proposition 1

We first prove part 4.

Define $G(C, N) = N^\eta u\left(\frac{C}{N}\right)$. We then have

$$
\begin{aligned}
G_C(C,N) &= N^{\eta-1} u'\left(\frac{C}{N}\right) \\
G_N(C,N) &= \eta N^{\eta-1} u\left(\frac{C}{N}\right) - N^{\eta-2} C u'\left(\frac{C}{N}\right) \\
G_{CC}(C,N) &= N^{\eta-2} u'\left(\frac{C}{N}\right) \\
G_{CN}(C,N) &= (\eta-1) N^{\eta-2} u'\left(\frac{C}{N}\right) - N^{\eta-3} C u''\left(\frac{C}{N}\right) \\
G_{NN}(C,N) &= \eta(\eta-2) N^{\eta-2} u\left(\frac{C}{N}\right) - 2(\eta-1) N^{\eta-3} C u'\left(\frac{C}{N}\right) + N^{\eta-4} C^2 u''\left(\frac{C}{N}\right)
\end{aligned}
$$

For strict concavity of G, we must have $G_{CC} < 0$, $G_{CC}G_{NN} - G_{NC}^2 > 0$. This implies that

$$G_{CC}(C,N) = N^{\eta-2}u''\left(\frac{C}{N}\right) < 0$$
$$G_{CC}G_{NN} - G_{CN}^2 = (\eta - 1)\left(\eta u\left(\frac{C}{N}\right)u''\left(\frac{C}{N}\right) - (\eta-1)u'\left(\frac{C}{N}\right)^2\right) > 0$$

The first inequality implies that $u(c)$ is a concave function. As for the second inequality, notice that when η is greater than 1, then since $u'' < 0$, the term in the brackets is negative which contradicts the inequality. Therefore, we have - replacing c for $\frac{C}{N}$:

$$\eta < 1 \quad , \quad \eta u(c)u''(c) - (\eta - 1)u'(c)^2 < 0$$

Now, we can rewrite (8) as follows:

$$\eta \frac{u(c_2(\theta))}{u'(c_2(\theta))} - c_2(\theta) = b\theta R$$

taking derivative of the left hand side with respect to c_2 gives the following:

$$\eta \frac{u'(c)^2 - u'(c)u''(c)}{u'(c)^2} - 1 = \frac{(\eta-1)u'(c)^2 - \eta u(c)u''(c)}{u'(c)^2} > 0$$

where the inequality is from above. Therefore $c_2(\theta)$ is increasing in θ and independent of W_0.
Equations (3) and (4), imply that $c_1(\theta)$, $n(\theta)$, and $m(\theta)$-leisure, are positively correlated. Hence the binding feasibility implies that they are strictly increasing in W_0 and $l(\theta) = 1 - m(\theta) - bn(\theta)$ is st. decreasing in W_0.
Moreover, since by (2), $c_1(\theta)$ is independent of θ, equation (4) implies that $n(\theta)$ is negatively correlated with $c_2(\theta)$ and hence $n(\theta)$ is st. decreasing in θ. Also by equation (3), leisure is decreasing in θ and therefore, labor supply is increasing in θ. ■

A.2 Proof of Corollary 2

Proof. We know from above that $V(N,W) = Nv(N^{-\eta}W)$. Strict concavity of $V(N,W)$ implies that $V_{WW} > 0, V_{NN} > 0, V_{WW}V_{NN} > V_{WN}^2$. We have

$$\begin{aligned}
V_{WW} &= N^{1-2\eta}v''(N^{-\eta}W) = N^{1-2\eta}v''(w) \\
V_{WN} &= (1-\eta)N^{-\eta}v'(N^{-\eta}W) - \eta N^{-2\eta}Wv''(N^{-\eta}W) = N^{-\eta}\left((1-\eta)v'(w) - \eta wv''(w)\right) \\
V_{NN} &= \eta(\eta-1)N^{-\eta-1}Wv'(N^{-\eta}W) + \eta^2 N^{-2\eta-1}W^2v''(N^{-\eta}W) \\
&= \eta N^{-1}w\left((\eta-1)v'(w) + \eta wv''(w)\right)
\end{aligned}$$

After some algebra, we have

$$V_{WW}V_{NN} - V_{WN}^2 = N^{-2\eta}\left((\eta-1)v'(w) + \eta wv''(w)\right)(1-\eta)v'(w)$$

Therefore, the strict convexity of $V(\cdot,\cdot)$ implies that:

$$\begin{aligned}
&v''(w) > 0 \\
&(\eta-1)v'(w) + \eta wv''(w) = \frac{d}{dw}\left(\eta wv'(w) - v(w)\right) > 0 \\
&v'(w) > 0
\end{aligned}$$

■

A.3 Proof of Proposition 2

Proof.

1. We have shown the results for $w'(\cdot,\cdot)$ above. (18) immediately implies that $c(w,\theta)$ is independent of θ. Now, by increasing θ, $w'(w,\theta)$ goes up as well as $v'(w'(w,\theta))$ since v is concave. Because, $c(w,\theta)$ is independent of θ, (16) implies that $n(w,\theta)^{\eta-1}$ is increasing in θ, and since $\eta < 1$, we have that $n(w,\theta)$ is st decreasing in θ. (15) implies that leisure is st. decreasing in θ and since fertility is decreasing, hours worked will be increasing in θ.

2. Convexity of $v(w)$ and (18) implies that $c(w,\theta)$ is increasing in w. An increase in w, causes a decrease in marginal utility of consumption and (16) implies that fertility is increasing in w. Moreover, (15) means that leisure is increasing in w and therefore hours worked is decreasing in w.

■

A.4 Closed Form Solutions for Fertility

A.4.1 Two Period Example

Now if we assume that utility from consumption is CRRA, $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$, then (8) becomes

$$\frac{\eta+\sigma-1}{1-\sigma}c_2(\theta) = bR\theta \Rightarrow c_2(\theta) = \frac{1-\sigma}{\sigma+\eta-1}bR\theta$$

Equation (3) then implies that

$$n(\theta) = \left(\left(\frac{(1-\sigma)bR}{(\sigma+\eta-1)}\right)^{\sigma} c_1^{-\sigma}\right)^{\frac{1}{\eta-1}} \theta^{\frac{\sigma}{\eta-1}}$$

The above equation implies that fertility is a constant elasticity function of productivity, an observation that is in accordance with the data[15].

A.4.2 Fully Dynamic Version

To gain more intuition about how the model works, it is worth looking at the sequence problem directly and characterize its solution. Here, we will focus on problem (P1')[16]. The problem is the following:

$$\max \sum_{t=0}^{\infty} \beta^t \sum_{\theta^t} \pi(\theta^t) N_t^{\eta}(\theta^{t-1}) \left(u\left(\frac{C_t(\theta^t)}{N_t(\theta^{t-1})}\right) + h\left(1 - \frac{L_t(\theta^t)}{N_t(\theta^{t-1})} - b\frac{N_{t+1}(\theta^t)}{N_t(\theta^{t-1})}\right)\right) \qquad \text{(P1')}$$

$$\text{s.t.} \sum_{t=0}^{\infty} \frac{1}{R^t} \sum_{\theta^t \in \Theta^{t+1}} \pi(\theta^t)\left(C_t(\theta^t) - \theta_t L_t(\theta^t)\right) \leq K_0$$

$$N_0(\theta^{-1}) = N_{-1} : \text{given}$$

Notice that N does not appear in the constraint set since we are looking at total allocations for each cohort with the same history. Assuming that the multiplier for feasibility constraint

[15]See Jones and Tertilt (2008)

[16]Solving (P1) instead of (P1') is more complicated since each fertility level appears in infinitely many terms.

is λ, we will have the following equations:

$$\beta^t R^t N_t^{\eta-1} u'\left(\frac{C_t}{N_t}\right) = \lambda \tag{28}$$

$$\beta^t R^t N_t^{\eta-1} h'\left(1 - \frac{L_t}{N_t} - b\frac{N_{t+1}}{N_t}\right) = \theta_t \lambda \tag{29}$$

$$\begin{aligned}\beta^{t+1}\eta N_{t+1}^{\eta-1} E_t\left\{u\left(\frac{C_{t+1}}{N_{t+1}}\right) + h\left(1 - \frac{L_{t+1}}{N_{t+1}} - b\frac{N_{t+2}}{N_{t+1}}\right)\right\} = b\beta^t N_t^{\eta-1} h'\left(1 - \frac{L_t}{N_t} - b\frac{N_{t+1}}{N_t}\right)\\ +\beta^{t+1} N_{t+1}^{\eta-2} E_t\left\{C_{t+1} u'\left(\frac{C_{t+1}}{N_{t+1}}\right) - \left(L_{t+1} + bN_{t+1}\right) h'\left(1 - \frac{L_{t+1}}{N_{t+1}} - b\frac{N_{t+2}}{N_{t+1}}\right)\right\}\end{aligned} \tag{30}$$

The above equations have the same interpretation as (15)-(18). Now assume that $u(c) = \frac{c^{1-\sigma}}{1-\sigma}, h(m) = \psi\frac{m^{1-\sigma}}{1-\sigma}$[17]. Combining (28) and (29) leads the following:

$$\frac{1}{\theta_t}\beta^t R^t N_t^{\eta-1} h'(m_t) = \beta^{t+1} R^{t+1} N_{t+1}^{\eta-1} u'(c_{t+1})$$

Then (30) becomes the following:

$$\beta^{t+1} N_{t+1}^{\eta-1} E_t\left\{\eta(u(c_{t+1}) + h(m_{t+1})) - c_{t+1}u'(c_{t+1}) + (1 - m_{t+1})h'(m_{t+1}))\right\} = b\beta^{t+1} R N_{t+1}^{\eta-1}\theta_t u'(c_{t+1})$$

and hence

$$\eta\frac{c_{t+1}^{1-\sigma}}{1-\sigma} - c_{t+1}^{1-\sigma} + E_t\left\{\eta\psi\frac{m_{t+1}^{1-\sigma}}{1-\sigma} + \psi(1 - m_{t+1})m_{t+1}^{-\sigma}\right\} = bR\theta_t c_{t+1}^{-\sigma}$$

From the trade-off between leisure and consumption we have that $\psi m_{t+1}^{-\sigma} = \theta_{t+1} c_{t+1}^{-\sigma}$ and

$$\frac{\eta+\sigma-1}{1-\sigma} c_{t+1} + E_t\frac{\eta+\sigma-1}{1-\sigma} c_{t+1}\theta_{t+1}^{1-1/\sigma}\psi^{1/\sigma} + E_t\theta_{t+1} = bR\theta_t$$

Therefore,

$$c_{t+1}(\theta^t) = \underbrace{\left(\frac{1-\sigma}{\eta+\sigma-1}\frac{bR}{1+\psi^{1/\sigma}E\theta^{1/\sigma}}\right)}_{A}\theta_t - \underbrace{\frac{1-\sigma}{\eta+\sigma-1}\frac{\psi}{1+\psi^{1/\sigma}E\theta^{1/\sigma}}}_{B}$$

and for fertility we have

$$n_t(\theta^t) = \left(\frac{1}{\beta R}\right)^{\frac{1}{\eta-1}}\left(\frac{A\theta_t - B}{A\theta_{t-1} - B}\right)^{\frac{\sigma}{\eta-1}}$$

[17] we will use m for leisure onward.

A.5 Efficient Allocation is not Incentive Compatible

In this section, we show that the efficient allocation with full information does not satisfy the incentive compatibility constraints for the maximization problem in section 4.

Intuitively, from intra-family risk sharing, equation (15), we know that per capita consumption among siblings is equal. Moreover, efficiency requires that leisure is decreasing in productivity, equation (16). It is therefore sufficient to show that future utility for a low productivity agent is higher than a high productivity agent.

This is shown below.

One intuition for this comes from the curvature properties of the cost function, $v(w)$. In the unconstrained efficient allocation, the planner equates per capita marginal cost $n(\theta)^{1-\eta}v'(w'(\theta))$ across various types. We know from Corollary 2 that $v'(\cdot)$ has a curvature higher than $\frac{1-\eta}{\eta}$. Therefore, for a given relative fertility $\Delta > 1$, equating per capita marginal cost implies that relative promised utility is at most Δ^{η} which implies that $n(\theta)^{\eta}w'(\theta)$ has the same direction as $n(\theta)$ does. Hence, overall promised value, $n(\theta)^{\eta}w'(\theta)$, is higher for lower productivity agents.

Formally, consider the efficient allocation, which is the solution to the dynamic programming problem (P4). From (16) and (18), we have that

$$n(\theta)^{1-\eta}v'(w'(\theta)) = n(\theta')^{1-\eta}v'(w'(\theta')), \forall\theta, \theta' \tag{31}$$

Moreover, from Corollary (2), we know that

$$\frac{wv''(w)}{v'(w)} > \frac{1-\eta}{\eta} \Rightarrow \frac{v''(w)}{v'(w)} > \frac{1-\eta}{\eta}\frac{1}{w}$$

If we assume that $\theta > \theta'$, then $w'(\theta) > w'(\theta')$ and we can integrate the above equation to obtain that

$$\log\left(\frac{v'(w'(\theta))}{v'(w'(\theta'))}\right) = \int_{w'(\theta')}^{w'(\theta)} \frac{v''(w)}{v'(w)}dw > \frac{1-\eta}{\eta}\log\left(\frac{w'(\theta)}{w'(\theta')}\right)$$

and therefore

$$\frac{v'(w'(\theta))}{v'(w'(\theta'))} > \left(\frac{w'(\theta)}{w'(\theta')}\right)^{\frac{1-\eta}{\eta}}$$

Combining the above with (31), we get the following inequality

$$\left(\frac{n(\theta')}{n(\theta)}\right)^{1-\eta} = \frac{v'(w'(\theta))}{v'(w'(\theta'))} > \left(\frac{w'(\theta)}{w'(\theta)}\right)^{\frac{1-\eta}{\eta}} \Rightarrow n(\theta')^{\eta}w'(\theta') > n(\theta)^{\eta}w'(\theta) \tag{32}$$

Moreover, from proposition 1, we know that $c(\theta)$ does not depend on θ and leisure is decreasing in θ. Therefore $1 - l(\theta) - bn(\theta) < 1 - l(\theta') - bn(\theta') < 1 - \theta' l(\theta')/\theta - bn(\theta')$, when $\theta > \theta'$. These properties together with (32) gives us the following inequality

$$u(c(\theta)) + h(1 - l(\theta) - bn(\theta)) + \beta n(\theta)^\eta w'(\theta) \leq$$
$$< u(c(\theta')) + h(1 - \frac{\theta' l(\theta')}{\theta} - bn(\theta')) + \beta n(\theta')^\eta w'(\theta'), \forall \theta > \theta'$$

which means that under the efficient allocation, agents with higher productivity would like to pretend to be low productivity. So the unconstrained efficient allocation is not incentive compatible. ■

A.6 Sufficient Condition for Slackness of Upward Incentive Constraints

In this section, we give sufficient conditions for slackness of upward incentive constraints. That is we show that if the efficient allocation satisfies certain constraint, then downward incentive constraints are sufficient. We summarize the sufficient conditions in the following lemma:

Lemma 1 *Suppose an allocation $(c(\theta), l(\theta), n(\theta), w'(\theta))$ satisfies the following:*

1. *$l(\theta)\theta$ is increasing in θ ,*
2. *$1 - l(\theta) - bn(\theta) \leq 1 - \frac{\theta' l(\theta')}{\theta} - bn(\theta')$, for all $\theta > \theta'$*
3. *Local downward incentive constraints are binding:*

$$u(c(\theta_i)) + h(1 - l(\theta_i) - bn(\theta_i)) + \beta n(\theta_i)^\eta w'(\theta_i) =$$
$$u(c(\theta_{i-1})) + h(1 - \frac{\theta_{i-1} l(\theta_{i-1})}{\theta_i} - bn(\theta_{i-1})) + \beta n(\theta_{i-1})^\eta w'(\theta_{i-1})$$

Then, incentive compatibility holds for any θ, θ'.

Proof. By part 3 of the assumption, we have

$$u(c(\theta_{i-1})) + \beta n(\theta_{i-1})^\eta w'(\theta_{i-1}) - u(c(\theta_i)) - \beta n(\theta_i)^\eta w'(\theta_i) =$$
$$= h(1 - l(\theta_i) - bn(\theta_i)) - h(1 - \frac{\theta_{i-1} l(\theta_{i-1})}{\theta_i} - bn(\theta_{i-1}))$$

By part 2 and 3 of the assumption we have

$$\frac{1}{\theta_{i-1}}(\theta_i l(\theta_i) - \theta_{i-1} l(\theta_{i-1})) \geq \frac{1}{\theta_i}(\theta_i l(\theta_i) - \theta_{i-1} l(\theta_{i-1})) \geq b(n(\theta_{i-1}) - n(\theta_i))$$

Hence, for any $x \in [1/\theta_i, 1/\theta_{i-1}]$

$$\begin{aligned} &x(\theta_i l(\theta_i) - \theta_{i-1} l(\theta_{i-1})) \geq b(n(\theta_{i-1}) - n(\theta_i)) \\ &\Rightarrow 1 - x\theta_{i-1} l(\theta_{i-1}) - bn(\theta_{i-1}) \geq 1 - x\theta_i l(\theta_{i-1}) - bn(\theta_i) \end{aligned}$$

Therefore, using part 1 and concavity of $h(\cdot)$,

$$-h'(1 - x\theta_{i-1} l(\theta_{i-1}) - bn(\theta_{i-1}))\theta_{i-1} l(\theta_{i-1}) \geq -h'(1 - x\theta_i l(\theta_{i-1}) - bn(\theta_i))\theta_i l(\theta_i)$$

Integrating both sides from $1/\theta_i$ to $1/\theta_{i-1}$, we get

$$h(1-l(\theta_{i-1})-bn(\theta_{i-1}))-h(1-\frac{\theta_{i-1}l(\theta_{i-1})}{\theta_i}-bn(\theta_{i-1})) \geq h(1-\frac{\theta_i l(\theta_i)}{\theta_{i-1}}-bn(\theta_i))-h(1-l(\theta_i)-bn(\theta_i))$$

Therefore,

$$\begin{aligned} u(c(\theta_{i-1})) + \beta n(\theta_{i-1})^\eta w'(\theta_{i-1}) - u(c(\theta_i)) - \beta n(\theta_i)^\eta w'(\theta_i) \geq \\ \geq h(1 - \frac{\theta_i l(\theta_i)}{\theta_{i-1}} - bn(\theta_i)) - h(1 - l(\theta_{i-1}) - bn(\theta_{i-1})) \end{aligned}$$

Hence, the local upward incentive constraints are satisfied.

Now, we will show that other upward incentive constraints are satisfied . To illustrate we show the argument for i and $i+2$ and a similar inductive argument works for higher differences. By condition 2, we know that:

$$\frac{1}{\theta_{i+2}}(\theta_{i+2} l(\theta_{i+2}) - \theta_{i+1} l(\theta_{i+1})) \geq b(n(\theta_{i+1}) - n(\theta_{i+2}))$$

and therefore,

$$\frac{1}{\theta_i}(\theta_{i+2} l(\theta_{i+2}) - \theta_{i+1} l(\theta_{i+1})) \geq \frac{1}{\theta_{i+1}}(\theta_{i+2} l(\theta_{i+2}) - \theta_{i+1} l(\theta_{i+1})) \geq b(n(\theta_{i+1}) - n(\theta_{i+2}))$$

Hence, for any $x \in [1/\theta_{i+1}, 1/\theta_i]$,

$$1 - x\theta_{i+1} l(\theta_{i+1}) - bn(\theta_{i+1}) \geq 1 - x\theta_{i+2} l(\theta_{i+2}) - bn(\theta_{i+2})$$

and we have,

$$\begin{aligned} h'(1-x\theta_{i+1}l(\theta_{i+1})-bn(\theta_{i+1})) &\leq h'(1-x\theta_{i+2}l(\theta_{i+2})-bn(\theta_{i+2})) \\ \Rightarrow -h'(1-x\theta_{i+1}l(\theta_{i+1})-bn(\theta_{i+1}))\theta_{i+1}l(\theta_{i+1}) &\geq -h'(1-x\theta_{i+2}l(\theta_{i+2})-bn(\theta_{i+2}))\theta_{i+2}l(\theta_{i+2}) \end{aligned}$$

So,

$$\begin{aligned} & h(1-\frac{\theta_{i+1}l(\theta_{i+1})}{\theta_i}-bn(\theta_{i+1}))-h(1-l(\theta_{i+1})-bn(\theta_{i+1})) \\ & \quad \geq h(1-\frac{\theta_{i+2}l(\theta_{i+2})}{\theta_i}-bn(\theta_{i+2}))-h(1-\frac{\theta_{i+2}l(\theta_{i+2})}{\theta_{i+1}}-bn(\theta_{i+2})) \end{aligned} \tag{33}$$

Rewriting local IC's for $i, i+1$ and $i+1, i+2$:

$$\begin{aligned} & u(c(\theta_i))+\beta n(\theta_i)^\eta w'(\theta_i)-u(c(\theta_{i+1}))-\beta n(\theta_{i+1})^\eta w'(\theta_{i+1}) \geq \\ & \quad \geq h(1-\frac{\theta_{i+1}l(\theta_{i+1})}{\theta_i}-bn(\theta_{i+1}))-h(1-l(\theta_i)-bn(\theta_i)) \end{aligned} \tag{34}$$

$$\begin{aligned} & u(c(\theta_{i-1}))+\beta n(\theta_{i+1})^\eta w'(\theta_{i+1})-u(c(\theta_{i+2}))-\beta n(\theta_{i+2})^\eta w'(\theta_{i+2}) \geq \\ & \quad \geq h(1-\frac{\theta_{i+2}l(\theta_{i+2})}{\theta_{i+1}}-bn(\theta_{i+2}))-h(1-l(\theta_{i+1})-bn(\theta_{i+1})) \end{aligned} \tag{35}$$

Summing over inequalities (33)-(34), we get:

$$\begin{aligned} & u(c(\theta_{i-1}))+\beta n(\theta_i)^\eta w'(\theta_i)-u(c(\theta_{i+2}))-\beta n(\theta_{i+2})^\eta w'(\theta_{i+2}) \geq \\ & \quad \geq h(1-\frac{\theta_{i+2}l(\theta_{i+2})}{\theta_i}-bn(\theta_{i+2}))-h(1-l(\theta_i)-bn(\theta_i)) \end{aligned}$$

which is the upward incentive constraint for $i, i+2$. The rest of the upward and downward incentive constraints can be proved in a similar way. ■

A.7 Proof of Proposition 6

We first prove the following lemma:

Lemma 2 *Suppose Assumptions 3 and 4 hold, then the value function and the policy functions satisfy the following properties:*

$$\begin{aligned}
\lim_{w\to-\infty} v(w) &= -\sum_i \pi_i\theta_i \\
\lim_{w\to-\infty} c(w,\theta_i) &= 0 \\
\lim_{w\to-\infty} n(w,\theta_i) &= 0 \\
\lim_{w\to-\infty} l(w,\theta_i) &= 1
\end{aligned}$$

Proof. Consider the following set of function:

$$S = \left\{ \hat{v}; \hat{v} \in C(\mathbb{R}_-),\ \hat{v}\text{: weakly increasing } \lim_{w\to-\infty} \hat{v}(w) = -\sum_i \pi_i\theta_i \right\}$$

Moreover define the following mapping on S as

$$\begin{aligned}
T\hat{v}(w) = \quad \min \quad & \sum_j \pi_j \left[c_j - \theta_j l_j + \frac{1}{R} n_j \hat{v}(w'_j) \right] \\
\text{s.t.} \quad & \sum_j \pi_j \left[u(c_j) + h(1 - l_j - bn_j) + \beta n_j^\eta w'_j \right] \geq w \\
& u(c_j) + h(1 - l_j - bn_j) + \beta n_j^\eta w'_j \geq \\
& \qquad u(c_i) + h(1 - \frac{\theta_i l_i}{\theta_j} - bn_i) + \beta n_i^\eta w'_i,\ \forall j > i \\
& c_j, l_j, n_j \geq 0 \\
& 1 \geq l_j + bn_j
\end{aligned}$$

We first show that the solution to the above program has the claimed property for the policy function and that $T\hat{v}$ satisfies the claimed property. Then, since S is closed and T preserves S, by Contraction Mapping Theorem we have that the fixed point of T belongs to S.

Now, suppose the claim about policy function for fertility, does not hold. Then there exists a sequence $w_n \to -\infty$ such that for some i, $n(w_n, \theta_i) \to \bar{n}_i > 0$. For each $j \neq i$, define $\bar{n}_j = \liminf_{n\to\infty} n(w_n, \theta_j)$, then we must have

$$\liminf_{n\to\infty} T\hat{v}(w_n) \geq \sum_j \pi_j [-\theta_j(1 - b\bar{n}_j) + \frac{1}{R}\bar{n}_j \left[-\sum_k \pi_k\theta_k \right]$$

Note that by Assumption 4, we have

$$b\theta_j > \frac{1}{R}\sum_k \pi_k\theta_k,\ \forall j$$

and therefore, if $\bar{n}_j \geq 0$, we must have

$$-\theta_j + b\bar{n}_j\theta_j - \frac{1}{R}\bar{n}_j\sum_k \pi_k\theta_k \geq -\theta_j$$

with equality only if $\bar{n}_j = 0$. This implies that

$$\liminf_{n\to\infty} T\hat{v}(w_n) > -\sum_j \pi_j\theta_j$$

since $\bar{n}_i > 0$.

Now,we construct a sequence of allocation and show that the above cannot be an optimal one. Consider a sequence of numbers ϵ_m that converges to zero. Define

$$\begin{aligned} c_m(\theta_i) &= u^{-1}(-\epsilon_m^\eta) \\ n_m(\theta_i) &= (n-i)^{\frac{1}{\eta}}\epsilon_m \\ w'_m(\theta_i) &= \tilde{w} < 0 \end{aligned}$$

If h is bounded above and below, define

$$l_m(\theta_i) = 1 - \epsilon_m - bn_m(\theta_i)$$

By construction,

$$\begin{aligned} c_m(\theta_i) &\to 0 \\ n_m(\theta_i) &\to 0 \\ l_m(\theta_i) &\to 1 \end{aligned}$$

Moreover,

$$u(c_m(\theta_j)) + \beta n_m(\theta_j)^\eta w'_m(\theta_j) - u(c_m(\theta_i)) - \beta n_m(\theta_i)^\eta w'_m(\theta_i) = \beta\tilde{w}\epsilon_m^\eta(i-j),\ \forall j > i$$

This expression converges to ∞ and therefore, since h is bounded above and below for m large enough, the allocations are incentive compatible.

When, h is unbounded below, since the utility of deviation is bounded away from $-\infty$, it is possible to construct a sequence for l_m that converges to 1. Find $l_m(\theta_i)$ such that

$$h(1 - l_m(\theta_i) - bn_m(\theta_i)) = \frac{1}{2}\tilde{w}\epsilon_m^\eta$$

Hence, we have:

$$\begin{aligned} &u(c_m(\theta_j)) + h(1 - l_m(\theta_j) - bn_m(\theta_j)) + \beta n_m(\theta_j)^\eta w'_m(\theta_j) - u(c_m(\theta_i)) - \beta n_m(\theta_i)^\eta w'_m(\theta_i) \\ &\quad = \tilde{w}\epsilon_m^\eta(i - j + \frac{1}{2}) \end{aligned}$$

converges to ∞. Moreover, by definition $l_m(\theta_i)$ converges to 1 and $n_m(\theta_i)$ converges to zero and therefore the deviation value for leisure, $h(1 - \frac{\theta_i l_m(\theta_i)}{\theta_j} - bn_m(\theta_i))$, converges to $h(1 - \frac{\theta_i}{\theta_j})$. This implies that for m large enough

$$\begin{aligned} &u(c_m(\theta_j)) + h(1 - l_m(\theta_j) - bn_m(\theta_j)) + \beta n_m(\theta_j)^\eta w'_m(\theta_j) - u(c_m(\theta_i)) - \beta n_m(\theta_i)^\eta w'_m(\theta_i) \\ &\quad \geq h(1 - \frac{\theta_i l_m(\theta_i)}{\theta_j} - bn_m(\theta_i)) \end{aligned}$$

and for m large enough the allocation is incentive compatible.

Therefore, The utility from the constructed allocation is the following:

$$\hat{w}_m = \epsilon_n^\eta \left[-1 + \beta \sum A_j^\eta \tilde{w}\right] + \sum_k \pi_j[h(1 - l_m(\theta_j) - bn_m(\theta_j))]$$

It is clear that $\hat{w}_m$'s converge to $-\infty$ and the allocation's cost converges to $-\sum_j \pi_j\theta_j$. Now since $\hat{w}_m$and $w_n$converge to $-\infty$, there exists subsequences $\hat{w}_{m_k}$ and w_{n_k} such that $\hat{w}_{m_k} \geq w_{n_k}$ and therefore by optimality:

$$\sum_j \pi_j \left[c_{m_k}(\theta_j) - \theta_j l_{m_k}(\theta_j) + \frac{1}{R} n_{m_k}(\theta_j)\hat{v}(\tilde{w})\right] \geq T\hat{v}(w_{n_k})$$

and therefore,

$$-\sum_k \pi_k\theta_k \geq \liminf_{n\to\infty} T\hat{v}(w_n) > -\sum_k \pi_k\theta_k$$

and we have a contradiction. This completes the proof. ■

Since h is unbounded below, given the above lemma for $w \in \mathbb{R}_-$ low enough, allocations should be interior and since v is differentiable, positive lagrange multipliers $\lambda, \mu(i,j)|_{i>j}$

must exists such that

$$u'(c(w,\theta_i))\left[\pi_i\lambda(w)+\sum_{j<i}\mu(i,j;w)-\sum_{j>i}\mu(j,i;w)\right]=\pi_i$$

$$\beta n(w,\theta_i)^{\eta-1}\left[\pi_i\lambda(w)+\sum_{j<i}\mu(i,j;w)-\sum_{j>i}\mu(j,i;w)\right]=\pi_i\frac{1}{R}v'(w'(w,\theta_i))$$

$$h'(1-l(w,\theta_i)-bn(w,\theta_i))\left[\pi_i\lambda(w)+\sum_{j<i}\mu(i,j;w)\right]$$
$$-\sum_{j>i}\mu(j,i;w)\frac{\theta_i}{\theta_j}h'(1-\frac{\theta_i l(w,\theta_i)}{\theta_j}-bn(w,\theta_i))=\pi_i\theta_i$$

$$\left\{-bh'(1-l(w,\theta_i)-bn(w,\theta_i))+\beta\eta n(w,\theta_i)^{\eta-1}w'(w,\theta_i)\right\}\left[\pi_i\lambda(w)+\sum_{j<i}\mu(i,j;w)\right]$$
$$-\sum_{j>i}\mu(j,i;w)\left\{-bh'(1-\frac{\theta_i l(w,\theta_i)}{\theta_j}-bn(w,\theta_i))+\beta\eta n(w,\theta_i)^{\eta-1}w'(w,\theta_i)\right\}=\pi_i\frac{1}{R}v(w'(w,\theta_i))$$

By Lemma 2, we must have:

$$\begin{aligned}\lim_{w\to-\infty}c(w,\theta_j) &= 0\\ \lim_{w\to-\infty}n(w,\theta_j) &= 0\\ \lim_{w\to-\infty}l(w,\theta_j) &= 1\end{aligned}$$

Then for every $\epsilon>0$, there exists W such that for all $w<W$, we have $u'(c(w,\theta_j))>\frac{N}{\epsilon}, h'(1-l(w,\theta_j)-bn(w,\theta_j))>\frac{N}{\epsilon}$. This implies that

$$\begin{aligned}\lambda(w) &= \sum_j\frac{\pi_j}{u'(c(w,\theta_j))}<\frac{\epsilon}{N}\\ \pi_n\lambda(w)+\sum_{j<n}\mu(n,j;w) &= \frac{\pi_n}{u'(c(w,\theta_n))}<\frac{\epsilon}{N}\\ \Rightarrow\mu(n,j;w) &< \frac{\epsilon}{N}\end{aligned}$$

In addition,

$$\begin{aligned}\pi_{n-1}\lambda(w)+\sum_{j<n-1}\mu(n-1,j;w)-\mu(n,n-1;w) &= \frac{\pi_{n-1}}{u'(c(w,\theta_{n-1}))}<\frac{\epsilon}{N}\\ \Rightarrow\mu(n-1,j;w) &< \frac{2\epsilon}{N}\end{aligned}$$

By an inductive argument, we have

$$\mu(i,j;w) < \frac{a_i \epsilon}{N}$$

where $a_{n-1} = 1, a_{n-2} = 2, a_{n-i} = a_{n-1} + \cdots + a_{n-i+1} + 1$. If we pick N so that $a_1 < N$, we have that

$$\mu(i,j;w) < \epsilon, \ \forall w < W$$

Moreover, by substituting first order conditions, we get

$$\begin{aligned}
\pi_i b\theta_i \geq \ & \pi_i \frac{1}{R}\eta v'(w'(w,\theta_i))w'(w,\theta_i) - \pi_i \frac{1}{R} v(w'(w,\theta_i)) \\
& = \pi_i b\theta_i - b\sum_{j>i}\left(1 - \frac{\theta_i}{\theta_j}\right)\mu(i,j)h'(1 - \frac{\theta_i l(w,\theta_i)}{\theta_j} - bn(w,\theta_i)) \\
& > \pi_i b\theta_i - b\epsilon\sum_{j>i}\left(1 - \frac{\theta_i}{\theta_j}\right)h'(1 - \frac{\theta_i l(w,\theta_i)}{\theta_j} - bn(w,\theta_i))
\end{aligned}$$

Since hours converges to 1, the term multiplied by ϵ in the above expression is bounded away from ∞. This implies that

$$\lim_{w \to -\infty} w'(w,\theta_i) = \underline{w}_i$$

where $\underline{w}_i$ satisfies [18]

$$\eta v'(\underline{w}_i)\underline{w}_i - v(\underline{w}_i) = bR\theta_i$$

■

[18]Notice that by assumption 4, $\bar{w}_i \in \mathbb{R}$, since $bR\theta_i + v(-\infty) > 0$. In addition, it can be shown that $v'(w)w \to 0$ as $w \to -\infty$.

A.8 Proof of Corollary 3

By Proposition 6, we know that

$$\lim_{w \to -\infty} w'(w, \theta_i) = \underline{w}_i$$

This implies that there exists a w_ϵ such that

$$\forall w \leq w_\epsilon, \ |w'(w, \theta_i) - \underline{w}_i| < \epsilon$$

By assumption 6, $\underline{w}_i < \bar{w}$. Now define,

$$\underline{w} = \min \left\{ w_\epsilon, \underline{w}_1 - \epsilon, \underline{w}_2 - \epsilon, \cdots, \underline{w}_n - \epsilon, \inf_{w \in [w_\epsilon, \overline{w}], i} w'(w, \theta_i) \right\}$$

Notice that since w' is a continuous function that is always in $\mathbb{R}$ and the infimum is taken over a compact set, $\underline{\mathrm{w}}$ is well-defined. Pick $\epsilon > 0$ small enough so that $\inf_{w \in [w_\epsilon, \overline{w}], i} w'(w, \theta_i) < \underline{w}_j - \epsilon$ for all j. Then by definition of w_ϵ we must have

$$w'(w, \theta_i) \in [\underline{w}, \bar{w}], \ \forall w \in [\underline{w}, \bar{w}]$$

■

Since utility is unbounded below and η is negative, $n(w, \theta_i)$ must be positive. Hence we can have the following corollary:

Corollary 6 *For all $w \in [\underline{w}, \bar{w}]$, we must have $n(w, \theta_i) \geq \underline{n}$ and $\frac{n(w, \theta_n)}{n(w, \theta_i)} \geq A$, for all $i \in \{1, \cdots, n\}$ and for some $\underline{n}, A > 0$.*

A.9 Assumption 6

Here we show that, under some additional assumptions, Assumption 6 can be shown to hold from primitives.

Suppose there are I types $\Theta = \{\theta_1, \ldots, \theta_I\}$ and $\theta_{i+1} > \theta_i$ for all $1 < i \leq I$. We break the proof into few lemmas.

Throughout this section, we make two additional assumptions:

1. $V(N, W)$ is convex;

2. If $l(\theta) = 0$ for some θ, then $l(\theta') = 0$ for all $\theta' < \theta$;

3. If all local downward incentive constraints are satisfied, all downward incentive constraints are satisfied.

Consider the following problem:

$$
\begin{aligned}
v(w) = \min_{c(\theta), l(\theta), n(\theta)} \quad & \sum_{\theta \in \Theta} \pi(\theta) \left(c(\theta) - \theta l(\theta) + \frac{1}{R} n(\theta) v(w'(\theta)) \right) \\
\text{s.t.} \quad & \sum_{\theta \in \Theta} \pi(\theta) \left(u(c(\theta)) + h(1 - l(\theta) - bn(\theta)) + \beta n(\theta)^{\eta} w'(\theta) \right) \geq w \\
& u(c(\theta_i)) + h(1 - l(\theta_i) - bn(\theta_i)) + \beta n(\theta_i)^{\eta} w'(\theta_i) \geq \\
& u(c(\theta_{i-1})) + h\left(1 - \frac{\theta_{i-1} l(\theta_{i-1})}{\theta_i} - bn(\theta_{i-1})\right) + \beta n(\theta_{i-1})^{\eta} w'(\theta_{i-1})
\end{aligned}
$$

Lemma 3 *For any w such that $l(w, \theta_I) > 0$ we have $w'(w, \theta_i) < w'(w, \theta_I)$ for all $i < I$.*

Proof. Let λ and μ_i be multipliers on promise keeping and incentive constraints. For now suppose $l(w, \theta_i) > 0$ for all i. First order conditions are (we suppress the dependence of the allocation on w, it plays no role in the following arguments):

$$\left(\lambda + \frac{\mu_I}{\pi(\theta_I)}\right) h'(1 - l(\theta_I) - bn(\theta_I)) = \theta_I \quad (36)$$

$$\left(\lambda + \frac{\mu_i}{\pi(\theta_i)}\right) h'(1 - l(\theta_i) - bn(\theta_i)) - \frac{\mu_{i+1}}{\pi(\theta_i)} \frac{\theta_i}{\theta_{i+1}} h'\left(1 - \frac{\theta_i l(\theta_i)}{\theta_{i+1}} - bn(\theta_i)\right) = \theta_i \quad (37)$$

$$v'(w'(\theta_I)) = \left(\lambda + \frac{\mu_I}{\pi(\theta_I)}\right) \beta R n(\theta_I)^{\eta - 1} \quad (38)$$

$$v'(w'(\theta_i)) = \left(\lambda + \frac{\mu_i}{\pi(\theta_i)} - \frac{\mu_{i+1}}{\pi(\theta_i)}\right) \beta R n(\theta_i)^{\eta - 1} \quad (39)$$

$$\left(\lambda + \frac{\mu_I}{\pi(\theta_I)}\right)\eta\beta R n(\theta_I)^{\eta-1} w'(\theta_I) = v(w'(\theta_I)) + Rb\left(\lambda + \frac{\mu_I}{\pi(\theta_I)}\right)h'(1 - l(\theta_I) - bn(\theta_I)) \tag{40}$$

$$\left(\lambda + \frac{\mu_i}{\pi(\theta_i)} - \frac{\mu_{i+1}}{\pi(\theta_i)}\right)\eta\beta R n(\theta_i)^{\eta-1} w'(\theta_i) = v(w'(\theta_i)) + Rb\left(\lambda + \frac{\mu_i}{\pi(\theta_i)}\right)h'(1 - l(\theta_i) - bn(\theta_i)) - \frac{\mu_{i+1}}{\pi(\theta_i)} Rbh'\left(1 - \frac{\theta_i l(\theta_i)}{\theta_{i+1}} - bn(\theta_i)\right) \tag{41}$$

for $1 < i < I$.

Combining these equations we can get the following two equations

$$\eta w'(\theta_I) v'(w'(\theta_I)) - v(w'(\theta_I)) = Rb\theta_I \tag{42}$$

$$\eta w'(\theta_i) v'(w'(\theta_i)) - v(w'(\theta_i)) = Rb\left(\lambda + \frac{\mu_i}{\pi(\theta_i)}\right)h'(1 - l(\theta_i) - bn(\theta_i)) - \frac{\mu_{i+1}}{\pi(\theta_i)} Rbh'\left(1 - \frac{\theta_i l(\theta_i)}{\theta_{i+1}} - bn(\theta_i)\right) \tag{43}$$

Since $\eta wv'(w) - v(w)$ is increasing in w, to establish the claim of the lemma it is enough to show that the right hand side of the equation (43) is smaller that $Rb\theta_I$. But notice that

$$\begin{aligned} \left(\lambda + \frac{\mu_i}{\pi(\theta_i)}\right)h'(1 - l(\theta_i) - bn(\theta_i)) - \frac{\mu_{i+1}}{\pi(\theta_i)} h'\left(1 - \frac{\theta_i l(\theta_i)}{\theta_{i+1}} - bn(\theta_i)\right) &< \\ \left(\lambda + \frac{\mu_i}{\pi(\theta_i)}\right)h'(1 - l(\theta_i) - bn(\theta_i)) - \frac{\mu_{i+1}}{\pi(\theta_i)} \frac{\theta_i}{\theta_{i+1}} h'\left(1 - \frac{\theta_i l(\theta_i)}{\theta_{i+1}} - bn(\theta_i)\right) &= \theta_i \\ &< \theta_I \end{aligned}$$

And this finishes the proof for the case in which $l(w, \theta_i) > 0$ for all i.

Now consider the case in which the non-negativity constraint on hours is binding for some types. Let $1 \leq j < I$ and suppose for all types θ_i ,$1 \leq i \leq j$ we have $l(w, \theta_i) = 0$, and for all types θ_i, $i > j$ we have $l(w, \theta_i) => 0$.Then all types θ_i, $1 \leq i \leq j$ receive the same allocations and therefore $\mu_i = 0$ for $1 \leq i \leq j$. The equations (43) and (37) for type θ_j change to

$$\eta w'(\theta_j) v'(w'(\theta_j)) - v(w'(\theta_j)) = Rb\left(\lambda - \frac{\mu_{j+1}}{\pi(\theta_j)}\right)h'(1 - bn(\theta_j))$$

and

$$\left(\lambda - \frac{\mu_{j+1}}{\pi(\theta_j)} \frac{\theta_j}{\theta_{j+1}}\right) h'(1 - bn(\theta_i)) > \theta_i$$

Suppose $w'(\theta_j) > w'(\theta_I)$, then

$$Rb\left(\lambda - \frac{\mu_{i+1}}{\pi(\theta_j)}\right) h'(1 - bn(\theta_j)) = \eta w'(\theta_j)v'(w'(\theta_j)) - v(w'(\theta_j)) > Rb\theta_I$$

and therefore

$$\begin{aligned} h'(1 - bn(\theta_j)) &> \frac{\theta_I}{\lambda - \frac{\mu_{j+1}}{\pi(\theta_j)}} \\ &> \frac{\theta_I}{\lambda + \frac{\mu_I}{\pi(\theta_I)}} = h'(1 - bn(\theta_I) - l(\theta_I)) \end{aligned}$$

Hence

$$1 - bn(\theta_j) < 1 - l(\theta_I) - bn(\theta_I)$$

On the other hand $w'(\theta_j) > w'(\theta_I)$ implies

$$\left(\lambda - \frac{\mu_{i+1}}{\pi(\theta_j)}\right) n(\theta_j)^{\eta-1} = \frac{v'(w'(\theta_j))}{\beta R} > \frac{v'(w'(\theta_I))}{\beta R} = \left(\lambda + \frac{\mu_I}{\pi(\theta_I)}\right) n(\theta_I)^{\eta-1}$$

and therefore

$$n(\theta_I) > n(\theta_j)$$

Then $l(\theta_I)$ has to be negative which is contradiction. Therefore, we must have $w'(w, \theta_j) < w'(w, \theta_I)$ for all w such that $l(w, \theta_I) > 0$. Since $l(w, \theta_i) = 0$ for all $i < j$, we also know that $w'(w, \theta_i) = w'(w, \theta_j) < w'(w, \theta_I)$ for all $i < j$. ■

Next we will find the promised utility at which the non-negativity for the the type θ_I just binds. At this point, type θ_L also works zeros hours and therefore both types receive the same allocations.

Let $\hat{c}$ and $\hat{n}$ be the solution to the following equations

$$\begin{aligned} \theta_I u'(\hat{c}) &= h'(1 - b\hat{n}) \\ v'(w_I) &= \frac{\beta R \hat{n}^{\eta-1}}{u'(\hat{c})} \end{aligned}$$

in which w_H is the solution to the following equation

$$\eta w_I v'(w_I) - v(w_I) = Rb\theta_I$$

Define

$$\hat{w} = u(\hat{c}) + h(1 - b\hat{n}) + \beta \hat{n}^{\eta} w_I$$

Note that $w'(\hat{w}, \theta_I) = w'(\hat{w}, \theta_i) = w_I$ for all i.

Next we show that for $w > \hat{w}$, both types work zero hours and $w'(w, \theta) > w_I$ for all θ. Then we prove the claim of the proposition for two cases on $w_I > \hat{w}$ and $w_I < \hat{w}$.

Lemma 4 *If $w > \hat{w}$, then $l(w, \theta_I) = 0$.*

Proof. Suppose otherwise and consider the following equations

$$w = u(c) + h(m) + \beta n^{\eta} w_I$$

and

$$\begin{aligned} \theta_H u'(c) &= h'(m) \\ v'(w_H) &= \frac{\beta R n^{\eta-1}}{u'(c)} \end{aligned}$$

in which $m = 1 - l - bn$. Note that

$$u'(c)\frac{\partial c}{\partial w} - h'(m)\frac{\partial m}{\partial w} + \beta w_H \eta n^{\eta-1}\frac{\partial n}{\partial w} = 1$$

$$\theta_H u''(c)\frac{\partial c}{\partial w} = h''(m)\frac{\partial m}{\partial w}$$

$$v'(w_H)u''(c)\frac{\partial c}{\partial w} = \beta R\eta - 1 n^{\eta-2}\frac{\partial n}{\partial w}$$

The last two equations imply that $\frac{\partial c}{\partial w}$, $\frac{\partial n}{\partial w}$ and $\frac{\partial m}{\partial w}$ must have the same sign. The only way that this can be consistent with the first equation is that all have positive sign. If $\frac{\partial m}{\partial w} > 0$ and $\frac{\partial n}{\partial w} > 0$, then we must have $\frac{\partial l}{\partial w} < 0$. Evaluating this at $w = \hat{w}$ implies that $l(w, \theta) < 0$ for $w > \hat{w}$. This implies that for $w > \hat{w}$ the non-negativity binds. ■

Next, we show that if $w_I > \hat{w}$ the claim in the proposition is satisfied.

Lemma 5 *If $w_I > \hat{w}$, then there exist $\hat{w} \leq w^* \leq 0$ such that $w'(w^*, \theta) = w^*$.*

Proof. Recall that that since $l(w, \theta) \geq 0$ is binding, both types work zero and receive the same allocations. Therefore, the incentive constraint is slack. The first order conditions for type θ_I are

$$\lambda h'(1 - bn(\theta_I)) > \theta_I$$

and

$$v(w'(\theta_I)) + Rb\lambda h'(1 - bn(\theta_I)) = \lambda\eta\beta Rn(\theta_I)^{\eta-1}w'(\theta_I)$$

therefore

$$\eta w'(\theta_I)v'(w'(\theta_I)) - v(w'(\theta_I)) = Rb\lambda h'(1 - bn(\theta_I)) > Rb\theta_I$$

This implies, $w'(w, \theta_H) > w_H > \hat{w}$. Define the function $w'_\epsilon(\cdot, \theta) : [\hat{w}, -\epsilon] \to [\hat{w}, -\epsilon]$ as

$$w'_\epsilon(\cdot, \theta) = \begin{cases} w'(\cdot, \theta) & \text{if } w'_\epsilon(\cdot, \theta) \leq -\epsilon \\ -\epsilon & \text{if } w'_\epsilon(\cdot, \theta) > -\epsilon \end{cases}$$

This function must have a fixed point $w^*_\epsilon \in [\hat{w}, -\epsilon]$. We know that $w'(\cdot, \theta) = \lim_{\epsilon \to 0} w'_\epsilon(\cdot, \theta)$. Then, either a $\hat{w} < w^* < 0$ exists such that $w'(w^*, \theta) = w^*$ or $\lim_{w \to 0} w'(w, \theta) = 0$. (Note that because no one works all types receive the same allocations). ■

So far we have established that if $w_I > \hat{w}$, then we can choose $\underline{w} = \bar{w} = w^*$ and the proposition is proved.

Now suppose $w_I \leq \hat{w}$. Then, by Lemma 3, $w(w, \theta) \leq w_I \leq \hat{w}$ for all w. Let $\bar{w} = \hat{w}$. Then, using this along with Corollary 3, it follows that $w'(w, \theta) \in [\underline{w}, \bar{w}]$ for any $w \in [\underline{w}, \bar{w}]$, that is w' maps the compact set $[\underline{w}, \bar{w}]$ into itself – i.e., Assumption 6 is satisfied.

A.10 Proof of Remark 5

Since $l(w, \theta_n) > 0$ for all $w \in [\underline{w}, w_0]$, resetting property at the top holds. Therefore, by definition

$$
\begin{aligned}
E_\Psi(n) \cdot \Psi(\{w_0\}) &= \pi_n \int_S n(w, \theta_n) d\Psi(w) \\
E_\Psi(n) &= \int_S \sum_{i=1}^{n} \pi_i n(w, \theta_i) d\Psi(w) \\
&\leq \pi_n \int_S n(w, \theta_n) d\Psi(w) \\
&\quad + (1 - \pi_n) A^{-1} \int_S n(w, \theta_n) d\Psi(w) \\
&= \left(\pi_n + (1 - \pi_n) A^{-1}\right) \int_S n(w, \theta_n) d\Psi(w)
\end{aligned}
$$

Therefore,

$$
\Psi(\{w_0\}) \geq \frac{\pi_n A}{1 - \pi_n + \pi_n A}
$$

■

A.11 Implementation

A.11.1 Distortions

The constrained efficient allocation $c_1^*(\theta), l^*(\theta), n^*(\theta)$ solves the following problem

$$\begin{aligned} \max \quad & \sum_{i=H,L} \pi_i \left[u(c_i) + h(1 - l_i - bn_i) + \beta n_i^\eta u(c_2) \right] \\ \text{s.t.} \quad & \sum_{i=H,L} \pi_i \left[c_i + \frac{1}{R} n_i c_2 \right] \le \sum_{i=H,L} \pi_i \theta_i l_i + RK_0 \\ & u(c_H) + h(1 - l_H - bn_H) + \beta n_H^\eta u(c_2) \ge \\ & \qquad u(c_L) + h(1 - \frac{\theta_L l_L}{\theta_H} - bn_L) + \beta n_L^\eta u(c_2). \end{aligned}$$

Assuming it is interior, it satisfies the following first order conditions:

$$\begin{aligned} u'(c_H^*)(\pi_H + \mu) &= \lambda \pi_H \\ u'(c_L^*)(\pi_L - \mu) &= \lambda \pi_L \\ h'(1 - l_H^* - bn_H^*)(\pi_H + \mu) &= \lambda \theta_H \pi_H \\ h'(1 - l_L^* - bn_L^*)\pi_L - \mu \frac{\theta_L}{\theta_H} h'(1 - \frac{\theta_L l_L^*}{\theta_H} - bn_L^*) &= \lambda \theta_L \pi_L \\ \left[-bh'(1 - l_H^* - bn_H^*) + \beta \eta n_H^{*\eta - 1} u(c_2) \right] (\pi_H + \mu) &= \lambda \frac{1}{R} \pi_H c_2 \\ & \left[-bh'(1 - l_L^* - bn_L^*) + \beta \eta n_L^{*\eta - 1} u(c_2) \right] \pi_L \\ - \left[-bh'(1 - \frac{\theta_L l_L^*}{\theta_H} - bn_L^*) + \beta \eta n_L^{*\eta - 1} u(c_2) \right] \mu &= \lambda \frac{1}{R} \pi_L c_2 \end{aligned}$$

Now suppose that we want to implement the above allocation with a tax in first period of the form $T(y, n)$. Then consumer's problem is the following:

$$\begin{aligned} \max \quad & u(c_1) + h(1 - l - bn) + \beta n^\eta u(c_2) \\ \text{s.t.} \quad & c_1 + k_1 \le Rk_0 + \theta l - T(\theta l, n) \\ & nc_2 \le Rk_1 \end{aligned}$$

As a first step, we assume that T is differentiable and that y is interior for both types. Then the FOCs are the following:

$$
\begin{aligned}
u'(c_1) &= \lambda_1 \\
h'(1-l-bn) &= \lambda_1\theta(1-T_y(\theta l, n)) \\
R\lambda_1 &= \lambda_2 \\
-bh'(1-l-bn)+\beta\eta n^{\eta-1}u(c_2) &= \lambda_2 c_2 + \lambda_1 T_n(\theta l, n)
\end{aligned}
$$

Comparing the FOC's for the planner with these, we see immediately that $T_n(\theta_H l_H^*, n_H^*) = 0, T_y(\theta_H l_H^*, n_H^*) = 0$ – there are no (marginal) distortions on the decisions of the agent with the high shock. Moreover, from the FOC's of the planner's problem we get:

$$
\begin{aligned}
&\left[-bh'(1-l_L^*-bn_L^*)\pi_L + bh'(1-\frac{\theta_L l_L^*}{\theta_H}-bn_L^*)\mu\right]\frac{1}{\pi_L-\mu} \\
+\beta\eta n_L^{*\eta-1}u(c_2) \quad = \quad &\frac{1}{R}u'(c_1^*)c_2.
\end{aligned}
$$

We know that

$$
\begin{aligned}
1-\frac{\theta_L l_L^*}{\theta_H}-bn_L^* &> 1-l_L^*-bn_L^* \\
\Rightarrow h'(1-\frac{\theta_L l_L^*}{\theta_H}-bn_L^*)\mu &< h'(1-l_L^*-bn_L^*)\mu \\
bh'(1-\frac{\theta_L l_L^*}{\theta_H}-bn_L^*)\mu - bh'(1-l_L^*-bn_L^*)\pi_L &< bh'(1-l_L^*-bn_L^*)\mu - bh'(1-l_L^*-bn_L^*)\pi_L \\
\left[-bh'(1-l_L^*-bn_L^*)\pi_L + bh'(1-\frac{\theta_L l_L^*}{\theta_H}-bn_L^*)\mu\right]\frac{1}{\pi_L-\mu} &< -bh'(1-l_L^*-bn_L^*).
\end{aligned}
$$

Hence,

$$
\frac{1}{R}u'(c_1^*)c_2 - \beta\eta n_L^{*\eta-1}u(c_2) < -bh'(1-l_L^*-bn_L^*).
$$

By FOC of the consumer problem,we have

$$
0 < T_n(\theta_L l_L^*, n_L^*) = -bh'(1-l_L^*-bn_L^*) - \frac{1}{R}u'(c_1^*)c_2 + \beta n_L^{*\eta-1}u(c_2).
$$

Finally, we turn to $T_y(\theta_L l_L^*, n_L^*)$. From above, we have:

$$
\begin{aligned}
h'(1-l_L^*-bn_L^*)\pi_L-\mu\frac{\theta_L}{\theta_H}h'(1-\frac{\theta_L l_L^*}{\theta_H}-bn_L^*) &= \lambda\theta_L\pi_L \\
h'(1-l_L^*-bn_L^*)\left[\pi_L-\mu\frac{\theta_L}{\theta_H}\right] &< \lambda\theta_L\pi_L \\
h'(1-l_L^*-bn_L^*)\left[\pi_L-\mu\frac{\theta_L}{\theta_H}\right] &< \theta_L u'(c_L^*)(\pi_L-\mu) \\
h'(1-l_L^*-bn_L^*) &< \theta_L u'(c_L^*)\frac{(\pi_L-\mu)}{\left[\pi_L-\mu\frac{\theta_L}{\theta_H}\right]} \\
h'(1-l_L^*-bn_L^*) &< \theta_L u'(c_L^*)
\end{aligned}
$$

Thus, from the FOC's of the agent's problem, we see that $T_y(\theta_L l_L^*, n_L^*) > 0$.

A.11.2 Defining of Tax Function

Begin by defining the levels of utilities for the two types that are obtained at the socially efficient allocation:

$$
\bar{u}_L = u(c_{1L}^*) + h(1-\frac{y_L^*}{\theta_L}-bn_L^*) + \beta n_L^{*\eta}u(c_2);
$$

$$
\bar{u}_H = u(c_{1H}^*) + h(1-\frac{y_H^*}{\theta_H}-bn_H^*) + \beta n_H^{*\eta}u(c_2);
$$

We define two versions of the tax function. The first, T_L, is designed to make sure that the low type always gets utility $\bar{u}_L$ if they satisfy their budget constraint with equality. The second, T_H, is defined similarly:

$$
\bar{u}_L = u(y-T_L(y,n)-\frac{1}{R}nc_2) + h(1-\frac{y}{\theta_L}-bn) + \beta n^{\eta}u(c_2);
$$

$$
\bar{u}_H = u(y-T_H(y,n)-\frac{1}{R}nc_2) + h(1-\frac{y}{\theta_H}-bn) + \beta n^{\eta}u(c_2);
$$

It can be shown that the locus of the points $T_H(y,n) = T_L(y,n)$ is downward sloping in (y,n) space.

Next we show that $T_L(y_L^*, n_L^*) = T_H(y_L^*, n_L^*)$. We know that at the constrained efficient allocation, type θ_H is indifferent between the allocations $(c_{1H}^*, y_H^*, n_H^*, c_2^*)$ and $(c_{1L}^*, y_L^*, n_L^*, c_2^*)$.

Hence we have the following equality:

$$\bar{u}_H = u(c^*_{1H}) + h(1 - \frac{y^*_H}{\theta_H} - bn^*_H) + \beta n^\eta_H u(c_2) = u(c^*_{1L}) + h(1 - \frac{y^*_L}{\theta_H} - bn^*_L) + \beta n^{*\eta}_L u(c_2)$$

replace for c^*_{1H} and c^*_{1L} from budget constraints, we also get

$$\begin{aligned} \bar{u}_H &= u(y^*_H - T_H(y^*_H, n^*_H) - \frac{1}{R} n^*_H c_2) + h(1 - \frac{y^*_H}{\theta_H} - bn^*_H) + \beta n^\eta_H u(c_2) \\ &= u(y^*_L - T_L(y^*_L, n^*_L) - \frac{1}{R} n^*_L c_2) + h(1 - \frac{y^*_L}{\theta_H} - bn^*_L) + \beta n^{*\eta}_L u(c_2) \end{aligned}$$

Moreover, from the definition of T_H we know that

$$\bar{u}_H = u(y^*_L - T_H(y^*_L, n^*_L) - \frac{1}{R} n^*_L c_2) + h(1 - \frac{y^*_L}{\theta_H} - bn^*_L) + \beta n^{*\eta}_L u(c_2)$$

Hence, the last two equalities imply that $T_L(y^*_L, n^*_L) = T_H(y^*_L, n^*_L)$.

We can also show that $T_H(y^*_H, n^*_H) > T_L(y^*_H, n^*_H)$. This follows from the fact that type θ_L strictly prefers the allocation $(c^*_{1L}, y^*_L, n^*_L, c^*_2)$ to $(c^*_{1H}, y^*_H, n^*_H, c^*_2)$, i.e.:

$$\begin{aligned} \bar{u}_L &= u(c^*_{1L}) + h(1 - \frac{y^*_L}{\theta_L} - bn^*_L) + \beta n^{*\eta}_L u(c_2) \\ &> u(c^*_{1H}) + h(1 - \frac{y^*_H}{\theta_L} - bn^*_H) + \beta n^{*\eta}_H u(c_2) \end{aligned}$$

replace for c^*_{1H} and c^*_{1L} from budget constraints, we also get

$$\begin{aligned} \bar{u}_L &= u(y^*_L - T_L(y^*_L, n^*_L) - \frac{1}{R} n^*_L c_2) + h(1 - \frac{y^*_L}{\theta_L} - bn^*_L) + \beta n^{*\eta}_L u(c_2) \\ &> u(y^*_H - T_H(y^*_H, n^*_H) - \frac{1}{R} n^*_H c_2) + h(1 - \frac{y^*_H}{\theta_L} - bn^*_H) + \beta n^{*\eta}_H u(c_2) \end{aligned}$$

By definition of $T_L$we have

$$\bar{u}_L = u(y^*_H - T_L(y^*_H, n^*_H) - \frac{1}{R} n^*_H c_2) + h(1 - \frac{y^*_H}{\theta_L} - bn^*_H) + \beta n^{*\eta}_H u(c_2)$$

Hence, we have that $T_H(y^*_H, n^*_H) > T_L(y^*_H, n^*_H)$.

A.11.3 Proof of Proposition 7

Given the tax funtion, the consumer's problem is the following:

$$\begin{aligned} \max_{c_1,y,n} \quad & u(c_1) + h(1 - \frac{y}{\theta} - bn) + \beta n^\eta u(c_2) \\ & \text{s.t. } c_1 + \frac{1}{R} n c_2 \leq y - T(y,n) \end{aligned}$$

We know that $T(y_H^*, n_H^*) = T_H(y_H^*, n_H^*)$. Hence, type θ_H can afford (c_{1H}^*, y_H^*, n_H^*) and $u(c_{1H}^*, y_H^*, n_H^*; \theta_H) = \bar{u}_H$. Let (c_1, y, n) be any allocation that satisfy $c_1 + \frac{1}{R} c_2 = y - T(y,n)$. Then,

$$c_1 + \frac{1}{R} n c_2 = y - \max\{T_L(y,n), T_H(y,n)\} \leq y - T_H(y,n)$$

But by definition of T_H, $u(c_1, y, n)$ can be at most $\bar{u}_H$.

Using similar argument we can show that type θ_H can afford (c_{1L}^*, y_L^*, n_L^*) and $u(c_{1L}^*, y_L^*, n_L^*; \theta_L) = \bar{u}_L$. Moreover, any allocation that satisfy the budget constraint has utility at most $\bar{u}_L$.

The differnitbaility of T and properties of marginal taxes follows form the discussion above.

B Linear Utility of Leisure

The problem becomes the following - $h(m) = \psi m$:[19]

$$
\begin{aligned}
V(N,W) = \min_{C_i, L_i, N_i, W_i'} \quad & \sum_{i=1}^{n} \pi(\theta_i) \left[C_i - \theta_i L_i + \frac{1}{R} V(N_i, W_i') \right] \qquad (P5) \\
\text{s.t} \quad & \sum_{i=1}^{n} \pi(\theta_i) \left[N^\eta \left(u\left(\frac{C_i}{N}\right) + \psi \left(1 - \frac{L_i}{N} - b\frac{N_i}{N} \right) \right) + \beta W_i' \right] \geq W \\
& N^\eta \left(u\left(\frac{C_i}{N}\right) + \psi \left(1 - \frac{L_i}{N} - b\frac{N_i'}{N} \right) \right) + \beta W_i' \geq \\
& \quad N^\eta \left(u\left(\frac{C_j}{N}\right) + \psi \left(1 - \frac{\theta_j L_j}{\theta_i N} - b\frac{N_j'}{N} \right) \right) + \beta W_j' \\
& \quad \forall i, j.
\end{aligned}
$$

Notice that the set of reports is not restricted to lower reports since we can prove that general incentive compatibility is equivalent to local downward constraints being binding and output being increasing, a similar approach to Thomas and Worrall (1990). Notice that adding the IC constraint where j pretends to be i and the reverse implies that:

$$
\frac{\theta_i L_i}{\theta_j N} + \frac{\theta_j L_j}{\theta_i N} \geq \frac{L_i}{N} + \frac{L_j}{N}
$$

Therefore, if $\theta_i > \theta_j$, then $\theta_i L_i \geq \theta_j L_j$ which mean output is increasing.
Moreover, if we assume that local downward IC constraints are binding and output is increasing, it can be easily shown that the local upward constraints are satisfied and summing over local incentive constraints gives the general ones. We also assume that output being increasing is not binding so we can neglect it. Therefore the functional equation becomes the following:

$$
\begin{aligned}
V(N,W) = \min_{C_i, L_i, N_i', W_i'} \quad & \sum_{i=1}^{n} \pi(\theta_i) \left[C_i - \theta_i L_i + \frac{1}{R} V(N_i', W_i') \right] \\
\text{s.t} \quad & \sum_{i=1}^{n} \pi(\theta_i) \left[N^\eta \left(u\left(\frac{C_i}{N}\right) + \psi \left(1 - \frac{L_i}{N} - b\frac{N_i'}{N} \right) \right) + \beta W_i' \right] \geq W \\
& N^\eta \left(u\left(\frac{C_i}{N}\right) + \psi \left(1 - \frac{L_i}{N} - b\frac{N_i'}{N} \right) \right) + \beta W_i' \geq \\
& \quad N^\eta \left(u\left(\frac{C_{i-1}}{N}\right) + \psi \left(1 - \frac{\theta_{i-1} L_{i-1}}{\theta_i N} - b\frac{N_{i-1}'}{N} \right) \right) + \beta W_{i-1}'
\end{aligned}
$$

[19] Assume that $\Theta = \{\theta_1 < \theta_2 < ... < \theta_n\}$. We will index allocations by i instead of θ.

Let $-\lambda N^{1-\eta}$ be the lagrange multiplier on promise-keeping constraint and $-\mu_i N^{1-\eta}$ be the multiplier for i-th IC constraint. Then the first order condition for hours worked is the following:

$$\pi_n\theta_n = (\lambda\pi_n + \mu_n)\psi$$

$$\pi_i\theta_i = (\lambda\pi_i + \mu_i - \frac{\theta_i\mu_{i+1}}{\theta_{i+1}})\psi, \quad i = 2, \cdots, n \tag{44}$$

$$\pi_1\theta_1 = (\lambda\pi_1 - \frac{\theta_1\mu_2}{\theta_2})\psi \tag{45}$$

We can define $\mu_1 = \mu_{n+1} = 0$ and (44) holds for $i = 1, ..., n$. If we divide the i-th equation by θ_i and sum over all i's, the μ_i's will cancel and we have

$$\lambda = \frac{1}{\psi\sum_i \frac{\pi_i}{\theta_i}} = \frac{1}{\psi\,\mathbb{E}\,\frac{1}{\theta}} \tag{46}$$

Therefore,

$$\mu_i = \frac{1}{\psi}\theta_i\sum_{j\geq i}\pi_j - \frac{\theta_i\sum_{j\geq i}\frac{\pi_j}{\theta_j}}{\psi\sum_j\frac{\pi_j}{\theta_j}} = \theta_i\frac{\sum_j\frac{\pi_j}{\theta_j}\sum_{j\geq i}\pi_j - \sum_{j\geq i}\frac{\pi_j}{\theta_j}}{\psi\sum_j\frac{\pi_j}{\theta_j}}$$

Since θ_i's increasing, all the μ_i's are positive.
The first order conditions with respect to consumption are:

$$\pi_i = (\lambda\pi_i + \mu_i - \mu_{i+1})u'\left(\frac{C_i}{N}\right)$$

Obviously, we need consumption to be increasing as well as marginal utility to be positive. This gives us a condition on distribution of θ_i. Moreover, we can see that consumption is independent of state variable (N, W).
The first order conditions with respect to N_i', W_i' are the following:

$$\pi_i\frac{1}{R}V_N(N_i', W_i') = -b(\lambda\pi_i + \mu_i - \mu_{i+1})\psi \tag{47}$$

$$\pi_i\frac{1}{R}V_W(N_i', W_i') = N^{1-\eta}(\lambda\pi_i + \mu_i - \mu_{i+1})\beta \tag{48}$$

Now for every i, define *after-tax-productivity* as follows:

$$\tilde{\theta}_i = \psi\frac{\lambda\pi_i + \mu_i - \mu_{i+1}}{\pi_i}$$

Notice that we have $u'(C_i/N)\tilde{\theta}_i = \psi$ and $\tilde{\theta}_i$ does not depend on the state variables. From before, we know that there exists a function $v(\cdot)$ such that $V(N, W) = Nv(N^{-\eta}W)$. Therefore,

$$V_N(N,W) = v(w) - \eta w v'(w), V_W(N,W) = N^{1-\eta} v'(w)$$

where $w = N^{-\eta} W$. Hence, from (48) we have that:

$$\eta w_i' v'(w_i') - v(w_i') = bR\tilde{\theta}_i$$
$$N_i'^{1-\eta} v'(w_i) = \beta R N^{1-\eta} \tilde{\theta}_i$$

The above, implies that $n_i = N_i'/N, w_i'$ are also independent of the state. Moreover, from the Envelope condition we have that:

$$V_W(N,W) = \lambda N^{1-\eta} = \frac{N^{1-\eta}}{\psi \sum_i \frac{\pi_i}{\theta_i}} = N^{1-\eta} v'(w)$$

Therefore, $v(\cdot)$ is a linear function and we must:

$$v(w) = A + \frac{w}{\psi \sum_i \frac{\pi_i}{\theta_i}} \Rightarrow V(N,W) = AN + \frac{WN^{1-\eta}}{\psi \sum_i \frac{\pi_i}{\theta_i}}$$

Notice that to satisfy assumption 2, we need $N^\eta h(\frac{M}{N})$ to be concave and therefore, to have $N^{\eta-1}M$ be weakly concave, we must have $\eta = 1$. In this case $V(N,W)$ is linear in (N,W) and therefore weakly convex.

References

Albanesi, S. and C. Sleet (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73, 1–30.

Alvarez, F. (1999): "Social Mobility: The Barro–Becker Children Meet the Laitner–Loury Dynasties," *Review of Economic Dynamics*, 2, 65–103.

Atkeson, A. and R. Lucas Jr. (1992): "On Efficient Distribution with Private Information," *Review of Economic Studies*, 59, 427–53.

——— (1995): "Efficiency and Equality in a Simple Model of Efficient Unemployment Insurance," *Journal of Economic Theory*, 66, 64–88.

Barro, R. J. (1974): "Are government bonds net wealth?" *Journal of Political Economy*, 82, 1095.

Barro, R. J. and G. S. Becker (1989): "Fertility Choice in an Endogenous Growth Model," *Econometrica*, 57, 481–501.

Becker, G. S. and R. J. Barro (1988): "A Reformulation of the Economic Theory of Fertility," *The Quarterly Journal of Economics*, 103, 1–25.

Doepke, M. and R. M. Townsend (2006): "Dynamic mechanism design with hidden income and hidden actions," *Journal of Economic Theory*, 126, 235–285.

Dunford, N. and J. Schwartz (1958): *Linear Operators*, Interscience publishers.

Farhi, E. and I. Werning (2007): "Inequality and Social Discounting," *Journal of Political Economy*, 115, 365–402.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70, 569–587.

Golosov, M. and A. Tsyvinski (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114, 257–279.

——— (2007): "Optimal Taxation With Endogenous Insurance Markets," *The Quarterly Journal of Economics*, 122, 487–534.

Green, E. J. (1987): "Lending and the Smoothing of Uninsurable Income," *Contractual Arrangements for Intertemporal Trade*, 3–25.

Jones, L. E., A. Schoonbroodt, and M. Tertilt (2008): "Fertility Theories: Can They Explain the Negative Fertility-Income Relationship?" in *Demography and the Economy*, National Bureau of Economic Research, Inc.

Jones, L. E. and M. Tertilt (2008): "An Economic History of Fertility in the U.S.: 1826-1960," *Frontiers in Family Economics*, edited by Peter Rupert.

Khan, A. and B. Ravikumar (2001): "Growth and risk-sharing with private information," *Journal of Monetary Economics*, 47, 499–521.

Kocherlakota, N. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73, 1587–1621.

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.

Phelan, C. (1998): "On the Long Run Implications of Repeated Moral Hazard," *Journal of Economic Theory*, 79, 174–191.

——— (2006): "Opportunity and Social Mobility," *Review of Economic Studies*, 73, 487–504.

Phelan, C. and R. M. Townsend (1991): "Computing Multi-Period, Information-Constrained Optima," *The Review of Economic Studies*, 58, 853–881.

Rogerson, W. (1985): "Repeated Moral Hazard," *Econometrica*, 53, 69–76.

Rudin, W. (1991): *Functional analysis International series in pure and applied mathematics*, McGraw-Hill.

Sleet, C. and Ş. Yeltekin (2006): "Credibility and endogenous societal discounting," *Review of Economic Dynamics*, 9, 410–437.

Thomas, J. and T. Worrall (1990): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-agent Problem," *Journal of Economic Theory*, 51, 367–390.