Summer 2012

Essays on income taxation and idiosyncratic risk.

Martin Eduardo Lopez Daneri
University of Iowa

ESSAYS ON INCOME TAXATION AND IDIOSYNCRATIC RISK

by

Martin Eduardo Lopez Daneri

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Economics
in the Graduate College of
The University of Iowa

July 2012

Thesis Supervisor: Professor Gustavo Jaime Ventura

ABSTRACT

I study the role of heterogeneity and idiosyncratic risk in Macroeconomics, and their implications on problems of income taxation. In the first chapter, I study the effects of redistributive taxation in an incomplete market economy with heterogeneous agents and idiosyncratic risk. I focus on the role of distortions in labor supply decisions and the interplay of heterogeneity and uninsurable idiosyncratic shocks, conducting the first general equilibrium analysis of a Negative Income Tax (NIT). I show that a NIT is a serious candidate to replace the current income tax in the United States. I find that the optimal NIT has a marginal tax rate of 28% and a transfer of 10% of per capita GDP, roughly $4600.

The welfare gains of replacing the current US income tax with a NIT are equivalent to a 6.3% increase in annual consumption in every state of the world. Low-ability agents, in the bottom quintile of the productivity distribution, benefit the most, while high-ability agents are worse off. A consequence of the reform is that the composition of the labor force changes, with high-productivity agents working more, in relative terms, than low-productivity agents. Finally, I find that the riskier the economy, the higher the welfare gains of the NIT as a provider of public insurance.

In the second chapter, I study labor income dynamics over the life cycle and introduce a novel methodology that can detect the presence of patterns in the idiosyncratic earnings shocks and recognize economic forces in action. Using a sample from the Panel Study of Income Dynamics (PSID), I estimate a Bayesian Logistic

Smoothed Transition Autoregressive model of order 1 (LSTAR(1)) with a rich level of heterogeneity in the innovations. I find that there is a life-cycle pattern in the earning shocks: before the age 29, young workers experience shocks with higher variance and a positive probability of lower persistence than older workers. A comparison with conventional models shows that an incorrect model specification introduces bias in the estimates. The proposed model can be easily approximated with a discrete Markov process. This means that this model can be used by macroeconomists to calibrate income processes.

Abstract Approved: ______________________________
Thesis Supervisor

____
Title and Department

____
Date

ESSAYS ON INCOME TAXATION AND IDIOSYNCRATIC RISK

by

Martin Eduardo Lopez Daneri

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Economics
in the Graduate College of
The University of Iowa

July 2012

Thesis Supervisor: Professor Gustavo Jaime Ventura

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Martin Eduardo Lopez Daneri

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Economics at the July 2012 graduation.

Thesis Committee: Gustavo Ventura, Thesis Supervisor

Antonio F. Galvao

Srihari Govindan

Guillaume Vandenbroucke

Charles H. Whiteman

To my loving parents and dear family

All theories are valid and none is relevant.
What makes them relevant is what you do with them.

–Jorge Luis Borges, Complete Works

ACKNOWLEDGEMENTS

An exhaustive list of all of the people who contributed to this dissertation will never be complete, no matter how hard I try. There are many friends and professors to whom I owe a great deal of gratitude.

I will start with Gustavo Ventura, my main advisor and first macroeconomics professor at Iowa. He taught and pushed me to be a better economist, always curious and in search of good questions, fearless of the difficulty of any economic problem. He has been a major role model for me, and an outstanding mentor.

Charles Whiteman has also been an inspiration for me. He spent time with me, correcting my thinking, encouraging me to try new approaches and, above all, showing me that “you need a model to beat a model”.

The members of my Dissertation Committee have also been generous and helpful. Antonio Galvao, Srihari Govindan and Guillaume Vandenbroucke: thanks to them for giving me so much and asking nothing in return. They prove that economists are generous and disinterested souls. I hope to someday become as good an economist as each of them are.

My thanks also goes to the Economics Department at The University of Iowa and all its past and present members. Iowa has been a stimulating and wonderful place to do macroeconomics. One person deserves special mention: Renea Jay. Her cheerful enthusiasm and hard work has been a breath of fresh air all of these years.

Then there are my very good friends German Cubas, Alicia Devoto, Elisa

Keller, Yu-Chien Kong, Gagan Ghosh and Steven Stong. Not only did we share economics, but also beautiful and meaningful times together that I will always treasure. Thank you for letting me be your friend and colleague.

I thank Ana Cortes for her love and support, and for always being there for me. I hope to live up to her expectations, and never let her down.

Finally, thanks to my parents and family for their support and immense love and for showing me from the very beginning the importance of honesty and hard work.

ABSTRACT

I study the role of heterogeneity and idiosyncratic risk in Macroeconomics, and their implications on problems of income taxation. In the first chapter, I study the effects of redistributive taxation in an incomplete market economy with heterogeneous agents and idiosyncratic risk. I focus on the role of distortions in labor supply decisions and the interplay of heterogeneity and uninsurable idiosyncratic shocks, conducting the first general equilibrium analysis of a Negative Income Tax (NIT). I show that a NIT is a serious candidate to replace the current income tax in the United States. I find that the optimal NIT has a marginal tax rate of 28% and a transfer of 10% of per capita GDP, roughly $4600.

The welfare gains of replacing the current US income tax with a NIT are equivalent to a 6.3% increase in annual consumption in every state of the world. Low-ability agents, in the bottom quintile of the productivity distribution, benefit the most, while high-ability agents are worse off. A consequence of the reform is that the composition of the labor force changes, with high-productivity agents working more, in relative terms, than low-productivity agents. Finally, I find that the riskier the economy, the higher the welfare gains of the NIT as a provider of public insurance.

In the second chapter, I study labor income dynamics over the life cycle and introduce a novel methodology that can detect the presence of patterns in the idiosyncratic earnings shocks and recognize economic forces in action. Using a sample from the Panel Study of Income Dynamics (PSID), I estimate a Bayesian Logistic

Smoothed Transition Autoregressive model of order 1 (LSTAR(1)) with a rich level of heterogeneity in the innovations. I find that there is a life-cycle pattern in the earning shocks: before the age 29, young workers experience shocks with higher variance and a positive probability of lower persistence than older workers. A comparison with conventional models shows that an incorrect model specification introduces bias in the estimates. The proposed model can be easily approximated with a discrete Markov process. This means that this model can be used by macroeconomists to calibrate income processes.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
THE MACROECONOMIC EFFECTS OF A NEGATIVE INCOME TAX

1.1 Introduction

There has been a continuous demand for a reform of the U.S. Federal income tax, and numerous reform proposals have been floated. Some, like the flat tax (Hall and Rabushka, 1985), sank, while others did not—e.g., the significant reduction of marginal tax rates that took place in the 1980*s*. This study focuses on a particular reform proposal, the Negative Income Tax[1] (NIT; M. Friedman, 1962), and for the first time carries out the quantitative analysis of the tax in a general equilibrium setting.

A NIT works as follows. At the beginning of the fiscal year, all households receive a transfer from the government , say $2000. During the period, all income made is taxed at a constant rate, say 20%. Then, households with yearly income of less than $10000 ($2000/0.2) pay no taxes and receive a positive net transfer (negative tax). As income increases, the effect of the transfer declines. Under the NIT, all households have a guaranteed minimum income.

In this chapter, I ask the following questions: What are the general equilibrium effects of replacing the actual income tax with a NIT? Specifically, what are the macroeconomic effects on income and earnings, labor supply, savings and welfare?

[1]There was a failed attempt to introduce it as a legislation during Richard Nixon's Presidency, but after all the modifications introduced in the Parliamentary debate, Milton Friedman who was a candid advocate, withdrew his support.

Should we pick a NIT?

To tackle these questions, I study a life-cycle economy, in which agents are ex-ante homogeneous but grow different over time as a result of life uncertainty together with age-independent and idiosyncratic persistent productivity shocks. At any point in time, the resulting heterogeneity is characterized by the agents' shock history, their level of asset accumulation, and their age. There is a social security system and agents receive benefits, once they are retired, in the form of lump-sum transfers. In addition, accidental bequests may occur, and are distributed evenly among all living agents[2].

I calibrate this model to match features of the U.S. economy, reproducing the inequality of labor earnings across individuals. My model of the U.S. income tax recognizes the important role of transfers like the Earned Income Tax Credit (EITC). In addition, my tax function mimics the effective average tax rates paid by the American households. I focus on a stationary equilibrium and find the level of transfers and marginal tax rate such that the NIT reform is revenue neutral and maximizes ex-ante welfare (i.e. expected utility prior to birth and revelation of agents' types). My findings can be summarized as follows.

First and foremost, the NIT produces important welfare gains. A NIT with a marginal tax rate of 28% and a transfer of 10% of per capita GDP, roughly $4600, implies a welfare gain equivalent to 6.33% increase in annual consumption in every

[2]It is worth to notice that this benchmark economy has a level of transfers higher than what is actually seen in the data, especially for low-income households, where accidental bequests and social security benefits are not as important in relative terms as they are in the model. This matter is not irrelevant as the transfer in the NIT will be relatively more important for low-income than high-income households.

state of the world. Low-ability agents in the bottom quintile of the productivity distribution benefit the most, with welfare gains that range from 12% to 64%. However, there are losers under the NIT: those in the upper level of the productivity distribution –the "upper class". Therefore, the NIT is a "Robin Hood tax", taking from the rich and giving to the poor.

The level of transfers plays an important role in the results. Indeed, a proportional tax, i.e. a NIT with no transfers (a "non-negative income tax"), has a welfare loss of 4.3% relative to the current tax system. This result is unsurprising as redistribution by transfers is an important feature of the actual U.S. income tax. The elimination of transfers benefits only the highest productivity agents, who as a result face a lower marginal tax rate. Moreover, transfers are not only important, they are essential - a flat tax (a "non-negative income tax" with a fixed deduction) does not outperform the NIT. In particular, the optimal flat tax (characterized by a marginal tax rate of 19% and a fixed deduction of 33% of per capita GDP, roughly $15000[3]) produces a welfare loss of 0.12%[4]. Therefore, the replacement of lump-sum transfer by a fixed deduction is not welfare enhancing.

Second, there is a negative relationship between the size of the transfers and per capita GDP, which decreases by 13% under the optimal NIT. The reason is simple: leisure is a normal good and the presence of the transfer insures agents against

---

[3]Hall and Rabushka (1985) propose a Flat Tax with a deduction of $22500 and marginal tax rate of 19%.

[4]In order to put into perspective this welfare loss, it is necessary to note that I am not taking into account the transition dynamics, and the comparison made is between steady states.

periods of low productivity, enabling them to work only when they are productive. Therefore, the composition of the labor force changes: high-productivity agents increase participation at the expense of low productivity types. This is reflected by the fact that although labor supply measured in hours worked drops by 18%, labor supply measured in efficiency units falls by only 7%. Consequently, the Gini coefficient for labor earnings declines from 0.46 to 0.53.

Third, the transfer reduces the individual incentives to save, and the saving rate drops 11%, implying a reduction in the capital stock of 23% and the capital output ratio of 11%. The resulting decrease in the wage rate and increase in the interest rate produces an extra source of welfare gain for capital income earners (retirees and high-productivity agents) and a welfare loss for wage earners (the youngsters and low-ability agents). In order to isolate the role of price changes in producing the welfare gain, I employ a small open economy assumption (fixed interest rate and wages) and find that welfare increases by only 3% relative to the current system. Moreover, the decreases in the capital output ratio and per capita GDP are even larger (23% and 16%) than in the move to the optimal NIT.

This finding has an interesting implication that is worth pointing out. Usually, the welfare gains in a closed economy are higher than in an open economy but, in this case, the opposite is true which means that the agents in this economy are over-accumulating capital. Therefore, the welfare gains of this reform in a closed economy, taking into account the transitional dynamics, will be similar to the small open economy welfare gains.

Fourth, the way accidental bequests are modeled plays an important role in the results. If instead of the generous scheme in which all accidental bequests are returned as a lump-sum transfer to all agents, I assume a scenario in which all bequests are taxed away by the government, the resulting optimal NIT implies a 40% larger increase in welfare with respect to the benchmark case. As usual, the true state of the world lies somewhere in between the scheme considered and the extreme case of no accidental bequests.

Finally, there is a negative relationship between the persistence of the shocks and the size of the transfers in the optimal NIT. I have considered two cases, one in which the half life of the shocks is doubled relative to the benchmark case and another in which the half life is halved. In the more persistent case, the welfare gain is 8.7%, and the transfer level is 11% of per capita GDP, which are both higher than in the baseline scenario. In the less persistent case, the welfare gain is reduced to 2.6% and the transfer level is 9%. Clearly, the riskier the economy, the higher the gains from providing public insurance.

1.1.1 Related Literature

This quantitative approach to optimal taxation has been followed in several papers, with models similar to mine, in which artificial economies with heterogeneous agents and incomplete markets (e.g. M. Huggett, 1993; and S.R Aiyagari, 1994) are simulated and the individual and aggregate effects of tax reforms are studied (e.g. G. Ventura, 1999; D. Altig et al, 2001; D. Domeij and J. Heathcote, 2004; J. Diaz-

Gimenez and J. Pijon-Mas, 2005; S. Nishiyama and K. Smetters, 2005; among others). For instance, G. Ventura (1999) studies the effects of a flat tax reform of the U.S. income tax and finds that a flat tax has positive impacts on capital accumulation, labor supply measured in efficiency units, earnings, and income. D. Domeij and J. Heathcote (2004) study the distributional effects of reducing capital taxes and I. Correia (2010) shows the distributional and welfare effects of replacing the U.S. income tax with a flat tax on consumption plus lump-sum transfers. Her approach is different than mine as she studies an economy populated with infinitely lived agents differentiated by the initial level of wealth and life-long labor productivity, whereas in my model, agents are born with no assets and there are no fixed inborn differences in labor productivity.

J.C. Conesa and D. Krueger (2006) focus on the optimal level of progressivity in the U.S. income tax and find that a flat tax with a tax rate of 17% and a deduction of $9400 is optimal, with an ex-ante welfare gain of 1.7%. Their approach differs from mine because they restrict themselves to a particular set of tax functions which do not allow for transfers. Along the same line, J.C. Conesa, D. Krueger and Kitao (2009) extend the work of J.C. Conesa and D. Krueger (2006) and allow for differences in the tax rates on capital income and labor earnings and show that capital should be taxed at a positive rate - in accordance with the results of A. Erosa and M. Gervais (2002).

This chapter is also related to the literature on the effects of redistributive taxation, in particular, two strands, one involving the effects of earnings shocks and insurance (e.g. J. Eaton and S. Rosen, 1980; M. Flodén, 2001; M. Flodén and J. Lindé, 2001; and D. Krueger and F. Perri, 2009) and another involving the distortions to labor supply decisions (e.g. R. Rogerson, 2008; E. Prescott, 2002; and M. Feldstein, 1973; among others). M. Flodén and J. Lindé (2001) study the provision of insurance through government transfers in the U.S. and Sweden, and find that a transfer of 15% of per capita GDP in the U.S. and 1.6% of per capita GDP in Sweden are optimal with welfare gains of 8.5% and 1.6%. M. Flodén (2001) studies the effects on risk sharing of different combinations of Government debt and lump-sum transfers, and their effects in isolation. My work differs from these previous two papers in one important aspect: their aim is to find an optimal level of transfers or a combination of Government debt and transfers without replacing any of the present taxes. On the contrary, I focus on a particular revenue neutral tax reform that has a lump-sum transfer as an important component but has another source of welfare gain: the increase in efficiency produced by the replacement of increasing marginal tax rates with a constant tax rate.

This chapter is organized as follows. Section II introduces the model and the definition of equilibrium. Section III presents the calibration strategy and the quantitative results. Section IV shows a sensitivity analysis and Section V concludes.

1.2 Model

The modeling framework is a general equilibrium life-cycle economy, populated by J heterogeneous overlapping generations. Agents face idiosyncratic risk and life uncertainty. Time is discrete and there is no aggregate risk[5]. There are no explicit insurance arrangements.

1.2.1 Environment

At each date t, a continuum of ex-ante homogeneous agents is born. An agent of age j faces a conditional survival probability s_{j+1} of being alive in the next period but no one survives after age J. There is an exogenous retirement age R, adding the first dimension of heterogeneity in the model: agents can be classified as workers or retirees depending on whether their ages are higher or lower than $R+1$.

There is a fixed positive population growth rate n and the total measure of the population at time t is N_t. Despite the fact that the population size evolves through time, each age $j-$generation represents a constant fraction μ_j of the total population size, making the demographic structure stationary[6].

All agents share a time separable utility function and value the expected discounted stream of leisure and consumption:

$$\sum_{j=1}^{J} \widetilde{\beta}^{j-1} \left(\prod_{i=1}^{j} s_j \right) u\left(c_{j,t}, l_{j,t}\right)$$

where $c_{j,t}$ and $l_{j.t}$ denotes consumption and leisure at age j and period t respectively. The momentary utility function is Cobb-Douglas:

$$u\left(c_{j,t}, l_{j,t}\right) = \frac{\left[c_{j,t}^{\nu}\left(1 - l_{j,t}\right)^{1-\nu}\right]^{1-\sigma}}{1-\sigma}.$$

[5]Krusell and Smith (1998) show that the inclusion of aggregate uncertainty, besides of introducing an extra layer of difficulty in the model, does not significantly change the results of a model with no aggregate uncertainty.

[6]The weights μ_j are obtained by the recursive formula $\mu_{j+1} = \mu_j.s_{j+1}/(1+n)$

Consumption and leisure are not separable and the intratemporal elasticity of substitution is equal to 1. The parameter $\nu \in (0,1)$ influences the time spent working, and together with $\sigma > 0$ influences the degree of risk of aversion and the Frisch elasticity of labor supply[7](Rios-Rull, 1995)

1.2.2 Agents' endowments and labor productivities

Agents are born with no assets and during their working life they are endowed with one unit of time. They receive a competitive wage rate w_t and their labor productivity is a first-order Markov process given by $e(z', j)$, which is a function of the shock $z' \in Z$, and their age $j \in J$:

$$\begin{aligned} \ln e(z', j) &= \gamma_j + z' \quad \text{and} \\ z' &= \rho z + \varepsilon, \quad \text{where } \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right). \end{aligned}$$

Thus, agents differ in the efficiency units of labor they supply to the market depending on their age and their shock history. Therefore, the labor income of an agent of age j and shock z is equal to $w_t l_t e(z, j)$, where l_t is the amount of time that the agent decides to work. At age 1, the measure of agents with shock z is $q(z)$.

The possibilities for insurance in this economy are limited. There are no annuity markets and agents cannot trade contingent claims. Nevertheless, agents

[7]The Frisch elasticity, which gives the elasticity of hours worked to changes in wages, keeping the marginal utility of consumption constant, is given by:

$$\eta(\nu, \sigma, l) = \frac{(1-l)}{l} \frac{[1 - \nu(1-\sigma)]}{\sigma}$$

The Arrow-Pratt measure of Relative Risk Aversion $\rho = -\frac{c u''_{cc}(c)}{u'_c(c)}$ is $1 - \nu(1-\sigma)$.

trade a one-period risk-free asset $a_{j,t} \in A \subseteq \mathbb{R}_+$ that will help them partially insure against their idiosyncratic productivity shocks. Agents are not allowed to borrow.

1.2.3 Firms and Technology

There is a representative firm that produces total output Y_t with a Cobb-Douglas production function:

$$Y_t = K_t^{\alpha} (A_t L_t)^{1-\alpha} .$$

where K_t and L_t are the aggregate capital and labor (measured in efficiency units) at time t, and $A_t = A_0 (1+g)^t$. The resource constraint is:

$$C_t + K_{t+1} - K_t (1-\delta) + G_t \leq K_t^{\alpha} (A_t L_t)^{1-\alpha} .$$

Following conventional notation, δ is the depreciation rate, G_t is public consumption and C_t is total private consumption.

1.2.4 Government and tax structure

At time t, the government receives payments from the social security system and the income tax. The proceeds serve to finance government consumption G_t , pay social security benefits SS_t and transfers TR_t. The social security system is fully funded by social security taxes paid by the working agents at a constant marginal tax rate τ_{ss} on labor earnings. Benefits are distributed evenly among all retirees of a particular cohort and are kept constant through out the retirement period[8]. Accidental

[8]This setting will not let me capture the actual degree of risk sharing present in the actual social security system. Although, this assumption will underestimate the potential

bequests occasioned by deaths of agents are returned as a lump-sum transfers to all living agents. Agents do not derive any utility from government consumption G_t[9].

The actual U.S. income tax system is the benchmark case and I will aim to replicate two of its main features: the double taxation of dividends and the effective tax rates paid by households. For the case of the double taxation of dividends, I introduce a constant corporate income tax τ_k that is levied on capital income.

In the personal U.S. income tax, an agent pays taxes on his total income, defined as the sum of labor and capital income, according to an income scale given by six brackets. Each bracket has a different statutory marginal tax rate τ_i that increases with the bracket, making the tax progressive. Mimicking the income tax requires recognition that there exists a considerable number of tax credits, deductions and overlapping provisions, which together implies that the statutory tax rates faced by an agent are not necessarily the ones effectively paid. Moreover, the presence of the Earned Income Tax Credit (EITC) needs to be taken into account[10]. Therefore, I follow N.Guner et al (2008) and replicate the average tax rate paid with the following

benefits of the reform, it eliminates the need to include agents' past contributions as a state variable.

[9]This assumption is consistent with either two views: 1) all government consumption is wasteful; 2) the consumption of public goods enters linearly in the agent's utility function. In any case, the results would be the same.

[10]The Earned Income Tax Credit is a refundable tax credit for low and middle income families, who satisfy certain requirements and is calculated based upon the number of children in the household, among other things. It was enacted in 1975 and has been expanded ever since (Moffit, 2003).

function[11]:

$$Average\ Tax\ Rate\,(Normalized\ Income) = \eta_1 + \eta_2 \log\,(Normalized\ Income)\,. \tag{1.1}$$

where *Normalized Income* is *Income* divided by the *Mean Household Income*. Then, the total taxes paid by an Agent are:

$$T^{Benchkamrk}_{j,t}\,(Income) = Average\ Tax\ Rate\,(Normalized\ Income) \times Income.$$

In the reform scenario, the NIT replaces the U.S. personal income tax, leaving the rest of the taxes and the social security system unchanged. Now, all agents receive a fixed lump-sum transfer TR^{NIT}_t at the beginning of the period and pay a constant marginal tax rate τ for every unit of income earned. Then, the total tax liability for an agent of age j and shock z with income $I_{j,t} \equiv w_t e\,(z,j)\,l_{j,t} + a_{j,t} r$ is:

$$T^{NIT}_{j,t} = I_{j,t} \times \tau - \mathrm{TR}^{NIT}_t.$$

1.2.5 Agent's problem: recursive formulation

The sate of any agent is fully described by his/her assets holdings a, productivity shocks z and age j. Let $x = (a, z)$ be the non-age dependant variables of the state vector. The mathematical formulation of the state space is formalized as follows.

[11] Several papers use the Gouveia and Strauss tax function (Gouveia and Strauss, 1994) to approximate the average tax rates paid instead of the function depicted above. Even though, it approximates the average tax rate well, it implies a lower marginal tax rate for higher incomes than the ones seen in the U.S. income tax. Moreover, it behaves as a flat tax for incomes higher than twice the household mean income. On the other hand, the tax function that I am considering not only does approximate well the average tax rate but it also does a good job for the marginal tax rates.

Let $(A, \mathcal{A})$, $(Z, \mathcal{Z})$ and $(J, \mathcal{J})$ be measurable spaces, where $\mathcal{A}$ is the Borel $\sigma-$algebra defined on A; $\mathcal{Z}$ is the Borel $\sigma-$algebra defined on Z, and $\mathcal{J}$ is the Power set of J. Let $(X, \mathcal{X}) = (A \times Z, \mathcal{A} \times \mathcal{Z})$ be a product space and $(x, j) \in X \times J$ be the state vector. Let $(X, \mathcal{X}, \psi_j)$ be the probability space, where $\psi_j : \chi \rightarrow [0, 1]$ is a probability measure. The measure of agents with state $x = (a, z)$ within the cohort of age j is $\psi_j(x)$.

I need to do stationary inducing transformations of the variables in order to express the model in terms of a dynamic programming formulation. Let $a_j(x) \equiv \frac{a_{j,t}}{A_t}$, $l_j(x) \equiv l_{j,t}$, $c_j(x) \equiv \frac{c_{j,t}}{A_t}$ be the asset, labor supply and consumption decision rules; let $w \equiv \frac{w_t}{A_t}$ and $r \equiv r_t$ be the wage rate and interest rate and let $\beta \equiv \widetilde{\beta}(1+g)^{\nu(1-\sigma)}$; let $G \equiv \frac{G_t}{A_t}$, $K \equiv \frac{K_t}{A_t}$ and $L \equiv L_t$ be the aggregate government consumption, capital and labor supply and let $\text{TR} \equiv \frac{\text{TR}_t}{A_t}$ and $T_j(x) \equiv \frac{T_{j,t}}{A_t}$ be the transfers and tax collection, and $SS_j \equiv \frac{SS}{(1+g)^{j-(R+1)}}$, where $SS = \frac{SS_t}{A_t}$, be the social security benefits[12]. Finally, let $T : X \times J \rightarrow X \times J$ be an operator and let ν denote the expected discounted stream of consumption and leisure for an agent with state (x, j) behaving optimally from now onwards.

Then, given prices $\{w, r\}$ and a tax regime T_j^k with $k \in \{NIT,\ Benchmark\}$, an agent of age j with state x needs to choose the amount of labor $l_j(x)$ to supply to the market, how much to consume $c_j(x)$ and the amount of assets $a_{j+1}(x)$ to carry over the next period. Optimal decisions rules solve the following dynamic programming problem:

[12]Time subscripts have been dropped because I am interested in a stationary equilibrium.

1. Working agents:

$$\nu\left(x,j\right)=\left(Tv\right)\left(x,j\right)\equiv\sup_{\left(c_j,l_j,a_{j+1}\right)}\left\{u\left(c_j,l_j\right)+\beta E\left[v\left(x',\ j+1\right)\right]\right\}.$$

subject to

$$c_j+a_{j+1}\left(1+g\right)\leq a_j\left(1+r\right)+w\left(1-\tau_{ss}\right)e\left(i,j\right)l_j-T_j^k\left(x\right)+\mathrm{TR}\qquad if\ j\leq R$$

$$c_j\geq 0,\quad a_j\geq 0,\quad a_{j+1}\geq 0\,and\;l_j\in\left[0,1\right]\quad .$$

2. Retirees:

$$\nu\left(x,j\right)=\left(Tv\right)\left(x,j\right)\equiv\sup_{\left(c_j,a_{j+1}\right)}\left\{u\left(c_j,0\right)+\beta E\left[v\left(x',\ j+1\right)\right]\right\}.$$

$$c_j+a_{j+1}\left(1+g\right)\leq a_j\left(1+r\right)-T_j^k\left(x\right)+\mathrm{TR}+SS_j\qquad if\ j>R$$

$$c_j\geq 0,\quad a_j\geq 0,\quad a_{j+1}\geq 0\ \ .$$

with

$$v\left(x,J+1\right)\equiv 0$$

1.2.6 Stationary Equilibrium

Definition 1.1. *A stationary equilibrium is a collection of value functions* $v\left(x,j\right)$, *decision rules* $\left\{c_j\left(x\right),l_j\left(x\right),\ a_{j+1}\left(x\right)\right\}_{j=0}^J$, *factor prices* $\left\{w,r\right\}$, *a tax regime* T_j^k, *taxes paid* $T_j\left(x\right)$ *and transfers* TR, *aggregate capital* K *and labor* L, *government consumption* G *and social security benefits* SS_j, *with a collection of invariant distributions* $\left(\psi_1,\ldots,\psi_J\right)$ *such that:*

1. *Decision rules* $c_j(x)$, $l_j(x)$ *and* $a_{j+1}(x)$ *together with a value function* $v(x,j)$ *solve the decision problem for an agent of age* j *and state* x.

2. *Factor prices are competitive:*

$$w = F_2(K,L).$$

$$r = F_1(K,L) - \delta.$$

3. *Market clearing conditions are satisfied:*

   (a) $\sum_j \mu_j \left[\int_X (c_j(x) + a_{j+1}(x)(1+g))\, d\psi_j(x)\right] + G = F(K,L) + (1-\delta)K$

   (b) $\sum_j \mu_j \int_X a_{j+1}(x)\, d\psi_j(x) = (1+n)K$

   (c) $\sum_j \mu_j \int_X l_j(x)\, e(z,j)\, d\psi_j(x) = L$

4. *Law motion of distributions are consistent with individual decision rules:*

$$\psi_{j+1}(B) = \int_X P(x,j,B)\, d\psi_j(x).$$

   where $P(x,j,B) = 1$ *if* $a_{j+1}(x) \in B$, *and* $P(x,j,B) = 0$ *otherwise,* $\forall B \in \mathrm{X}$, $j = 1, \ldots, J$. $\psi_1(x)$ *is unequivocally determined by* $q(z)$ *as agents are born with no assets.*

5. *Government budget is balanced:*

$$G = \sum_j \mu_j \int_X T_j(x)\, d\psi_j(x).$$

6. *The social security system is fully funded:*

$$\tau_{ss} w L = \sum_{j=R+1}^{J} \mu_j SS_j.$$

7. *Transfers are equal to accidental bequests:*

$$(1+n)\,TR = \sum_j \mu_j\,(1-s_{j+1}) \int_X a_{j+1}(x)\,(1+r)\,d\psi_j(x)\,.$$

1.3 Results

1.3.1 Calibration

In this subsection, I discuss the calibration strategy and the assumptions made for the benchmark economy. I set the model period equal to 1 year. Table 1.1 summarizes the parameters values used in the calibration. Table 1.2 presents the results for the calibrated economy.

Table 1.1: Calibrated Parameters.

Parameters	**Value**	**Target.**
β	0.979	$K/Y = 2.89$
σ	4	$IES = 0.5.$
ν	0.383	Average time spent working= 1/3
J	81	Maximum Age 100
R	45	Retirement age 65
n	1.1%	Data
ρ *and* σ_ε^2	0.973 *and* 0.02	J. Heathcote et al (2010)
α	0.35	Capital share.
δ	4%	$I/Y = 21.38\%$.
g	2.22%	Data
η_1 and η_2	10.23% & 7.33%	N. Guner et al (2008)
τ_{ss} and τ_k	8.6% & 7.48%	Data

Table 1.2: Benchmark Economy

Variables	Values
GDP per capita	0.682
Capital Stock per capita	1.976
Labor Supply per working agent	0.385
Hours per working agent	0.33
Measure working age population	0.80
Saving Rate	9.66%
Capital Output Ratio	2.89
Capital per Labor	5.141
Wage	1.153
Interest Rate	0.080
Social Security benefits per capita	0.235
Bequests per capita	0.022
Mean Household Income	0.662

1.3.1.1 Demographics

In my model, agents are born at age 21 (model period 1), work until age 65 (model period 45, i.e. $R = 45$) and die for certain at age 100 (model period 81, i.e. $J = 81$). Survival probabilities s_{j+1} are taken from the National Vital Statistics System[13]. Population growth n is set equal to 1.09% which is the average population growth for the U.S. during $1990 - 2009$[14].

1.3.1.2 Preferences

I set σ equal to 4 and calibrate ν endogenously in order to achieve an average time spent working equal to a 1/3. The resulting value for ν is 0.383 which together with σ give an intertemporal elasticity of substitution approximately equal to 1/2,

[13] National Vital Statistics Report, volume 58, number 10, March 2010.

[14] Economic Report of the President 2010, Table $B34$.

and a Frisch elasticity of 1, consistent with previous macro estimates[15] (see D. Domeij and M. Flodén, 2006; and L. Pistaferri, 2003).

The discount factor β is calibrated endogenously to 0.979 in order to target a capital-output ratio equal to 2.89. This last figure is the average capital-output ratio for the period $1960-2007$. I calculate it following the Cooley and Prescott Methodology[16].

1.3.1.3 Technology

I set α equal to 0.35, which is the average of capital income over total income for the period $1960-2007$ (Cooley and Prescott, 1995). The parameters for the labor augmenting technology are calibrated as follows. The growth rate g is equal to 2.22% and is taken from the average growth rate of real per capita GDP during $1960-2007$[17]. The parameter A_0 is a free parameter and I set it equal to 1. Also, I set the depreciation rate δ equal to 4% to assure an investment-output ratio equal to 21.38%[18], which was the average for the $1960-2007$ period.

[15] The macro estimates of the Frisch elasticity tend to be higher than those from the labor literature for prime age male workers. This apparent discrepancy is the result that macro elasticities are unrelated to micro elasticities, as was documented by R. Rogerson and J. Wallenius (2009)

[16] Data for Residential and non-residential structures (equipment and software, structures) and consumer durable goods comes from Table 1.1. Current-Cost Net Stock of Fixed Assets and Consumer Durable Goods BEA April 2010 (http://www.bea.gov/national/FA2004/SelectTable.asp). Data for the stock of Land comes from Flow Funds Accounts, Table B.100, Table B.102 and Table B.103. Inventories are taken from the Economic Report of the President 2010, Table B1.

[17] Economics Report of the President 2010, Table B.26.

[18] Investment comes from Economic report of the President 2010, Table B1. Consumption of durables is taken from Economic report of the President 2010 Table B16.

1.3.1.4 Taxes

I need to calibrate three taxes: the social security tax τ_{ss} and the U.S. personal and corporate income taxes. For the first case, I calculate the average social security contribution as a fraction of total labor income for the $1990-2000$ period and set τ_{ss} equal to 8.6%.[19].

In the case of the U.S. personal income tax, I need to specify a parametric function to reproduce the effective average tax rate paid by an American household. For that purpose, I use the N. Guner et al's (2008) estimates for married households. They use data from the U.S. Internal Revenue Service for the year 2000 and calculate the average tax rate for every income bracket normalized by the mean household income for the period as:

$$average\ tax\ rate = \frac{\frac{total\ amount\ of\ income\ tax\ paid}{number\ of\ taxable\ returns}}{\frac{total\ adjusted\ gross\ income}{number\ of\ returns}}.$$

They fit the function (1.1) and obtain $\widehat{\eta}_1 = 10.23\%$ and $\widehat{\eta}_2 = 7.33\%$ with an $R^2 = 99\%$. As noted by N.Guner, this tax function fits the data a little better than the functional form employed by M. Gouveia and R. Strauss (1994). Further, Figure 1.1 shows that the two formulations lead to very similar average effective tax rates, while Figure 1.2 indicates that the Gouveia Strauss tax function displays a constant marginal tax rate for incomes higher than twice the mean household income. This situation does not correspond to what is seen in the data.

[19] I consider those contributions from Old Age, Survivors and DI programs. Social Security Bulletin, Annual Statistical Supplement, 2005, Tables $4.A.3$.

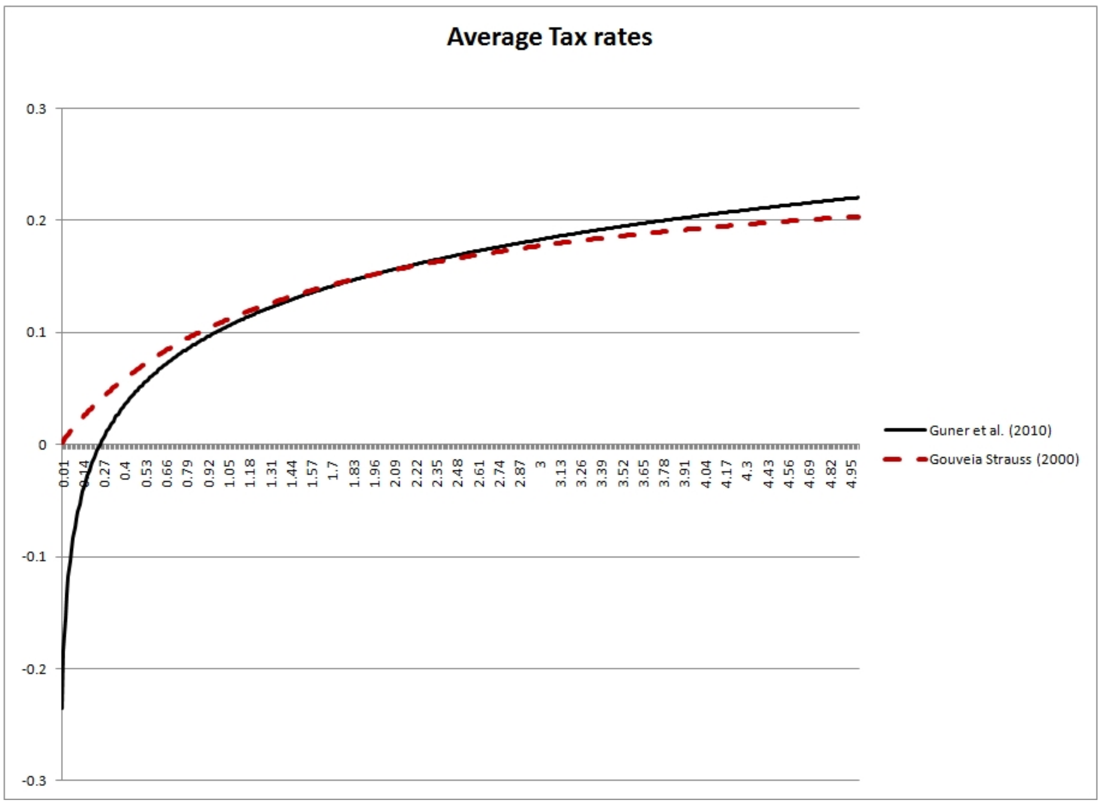

Figure 1.1: Average Tax Rates.

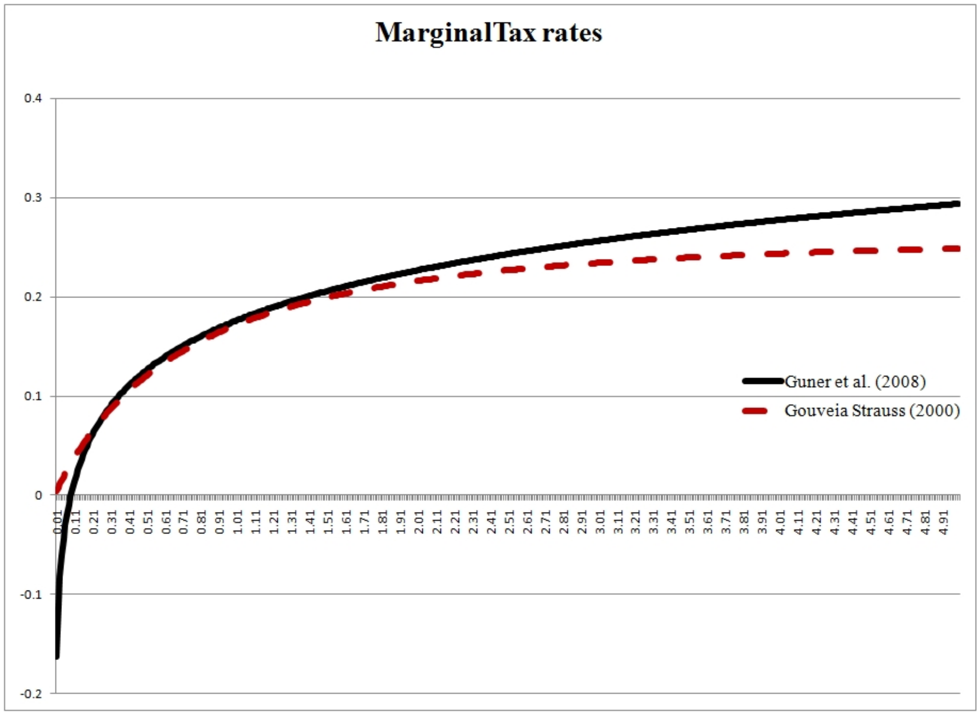

Figure 1.2: Marginal Tax Rates.

For the corporate income tax, I set τ_k equal to 7.48% in order to reproduce the 1.74% average ratio of capital net of depreciation to total income for the $1987-2007$ period (Cooley and Prescott, 1995).

1.3.1.5 Idiosyncratic Shocks

The efficiency profile $e(z', j)$ has an age-component γ_j, which is taken from G. Hansen (1993), and an idiosyncratic shock-component z', which follows an $AR(1)$ process whose values are taken from the J. Heathcote et al (2010) estimates using the PSID data from the period $1967-2000$[20]. They find a correlation coefficient ρ of

[20]The interesting feature of this paper -and the difference with K. Storeletten et al.(2004)- is that they allow in their model for an endogenous supply of labor, which enables me to

0.973 and a variance for the innovations σ_{ε}^2 equal to 0.02.

I use a Gaussian-Hermite quadrature procedure (Tauchen and Hussey, 1991) to approximate this $AR(1)$ with a 21 state Markov process. The transition-probability matrix is $Q \; -Q_{zz'} \; = \; P\left(Z = z'/Z = z\right)-$, where Q is aperiodic and irreducible, what insures an invariant distribution (Hopenhayn and Prescott, 1992). I follow M. Flondé's (2008) approach that consists of taking a weighted average of the conditional and unconditional variances of the AR(1) as variance of the process, and gives a good approximation for highly persistent processes.

The initial distribution of shocks $q(z)$ follows a Gaussian Distribution with mean zero and a variance σ_z^2 that is endogenously calibrated to match the 0.46 Gini coefficient for labor earnings in the U.S. for the year 2000[21]. Table 1.3 shows the 21 values of the shocks in log-scale with their initial and invariant distribution.

1.3.2 Tax Reform's results

In order to understand the effects of a transfer in the tax scheme, I eliminate the implicit/explicit transfers in the actual U.S. income tax through the introduction of a proportional tax, i.e. a NIT with no transfers (a "non-negative income tax"), and then I increase the transfer level in the NIT to 2.5% and 5% of per capita GDP in the benchmark economy. These quantitative exercises will let me evaluate the changes in the aggregate variables and understand how the transfer works. Table 1.4 summarizes the results.

---

take directly their estimates for the AR(1) process.

[21]US Census Bureau 2000.

Table 1.3: Markov process

Types	**States**	**Initial Distribution**	**Invariant Distribution**
1	-2.06	1.94%	0.27%
2	-1.77	1.81%	0.59%
3	-1.53	2.53%	1.17%
4	-1.31	3.35%	2.05%
5	-1.11	4.22%	3.23%
6	-0.91	5.09%	4.66%
7	-0.72	5.91%	6.23%
8	-0.54	6.63%	7.75%
9	-0.36	7.18%	9.04%
10	-0.18	7.52%	9.90%
11	0	7.64%	10.21%
12	0.18	7.52%	9.90%
13	0.36	7.18%	9.04%
14	0.54	6.63%	7.75%
15	0.72	5.91%	6.23%
16	0.91	5.09%	4.66%
17	1.11	4.22%	3.23%
18	1.31	3.35%	2.05%
19	1.53	2.53%	1.17%
20	1.77	1.81%	0.59%
21	2.06	1.94%	0.27%

Table 1.4: Aggregate variables for different levels of Transfers.

Variables	**Baseline**	**Proportional Tax**	2.5% TR	5% TR
GDP per capita	100	110.13	105.36	100.12
Capital Stock	100	116.90	108.08	98.77
K/Y	100	106.16	102.58	98.65
K/L	100	109.57	103.90	97.86
Saving rate	100	106.12	102.52	98.61
Labor supply	100	106.64	103.92	100.86
Hours	100	105.79	101.06	95.73
Wage	100	103.25	101.35	99.25
Interest Rate	100	91.31	96.30	102.13
Bequests	100	111.95	102.55	92.73
Social Security Benefits	100	110.23	105.42	100.23
Marginal tax rate	–	12.28%	15.49%	19.03%
CEV	–	**−4.26%**	**−0.55%**	**2.63%**

In a world with a proportional tax, the marginal tax rate drops to 12%, and the resulting dramatic decrease in the marginal tax rate faced by high-income households means that they benefit more than other households. The absence of the transfer eliminates any redistribution from high to low-productivity agents, making it possible to have a low tax rate. Additionally, the 10% increase in per capita GDP augments the size of the tax base, allowing for a further reduction of the tax rate. The tax bill is reduced for high and medium-productivity agents, while the opposite is true for low productivity types. As a result, labor supply measured in efficiency units increases 7%, while the number of hours worked increases less (6%).

A lower marginal tax rate induces medium and high-productivity agents to supply more hours to the market via a substitution effect, even though there is an income effect that works in the opposite direction. In contrast, low-productivity agents deprived of transfers face a negative income effect making them work more, while a substitution effect caused by a higher tax rate makes them want to work less. The effect is not symmetric and there is a change in the composition of the labor supply: high-productivity agents gain participation at the expense of low-productivity agents. Therefore, average labor productivity increases.

Without transfers agents wish to save more for precautionary reasons: the saving rate, defined as the interperiod change in household's assets holdings divided by GDP, increases 6% leaving this economy with a higher capital stock. Consequently, with more capital and more productive labor available, GDP is higher.

It is worth point out that that the total transfers received by the households can be divided into three different sources: the social security system, the income tax and the accidental bequests. The social security payments are a function of the total earnings in the economy, the income tax transfers depend explicitly on the tax scheme considered while the accidental bequests are proportional to the level of capital in the economy. Therefore, changes in earnings and the capital stock change the composition and size of the total transfers received by the households. The increase in the capital stock (17%), labor supply (7%) and the wage rate (3%) imply that the social security payments and accidental bequests has increased (10% and 12%).

Naturally, there has been a change in the composition of the total transfers received and it can be argued that there has been an improvement in the income of wage earners, particularly low-income households. However, the removal of the income tax transfers has shut down an important redistributive channel that previously benefited low-income households.

Does the increase in social security and accidental bequest transfers offset the loss of the income tax transfers? No, because welfare is lower under the proportional tax. Indeed, to make agents indifferent between a proportional tax and the US income tax regime, consumption under the proportional tax regime would need to increase by 4.26% in every state of the world (i.e. the Consumption Equivalent Variation -CEV- is -4.26%).

An analysis of the CEV by productivity types shows that losses are concentrated on the low types, with welfare losses as high as 8%, while more productive types are better off. Not surprisingly, the most productive agents in this economy have a striking welfare gain of 15.13%. It is clear that the trade-off between tax rates and transfers depends crucially on the productivity type: low-productivity types prefer hight transfers and high tax rates while the opposite is true for high-productivity types.

To illustrate this last point, I increase the transfer from 0 to 2.5% ($1150) and 5% ($2250). By doing so, welfare increases by 3.71 and 6.86 percentage points respectively. Naturally, the increase in the size of the transfer must be accompanied by an increase in tax rates in order to make the tax reform revenue neutral: the marginal tax rates for a NIT with 2.5% and 5% transfers are 16% and 19% respectively.

In the absence of complete markets, the transfer can be thought as a source of insurance: in any state of the world, the transfer is present reducing the need to save and work. Consequently, there is a negative correlation between the size of the transfer, and the saving rate and hours worked. The saving rate moves from 10.25% in a proportional tax regime or "non-negative income tax" to 9.5% in a NIT with 5% transfers (NIT 5%), while the original 6% increase of hours worked in the proportional tax subsides to a 1% increase in a NIT with 2.5% transfers (NIT 2.5%), to finally end in a decrease of 4% in the NIT 5%.

Labor supply measured in efficiency units moves from a 7% increase in the proportional tax case to a 4% and 1% increases in the NIT 2.5% and NIT 5% respectively, remaining in the last case practically at the same level as in the benchmark case, even though there is a drop in the hours worked. As the decrease in hours worked is higher than the decrease in labor supply, it is evident that there is a change

Table 1.5: Gini coefficients Pre-Tax Earnings

	$Gini\ Pre-Tax\ Earnings$
Benchmark	0.46
Proportional	0.47
NIT 2.5%	0.48
NIT 5%	0.49
Optimal NIT	0.53

in the composition of the labor supply and average labor productivity increases.

The reasoning behind this result is that the presence of the transfer enables agents to cope better with bad productivity shocks. Agents hit with a bad shock increase their consumption of leisure, which is a normal good; i.e. they work fewer hours. When positive shocks hit, they work more. This means that the transfer enables them to work when they are more productive.

As a result, there is an increase in the dispersion and concentration of labor earnings. The Gini coefficient on labor earnings deteriorates from 0.46 to 0.47 in the proportional tax setting, 0.48 in the NIT 2.5% and 0.49 in the NIT 5% (see Table 1.5).

In this economy, prices are a function of the capital labor ratio, being the wage rate an increasing function of the capital labor ratio and the interest rate a decreasing function. The capital labor ratio moves from 5.6 in the proportional tax regime to 5.3 in the NIT 2.5%, and 5 in the NIT 5%. This translates into a 4% decrease in the wage rate and a 12% increase in the interest rate from the proportional tax to a NIT 5%.

These changes on prices have a natural impact on the distribution on income. If we abstract from any other transfer present in this economy, it will be possible to argue that young and low-productivity agents, who receive most of their income from labor earnings, are worse off while capital income earners are better off. The final verdict depends on the size of the different transfers.

A conflicting picture emerges for retirees, who have their social security benefits practically unchanged in the NIT 5% after an initial increase of 10% in the proportional tax setting. As the income tax transfer and the interest rate increase, the social security benefits and the accidental bequests are reduced, attenuating the potential gains from the income tax transfer.

It is interesting to notice the relationship between the welfare profile and the introduction of transfers. In a "non-negative income tax" scenario, the welfare profile is monotone increasing in productivity types. Once the transfer is introduced, a U-shaped figure emerges with low productivity types benefitting directly from the transfer, while high-productivity type agents enjoy lower marginal tax rates. The "middle class" is caught in the middle: the transfer is not high enough for them and the tax rates are not as low as they want to (see Figure 1.3).

1.3.3 The Optimal NIT versus the Flat Tax

A NIT with a marginal tax rate of 28% and a transfer of 10% of the benchmark's economy per capita GDP, approximately $4600, is optimal in the sense that it maximizes the expected lifetime utility calculated before the agent is born and knows

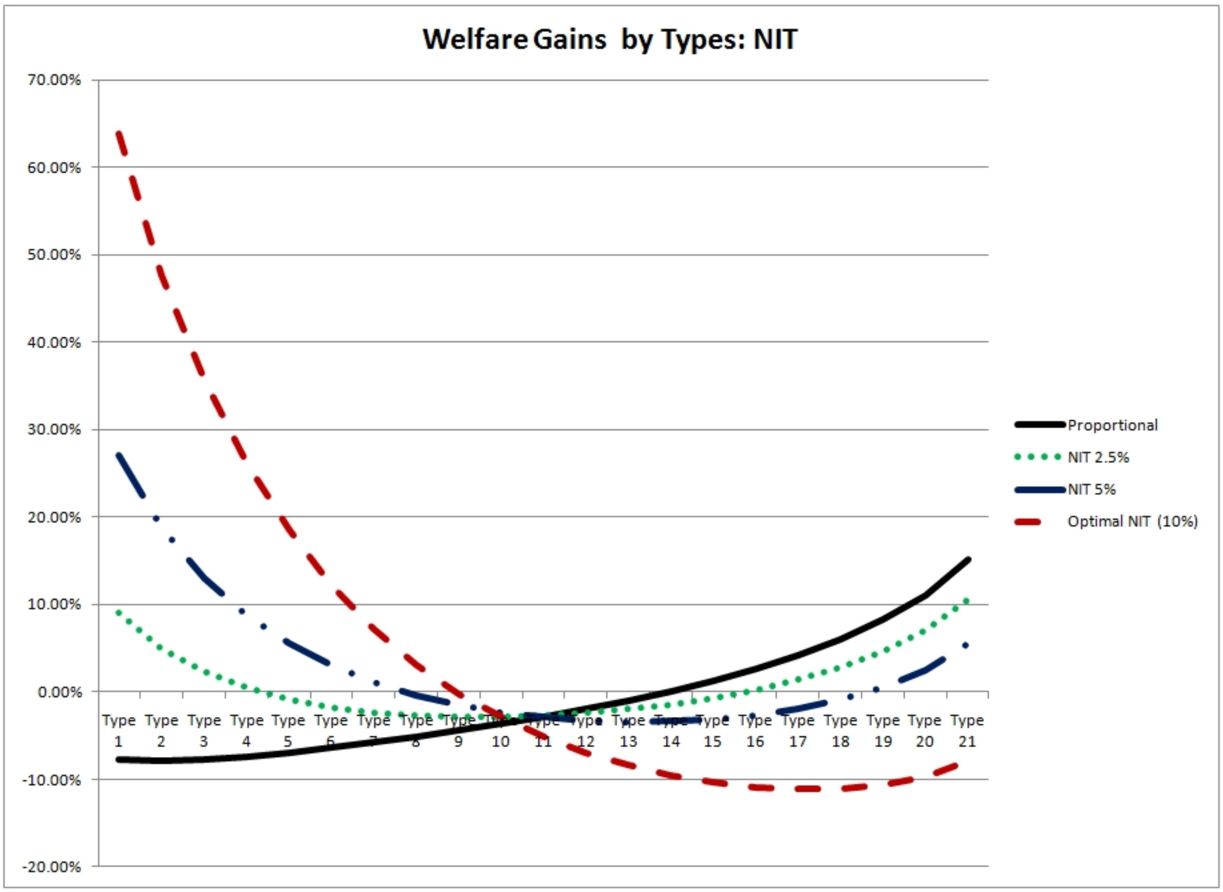

Figure 1.3: Welfare gains by types: Proportional Tax, NIT 2.5% and NIT 5%.

Table 1.6: Optimal NIT versus Optimal Flat Tax

Measures	*Optimal NIT* (10% TR)	*FlatTax* (33% *deduction*)
GDP per capita	87.18	103.81
Capital Stock	77.37	104.85
K/Y	88.74	101.00
K/L	83.22	101.37
Saving rate	88.74	100.89
Labor supply	92.97	103.26
Hours	82.41	101.96
Wage	93.77	100.48
Interest Rate	119.12	98.67
Bequests	70.76	105.91
Social Security Benefits	87.33	103.82
Marginal tax rate	27.95	18.47
CEV	**6.33%**	−0.12%

his true type. The expected welfare gain is an impressive 6.33% increase in individual consumption in every state of the world. The picture that emerges for this economy is similar to those associated with the sub-optimal NIT's of the previous subsection. Table 1.6 summarizes the results.

A striking result is that the optimal NIT causes per capita GDP to decline 13%. This is due to a 23% decrease in the capital stock and a 7% decrease in the labor supply measured in efficiency units. In the previous exercises, per capita GDP declines as the transfer increases. The transfer in the optimal NIT is even larger and per capita GDP declines further.

Next, I examine the determinants of the drop in the capital stock and labor supply. The transfer from the optimal NIT enables agents to save less in order to cope with the uncertainty they face, implying a lower level of capital. Indeed, the need to save for precautionary motives has subsided: the saving rate drops 11%; total savings falls more than the GDP.

The transfer enables agents to substitute leisure for work when they are hit by a bad shock, moving the threshold that splits the active population into working and non-working agents. The NIT provides agents with fewer incentives to adjust their labor supply decisions as means of insurance (see J. Pijoan-Mas, 2006): hours worked decrease 18%, a decrease 2.5 times greater than the reduction of labor supply. Further, the presence of the transfer means that the productivity of the least productive working agent under the NIT is higher than that under the US tax system.

Because less productive agents work less and more productive ones work more, income inequality goes up. The Gini coefficient for labor earnings jumps to 0.53 from 0.46, a deterioration of 15%.

The move to the optimal NIT also changes prices in this economy as a consequence of the 11% and 17% reduction in the capital output ratio and capital labor ratio. The wage rate falls 6% and the interest rate rises 19%, from 8% to 9.5%. For capital income earners, who are concentrated among high-income households, this increase in interest rates partially offset the increase in the tax rates.

As a result of the fall in savings, wage rates, and labor supply, social security benefits and accidental bequests fall 13% and 29% respectively. It is clear that the composition of the total transfers, given by the sum of income tax transfers, social security benefits and accidental bequests, has changed because of the increase in the income tax transfers and the decrease of the social security benefits and accidental bequests. However, it is not necessarily true that its size has diminished.

The welfare profile under the optimal NIT is different from the cases studied above. The clear U-shape relationship has disappeared and there is a single cut-off that separates winners from losers: the winners are agents from the productivity level 1 through 9 with welfare gains from 26% to 64%, while high-productivity agents lose (see Figure 1.4).

Even though, the optimal NIT has a lower marginal tax rate than the actual US income tax for the highest income households, their tax bill has increased as the result of replacing a structure of increasing marginal tax rates under the US income tax with a single tax rate under the optimal NIT. In relative terms, they are better off with respect to medium-income households but for both groups financing a sizeable transfer has increased their tax burden.

The large size of the income tax transfer is due to the persistence of the idiosyncratic shocks. The calibrated value for the parameter ρ implies that a shock has a half life of 25 periods. Thus, agents born with a low productivity shock are plagued by it for a long time, so they prefer a high transfer in order to smooth consumption. The opposite is true for agents born with a good shock. They prefer a low marginal tax rate instead of a high transfer. The natural trade-off between low marginal tax rates and high transfers is a question of efficiency versus insurance. A high transfer means that the insurance and redistributive aspects of the NIT erode the efficiency effects of the tax, i.e. the welfare gains come from the fact that low-ability agents are able to insure themselves against bad shocks.

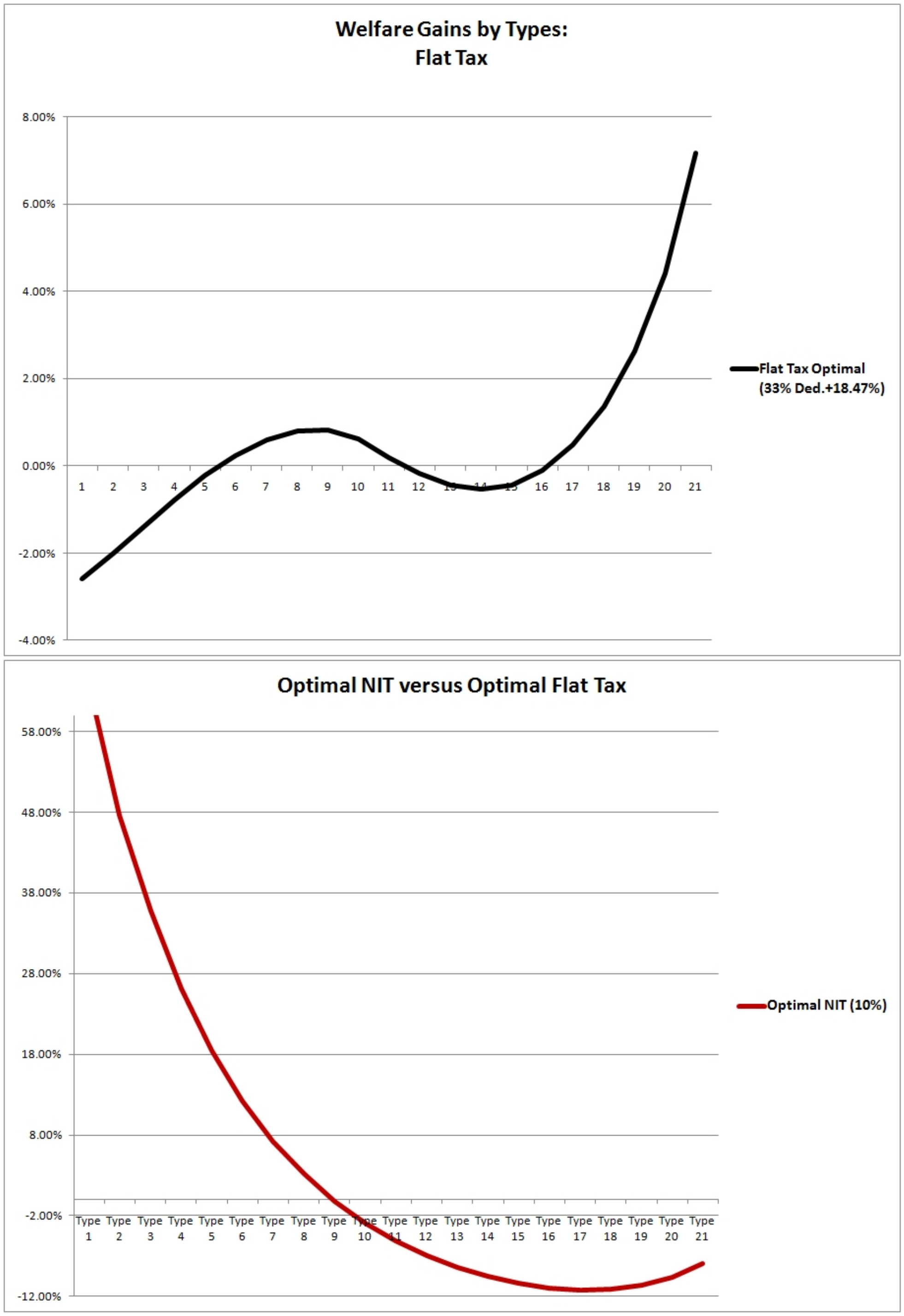

Figure 1.4: Welfare gains by types: Optimal NIT versus Optimal Flat Tax.

To further understand the NIT, I compare it with the popular flat tax and evaluate the desirability of the reform. I search for the optimal flat tax that maximizes ex-ante welfare and find, in accordance with the literature, that a 33% deduction ($15000) computed from the benchmark GDP and a marginal tax rate of 19% is optimal.

The optimal NIT outperforms the optimal flat tax, which has a welfare loss of 0.12%. It may seem surprising that an optimal flat tax implies a welfare loss but as I noted above, I am not taking into account the transition dynamics which may convert this welfare loss into a welfare gain. Even if this were to be the case, the steady state welfare gain under the NIT outweighs any potential transitional gains under the flat tax.

In a world with a flat tax, all the implicit/explicit transfers from the income tax are replaced with a fixed deduction. With no transfer to fund, the marginal tax rates are lower than the ones in the optimal NIT.

The most interesting result is related to the shape of the welfare profile. It is clear that high-productivity agents benefit but the picture is mixed for low-productivity agents: the lowest types are actually worse off. This result is at odds with previous studies (e.g. G. Ventura, 1999; and Diaz-Gimenez and J. Pijoan-Mas, 2005) but it is explained by the way I modeled the U.S. income tax to capture the actual level of transfers present in the system (e.g. EITC, among others). Such transfers represent an important source of income for low-income households. Naturally, these agents prefer a transfer to a deduction. As productivity increases, because of the U-shaped welfare gain profile, a fraction of the low-income households are better off. This exercise highlights the importance of modeling the income tax transfers carefully.

Under the flat tax, per capita GDP increases 4% as a result of an increase in capital (5%) and labor supply (3%). The latter increases more than hours (2%), a natural consequence of the lower marginal tax rate that gives high-productivity agents more incentives to work and the increase in the lowest ability type's tax bill, due to the replacement of the transfer by a deduction. Agents in the low part of the productivity distribution, between the lowest and the median, see a reduction in their tax bill, while the middle types see the opposite. This explains the different effects on hours and labor supply; in some groups an income effect prevails over a substitution effect, and vice versa.

The saving rate remains practically at the same level with a modest increase of 1%. Nevertheless, total savings increases as a result of the elimination of the transfers. This loss of transfers is partially offset by a 4% increase of social security benefits and the 6% increase of accidental bequests.

The wage rate increases 0.5% and the interest rate drops 1%. The increase in the wage rate partly offset the loss of the transfer for the low productivity types, whose main source of income comes from labor earnings.

In summary, a flat tax implies higher GDP, capital accumulation and labor supply than the optimal NIT, but considerably lower welfare.

1.4 Sensitivity Analysis

In the previous section, I have established that the optimal NIT requires a high level of transfers and produces important changes to prices, social security benefits and accidental bequests that could dampen the potential welfare gains from the reform. Therefore, in this section, I undertake a sensitivity analysis, disentangling the role of prices, accidental bequests and the nature of the shocks in the welfare gains reported[22].

1.4.1 The role of prices: An Open Economy

In this exercise, I make an open economy assumption, with free movements of capital, and therefore, the wage rate and the interest rate are kept fixed. Doing so will let us understand the direction and the role of prices in the tax reform. In this scenario, the optimal NIT has a welfare gain of 6.5% and a slightly lower transfer of 10% of the benchmark economy's per capita GDP. Remarkably, even though the transfer level slightly decreases, the marginal tax rate shows a moderate increase: the new marginal tax rate is 29% (see Table 1.7).

This increase can be explained by the shrinkage of the tax base: GDP decreases 16% versus the 13% shrinkage under the optimal NIT with flexible prices. This reduction is a direct consequence of the drop in the capital stock level. As the interest rate remains constant, the fall in savings is larger, and there is no increase in the interest rate that could partially offset the decrease in individual savings. The saving rate falls a dramatic 23%, doubling the 11% fall previously seen.

[22]Otherwise stated, all comparisons made are against the optimal NIT found in the previous section.

Table 1.7: Optimal NIT in an Open Economy

Measures	*Optimal NIT* (10% TR)
GDP per capita	83.64
Capital Stock	64.38
K/Y	76.98
K/L	66.69
Saving rate	76.85
Labor supply	96.30
Hours	86.86
Wage	100.00
Interest Rate	100.00
Bequests	52.21
Social Security Benefits	96.40
Marginal tax rate	28.74
CEV	6.48%

Even though GDP decreases more, labor supply decreases less: -4% versus -7%. As there is no change in the wage rate, leisure does not get cheaper as it did before, so labor supply does not decrease as much. The same effect can be seen in the total hours worked (-12% versus -17%). Naturally, leaving the wage rate in the same level gives high-productivity agents more incentive to supply more work, gaining participation in the labor supply and changing the composition of the labor input.

As there is no decrease in the wage rate, all the change in the social security benefits comes from the drop in the labor supply. However, the transfers are affected by the drop of capital: accidental bequests are reduced by half (-48% versus -29%).

As can be seen, despite the fact that income tax transfers remains at practically the same level, there is a higher welfare gain characterized by a smaller reduction in hours and social security benefits, but a much larger reduction in accidental bequests.

The direct consequence of shutting down the role of prices in the tax reform is that capital income earners and retirees do not benefit from a higher interest rate, while the wage earners are unaffected. This means that the welfare gains are concentrated at the beginning of the working life, particularly for those suffering low productivity shocks. These agents, despite suffering a dramatic decrease in accidental bequests, are better off if the wage rate does not change. In contrast, retirees receive higher social security benefits and this offsets the interest rate effect.

Moreover, a constant interest rate, lower than the one with flexible prices, gives agents less incentives to postpone consumption, explaining the increase in welfare.

1.4.2 The role of accidental bequests

In the benchmark economy, I treated accidental bequests in the usual fashion by returning them via equal lump-sum transfers to all living agents. Although the assumption is common in this type of model, it implies that bequests are higher on average than what is actually seen in the data. This is especially true for low-income households. Moreover, as most of the welfare gains in this income group come from the NIT explicit transfer, it is important to be careful in modeling the accidental bequests because their treatment could affect the potential gains from the reform. Therefore, in this exercise, I move to an opposite scenario and make the extreme assumption that all accidental bequests are taxed away by the government, and used to finance public consumption.

For this exercise, the definition of revenue neutrality that I am employing is

Table 1.8: The role of accidental bequests

Measures	*Optimal NIT* (10% TR)
GDP per capita	84.20
Capital Stock	72.11
K/Y	85.65
K/L	78.71
Saving rate	85.59
Labor supply	91.52
Hours	80.68
Wage	91.96
Interest Rate	125.34
Bequests	0.00
Social Security Benefits	84.21
Marginal tax rate	30.52
CEV	8.81%

different. Here, prior to the reform, there are two sources of income: the current tax system and the total of accidental bequests. The tax reform will be revenue neutral if it raises the same total revenue, accounting for any change in the equilibrium level of accidental bequests. Table 1.8 summarizes the results.

The optimal NIT now has a transfer level of 10% of the benchmark economy's per capita GDP and a marginal tax rate of 31%. The increase in the tax rate arises from the drop in the level of accidental bequests. As can be seen, the drop in individual savings implies a fall in the capital stock (-28% versus -23%) that leads to lower bequests and the need to raise the marginal tax rate to compensate the shrinkage of the tax base. In a world with no accidental bequests as lump-sum transfers, savings are higher from the very beginning and the introduction of the NIT transfer causes the saving rate to fall more (-14% versus -11%).

The welfare gains are impressive: the CEV is 8.8%, a 40% increment from the original experiment. Naturally, the accidental bequests are not as important as a source of income for the high-productivity agents, while the opposite is true for the low-productivity agents. Therefore, the introduction of an optimal NIT implies welfare gains for the latter group as high as 92%, while the changes in welfare for the former group are not as spectacular.

Labor supply deteriorates 20%, moving to -8.5% from -7%. This is a natural consequence of the higher marginal tax rate, which gives high productivity agents smaller incentives to supply work. Hours deteriorate 10% (-19% versus -18%) and a familiar message emerges again: high productivity types crowd out low productivity ones, increasing the average labor productivity by hour worked.

The greater fall in labor supply and capital implies a greater drop in GDP (-16% versus -14%). Naturally, there is a change in prices: the wage rate is lower (-8% versus -6%) and the interest rate higher (25% versus 19%). A lower wage rate negatively affects low productivity agents but their supply of labor is reduced by the transfer, mitigating the impact of the lower wage rate. However, the transfer is not large enough to compensate for the low wage rate to the medium-productivity types. On the other hand, high productivity types enjoy a higher interest rate which offsets the wage effect.

In a world with no accidental bequests returned evenly as lump-sum transfers to all living agents, there is an increase in the welfare gains due to the NIT as a provider of public insurance.

Table 1.9: Idiosyncratic Shocks

Measures	*Optimal*: 9% TR; $\rho = 0.946$	*Optimal*: 11% TR; $\rho = 0.986$
GDP per capita	89.48	84.49
Capital Stock	80.68	73.44
K/Y	90.16	86.92
K/L	85.08	80.40
Saving rate	90.03	86.78
Labor supply	94.62	91.12
Hours	86.87	77.64
Wage	94.50	92.65
Interest Rate	116.69	122.93
Bequests	73.82	67.59
Social Security Benefits	89.49	84.37
Marginal tax rate	25.71	30.31
CEV	2.60%	8.70%

1.4.3 The role of idiosyncratic shocks

In the final exercise, I analyze the effect of the persistence of the idiosyncratic shocks on the optimal NIT. I analyze two cases. In the first, I pick a new ρ (equal to 0.946) to reduce the half life of a shock by half. In the other, I do the opposite: I choose a new ρ (equal to 0.986) to double the half life of a shock (see Figure 1.5). In all cases, the variance of the error term is changed in order to keep the same mean of the shock process, and by log normality, the variance of the log of the shock. Table 1.9 summarizes the results.

The welfare gains from the move to the NIT is higher the more persistent the shocks are. For the high persistence shock, the welfare gain is 8.7%, an increment of 38% with respect to the benchmark case. In contrast, for the less persistent shock, the welfare gain is 2.6% or a 59% reduction in welfare from the original exercise. It is

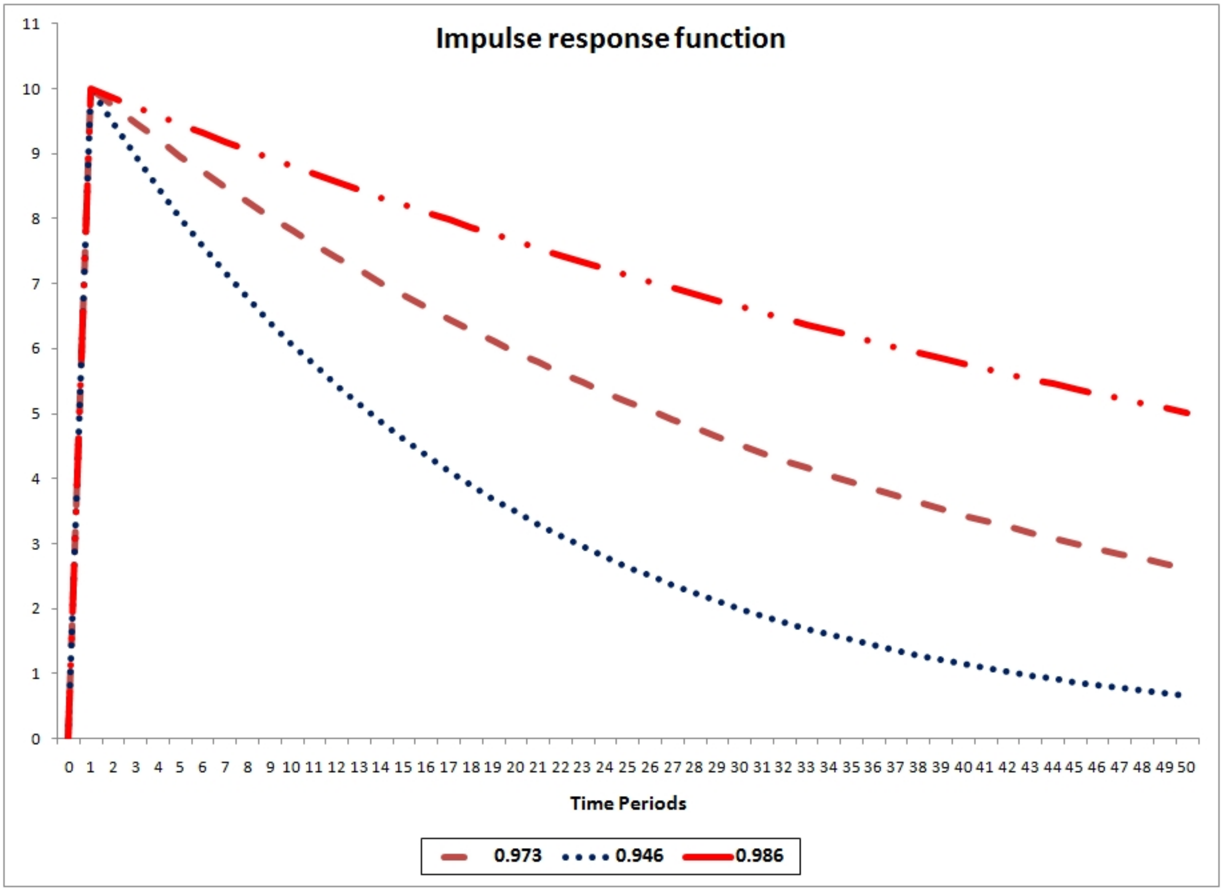

Figure 1.5: Impulse Response Function.

clear that the decrease of the persistence of the shock has a large effect on the optimal NIT and the welfare gains.

The results for transfers and tax rates are as expected. Naturally, the less persistent the shock, the lower the level of the transfer; for a ρ equal to 0.946, the transfer is just 9% while in the other case, it is 11% of the benchmark economy's per capita GDP. The positive association between marginal tax rates and transfers re-emerges. In the less persistent case (low transfer), the marginal tax rate is 26% against a marginal tax rate of 30% in the high persistence case (high transfer).

The relation between the size of the transfer and the aggregate variables is also similar to previous cases. With a lower transfer, GDP drops 11% while with a higher transfer, the reduction is 15%, confirming the negative relationship between the size of the transfer and GDP. A similar story appears in labor supply (-5% versus -9%) and hours (-13% versus -22%): the adjustments are larger the more persistent are the productivity shocks.

The less persistent the shocks, the less the need for higher individual savings. In the less persistent case, capital drops 19% against the 27% reduction when shocks are more persistent. Also, the wage rate falls less the lower the persistence (-6% versus -7%) and the interest rate rises less (17% versus 23%). Consequently, social security benefits (-11% versus -16%) and accidental bequests (-26% versus -31%) fall less the less persistent are the shocks.

The conclusion is clear: with more persistent shocks, the gains from providing public insurance are higher.

1.5 Conclusions

In this chapter, I provide a new, general equilibrium, analysis of a Negative Income Tax (NIT) in a model with ex-ante homogeneous agents beset by idiosyncratic shocks. The model reproduces key features of the U.S. economic data, and in a setting that explicitly takes into account the tax credits, overlapping provisions, and transfers from the actual U.S. income tax. The NIT is simple, consisting of a transfer and a constant marginal tax rate, and produces an outstanding welfare gain, equivalent to a 6.33% increase of individual consumption in every state of the world. The NIT outperforms the popular flat tax (a zero transfer NIT with deductions) by a huge margin.

The optimal NIT has an important insurance component and benefits most those agents who suffer low productivity shocks at the beginning of their working lives. Different and smaller levels of transfers in the NIT have non trivial welfare gains. The NIT has an effect on insurance and efficiency: regarding insurance, individual savings drops as the transfer replaces the need to save, while for efficiency, high-productivity agents face lower marginal tax rates, giving them incentives to work more hours, and the NIT transfer enables all agents to work when they are more productive. Therefore, the composition of labor supply changes and the average labor productivity by hours worked increases. In all cases, the medium-productivity agents are worse off. A similar result emerges from a flat tax.

I conduct sensitivity analyses and show that the persistence of the shocks and the way accidental bequests are modeled have non trivial effects on the welfare gains reported. Further, the more persistent the shocks, the more desirable the reform. Moreover, modeling the level of accidental bequests in two ways brackets the welfare gain that could be obtained from this tax reform: in the both cases considered, one with a generous scheme of bequests and the other one with none, the differences in welfare are not trivial, and in all cases the low-income households are better off.

My comparisons are of steady states and it could be argued that the transitions dynamics of the NIT could be important. However, the welfare gains are of such magnitude that the computation of transitions will not change the direction of the results. Moreover, the important drop in the capital stock in the optimal NIT implies that moving from one steady state to the other one will increase welfare, as agents will consume the capital they have already accumulated.

For future research, it will be interesting to model a NIT in a political economy model, in order to understand the reasons why a tax with such welfare gains had so many difficulties and obstacles at the time it was discussed in the U.S. Congress.

CHAPTER 2
LIFE-CYCLE PATTERNS OF EARNINGS SHOCKS

2.1 Introduction

There is a common perception that a middle-age worker faces risks, in terms of income volatility and job tenure, which are inherently different from those experienced by a twenty-year-old worker, and that these risks are distinct from those faced by a worker close to retirement. This popular belief that there is a life-cycle pattern in labor earning risk has not received much attention in the empirical literature on income dynamics. If such a pattern does exist, modeling it in a world of incomplete markets could significantly improve our understanding of consumption-saving decisions in the life cycle, wealth inequality, insurance, the welfare costs of business cycles and public policies, and other phenomena.

In this chapter, I study labor income dynamics in the life cycle and ask the following questions: Is there a life-cycle pattern in the idiosyncratic earnings shocks? What is the role of aging in these shocks? If there is a life-cycle pattern to earning risk, can it be modeled parsimoniously?

The starting point for addressing these questions is a reduced-form model of income dynamics, as it has been the standard in the literature since the early 1970s, with the early works on the matter by McCall (1973), Shorrocks (1976), Lillard and Willis (1978), Lillard and Weiss (1979), MaCurdy (1982), and Gottschalk (1982).

While the starting point is a reduced form, this chapter introduces a novel methodology that is able to accommodate both statistical fit and economic interpretation. Using a sample of male heads of households from the Panel Study of Income Dynamics (PSID) from the years 1968 to 1996, I estimate a Logistic Smoothed Transition Autoregressive model of order 1 (LSTAR(1)), with a rich level of heterogeneity in the innovations, that replaces the usual linear AR structure of the persistent component of the earnings shocks with a nonlinear structure. At any point in time in the life of a worker, the persistent shocks are a convex combination of two different AR processes, where the weights are a function of a threshold variable that has its own economic interpretation. This means that the model can detect the presence of patterns in the life cycle as well as recognize the economic forces in action.

Although I limit my analysis to the effects of age on the idiosyncratic earnings shocks in a Restricted Income Profile (RIP) setting, where there is no heterogeneity in the income growth rates of the individuals, one important contribution of this chapter is that I employ a Bayesian analysis. This makes it possible to extend the results to a Heterogeneous Income Profile (HIP), i.e. with heterogeneity in the income growth rates, and to cover a much wider range of years in the PSID. This was not possible in the previous models used in the literature, and it has nontrivial economic implications. My findings can be summarized as follows.

First, there is a life-cycle pattern in the idiosyncratic earnings shocks. Workers younger than 29 experience shocks with higher variance and a positive probability of having a lower persistence than older workers. The stationary variance of the persistent component of the earnings shocks for a young worker is 2.5 times higher than the stationary variance for an old worker.

Second, the usual GMM estimates for the persistence in the AR process, in a RIP setting, are higher than 0.95, but a comparison to a model with a rich structure of innovations and the same linear AR structure shows a lower persistence, with a mean of 0.89. Including nonlinearities reduces the persistence to a range of 0.82 to 0.84, a substantial reduction with respect to the standard model used in the literature. A potential explanation could be the unit root problem pointed out by Sims and Uhlig (1991).

Third, the posterior mean of the persistence of the shocks for young workers is higher than the persistence for older workers, but the difference is not as significant as one might expect. The posterior standard deviation of the persistence parameter for the young workers is almost twice that of the standard deviation for the older workers. There is an overlap in the distributions of both persistence parameters that suggests they might be equal. Moreover, this higher dispersion in the posterior distribution of the persistence suggests that using age as threshold variable might not give a complete picture of what is seen in the data.

Fourth, the introduction of nonlinearities shows the importance of correctly modeling the structure of the innovations. The results strongly suggest that there is dispersion in the individual variances of the innovations. The distribution of the individual-specific component of this variance points out a small proportion of individuals with a higher conditional variance than the rest of the sample. Models that ignore this fact introduce an upward bias on the variance estimations. Also, my model is in line with previous estimates of the transitory component of the earnings shocks.

Finally, in this model, the income process is defined as a convex combination of two AR processes and can be easily approximated with a discrete Markov process (see Vandekerkhove, 2005). This means the model is tractable and the usual calibration techniques can be applied.

2.1.1 Literature Review

The empirical literature on income dynamics has produced a variety of models with different levels of complexity and heterogeneity. These models can be classified in terms of the degree of heterogeneity present in the conditional mean and the conditional variance of the income process.

Recently, it has become popular in the macro literature to classify these models in terms of the heterogeneity in the growth rates of individual earning profiles. An income process with a common growth rate for all individuals is a RIP, while if these rates are different for every individual the process is a HIP. Consequently, a RIP and a HIP are just different degrees of heterogeneity in the conditional mean of the process. This classification is motivated by models of human capital where agents with different levels of ability have different returns on their investments in human capital accumulation.

Examples of HIP models with no heterogeneity in the persistence nor the conditional variance are Baker (1997); Baker and Solon (2003); Gottschalk and Moffitt (2002); Gottschalk and Moffitt (2002); Guvenen (2007); Guvenen (2009); Guvenen, Ozkan and Song (2012); and Haider (2001).

Illustrations of RIP models are the papers of Abowd and Card (1989), Hryshko (2012), Lillard and Willis (1978), Lillard and Weiss (1979), MaCurdy (1982), Storesletten, Telmer and Yaron (2004), just to name a few.

A second group of papers aims to explain different patterns seen in the data by modeling the conditional variance of the income process instead of the conditional mean. In this group, the persistent component of the earnings shocks is discarded, and the variance is allowed to vary across time and individuals. This is the case of the papers of Barsky et al. (1997), Chamberlain and Hirano (1999), Jensen and Shore (2012), Meghir and Pistaferri (2004), and Meghir and Windmeijer (1999).

This chapter presents a RIP model and is in the intersection of both groups because of its focus on the conditional mean as wells as the conditional variance of the income process, as in the papers of Browning, Ejrnæs and Alvarez (2010) and Hospido (2012) , to name two. It also contributes to the Bayesian literature on income dynamics that has started recently with Geweke and Keane (2000), Jensen and Shore (2012), Norets and Schulhofer-Wohl (2009), and Shore (2011).

Even though this Bayesian literature on income dynamics is in its initial stages, there are reasons to believe that it will experience a surge. The need for models that capture the right level of heterogeneity and, as Nakata and Tonetti (2012) point out, the good small-sample properties of Bayesian estimates predict a widespread use of Bayesian methods.

There have only been few attempts to model the role of aging in the structure of idiosyncratic earnings shocks. The most relevant papers on this matter are those of Hause (1980), Meghir and Pistaferri (2004), and Karahan and Ozkan (2011).

Hause (1980) aimed to disentangle the effect of "on-the-job training" on the individual earnings profiles in a sample taken from one cohort of Swedish white-collar workers. The residual earnings is modeled with a time-varying AR structure, and even though the variable age is not explicitly mentioned, the fact that there is only one cohort in his sample makes this time-varying process an age-varying AR process. Later, Meghir and Pistaferri (2004) focused on modeling the conditional variance of labor earnings. They introduced an ARCH(1) specification for the permanent and transitory component of the shocks. However, the variable age seems not to have been significant.

Karahan and Ozkan (2011) is to date the most complete paper to explicitly model the age profile of the earnings shocks. After making some identifying assumptions, they used GMM methods to model the residual earnings, including in their specification the individual fixed-effects, a transitory shock, and a persistent AR component, the last two being age dependent. They found that the transitory shocks do not exhibit an age profile, but the persistent component does. They estimated the persistence parameters and variances for every age-bin considered, and in order to simplify their model, they fit an age polynomial onto their estimates.

My work is related to Karahan and Ozkan in the sense that both of our papers are interested in the exploration of life-cycle patterns of the earnings shocks, and both papers are the first to focus on the effects of aging on the structure of these shocks.

However, there are methodological and conceptual differences that distinguish my paper from theirs.

First, I follow a Bayesian approach. This enables me to deal with a level of heterogeneity that is not present in Karahan and Ozkan's frequentist model. Moreover, my model can easily be extended to a setting where the individual income growth rates are different for each individual (HIP), while their model cannot. In addition, my model delivers a simple story of two distinct regimes that are connected by a smooth transition function instead of the multiple estimates that they had for which they needed to reduce their number by fitting an age polynomial. Also, the results are different: while they show that there are significant differences in the persistence and the variances, my results indicate that even though the persistence is lower, the bulk of the action is in the innovations.

Finally, the most important distinction between this chapter and Karahan and Ozkan's is that my model could take into account other variables besides age to explain the patterns seen in the data. This allows me to take a stand on different economic theories.

The chapter continues with a description of the statistical model. In Section III the estimation strategy is explained. Section IV describes the data set, and the selection criteria for the sample used. Section V presents the results and Section VI concludes.

2.2 Statistical Model

Let $y^i_{h(t),t}$ be the logarithm of labor earnings for an individual i of age $h(t)$ at time t. Individual i first appears in the PSID at time t_i, and leaves the sample after time T_i, so $t \in \{t_i, t_i+1, \ldots, T_i\}$. Labor age $h(t)$ is normalized to 0 for calendar age 18, and $h(t) \in \{1, \ldots, H\}$, where H is the maximum normalized labor age in the sample (which corresponds to the calendar age 64).

The logarithm of labor earnings $y^i_{h(t),t}$ depends on the vector $x^i_{h(t),t}$ that controls for the time or cohort effects, and the observable characteristics[1] of an individual i of age $h(t)$ at time t, a temporary shock $\omega^i_{h(t),t}$, and a persistent shock $\varepsilon^i_{h(t),t}$:

$$y^i_{h(t),t} = x^{i\prime}_{h(t),t}\beta + \varepsilon^i_{h(t),t} + \omega^i_{h(t),t}$$

where $\beta \in \Re^k$ and $\omega^i_{h(t),t}/\sigma^2_t \sim N(0, \sigma^2_t)$. The temporary shock reflects the measurement error present in the sample, and the temporary changes in the worker labor productivity. In the case of the persistent component of the innovations, the simplest assumption is to assume that $\varepsilon^i_{h(t),t}$ follows an ordinary AR process:

$$\varepsilon^i_{h(t+1),t+1} = \rho\varepsilon^i_{h(t),t} + \eta^i_{t+1}$$

I am interested in studying an alternative to this simple specification: the possibility that the persistent changes over the life cycle follow a nonlinear structure given by the presence of two regimes in $\varepsilon^i_{h(t+1),t+1}$:

[1]This treatment effect consists of regressing the logarithm of labor earnings $y^i_{h(t),t}$ on variables such as the level of education, a cubic polynomial for age, the interaction between the polynomial of age and education, and dummy variables for race, immigration and marital status.

$$\varepsilon^i_{h(t+1),t+1} = \begin{cases} \rho_1 \varepsilon^i_{h(t),t} + \eta^{1,i}_{t+1}, \text{ when } \tau^i_t \in (-\infty, c] \\ \rho_2 \varepsilon^i_{h(t),t} + \eta^{2,i}_{t+1}, \text{ when } \tau^i_t \in (c, +\infty) \end{cases}$$

or equivalently,

$$\varepsilon^i_{h(t+1),t+1} = I\left(\tau^i_t \in (-\infty, c]\right)\left(\rho_1 \varepsilon^i_{h(t),t} + \eta^{1,i}_{t+1}\right) + I\left(\tau^i_t \in (c, +\infty)\right)\left(\rho_2 \varepsilon^i_{h(t),t} + \eta^{2,i}_{t+1}\right) \tag{2.1}$$

The variable τ^i_t is an individual specific variable[2] and, depending on its value with respect to the threshold, c, triggers the switch of regimes. The nuisance terms $\eta^{1,i}_{t+1}$ and $\eta^{2,i}_{t+1}$ are independent, with the same mean but different precision.

The model 2.1 is a Threshold Autoregressive Model (TAR), and was introduced by Tong (1978), and further developed by Tong and Lim (1980), Tsay (1989), and Tong (1993).

The abrupt change between regimes that happens in a TAR model adds an extra degree of difficulty in maximum likelihood estimation, because the function is not differentiable and the usual maximization techniques cannot be applied. What is more, from an economic point of view, it is reasonable to think of a smooth function that connects both regimes. Therefore, I use the following formulation:

$$\varepsilon^i_{h(t+1),t+1} = \left(1 - G\left(\gamma, c, \tau^i_t\right)\right)\rho_1 \varepsilon^i_{h(t),t} + G\left(\gamma, c, \tau^i_t\right)\rho_2 \varepsilon^i_{h(t),t} + \eta^i_{t+1} \tag{2.2}$$

The specification 2.2 is called a **Smoothed Transition Autoregressive model of order 1 (STAR(1))**. It was first proposed by Chan and Tong (1986) and

[2]Some examples of τ^i_t are age, level of income, or number of weeks unemployed, to a name a few possible possibilities.

extensively studied by Luukkonen, Saikkonen and Teräsvirta (1988), and Teräsvirta (1994)[3].

My choice for the transition function is a logistic cumulative distribution and, as a result, my statistical model is also called a **Logistic Smoothed Transition Autoregressive model of order 1 (LSTAR(1))**.

Let $G(\gamma, c, \tau)$ be the logistic cumulative distribution function:

$$G(\gamma, c, \tau) = \frac{1}{1 + \exp\left[-\gamma\left(\tau_t^i - c\right)\right]}$$

where $\gamma \in (0, \infty)$ is a smoothing parameter that has the following property: the higher is γ, the less smooth the transition function. For instance, if $\gamma \to 0$, the weight function $G(\gamma, c, \tau) \to 1/2$ and there is an indeterminacy problem, because $\rho_2 - \rho_1$ can take any value. On the other hand, if $\gamma \to +\infty$, then $G(\gamma, c, \tau) \to I(\tau_t^i - c > 0)$, and $G(\gamma, c, \tau)$ becomes an indicator function, and the model is reduced to a TAR model.

I can rewrite the autoregressive component as:

$$\varepsilon^i_{h(t+1),t+1} = \varepsilon^i(\gamma, c)_{h(t),t}\, \widetilde{\rho} + \eta^i_{t+1}.$$

where $\varepsilon^i(\gamma, c)_{h(t),t} = \left(\varepsilon^i_{h(t),t}, G(\gamma, c, \tau^i_t)\, \varepsilon^i_{h(t),t}\right)$ and $\widetilde{\rho} = (\rho_1, \rho_2 - \rho_1)'$. Then, the LSTAR(1) model can be expressed as:

$$\begin{cases} y^i_{h(t),t} = x^{i\prime}_{h(t),t}\beta + \varepsilon^i_{h(t),t} + \omega^i_{h(t),t} \\ \varepsilon^i_{h(t+1),t+1} = \varepsilon^i(\gamma, c)_{h(t),t}\, \widetilde{\rho} + \eta^i_{t+1} \end{cases} \quad (2.3)$$

[3] For a good review on nonlinear time series, see Bauwens, Lubrano and Richard (2000), Dijk, Teräsvirta and Franses (2002), and Korenok (2009).

The nuisance term η_{t+1}^{i} in 2.3 is a linear combination of $\eta_{t+1}^{1,i}$ and $\eta_{t+1}^{2,i}$ from 2.1. I want the variance of η_{t+1}^{i} to be a convex combination of the variances of the two regimes. Therefore,

$$\eta_t^i = (\kappa_i)^{-.5}\,\xi_t^i\sqrt{1-G\left(\gamma,c,\tau_t^i\right)+\phi G\left(\gamma,c,\tau_t^i\right)}$$

with ϕ being the ratio of the variance of the second regime with respect to the first regime. The distribution of η_t^i needs to be flexible enough to accommodate the particular characteristics of the PSID (see Lillard and Willis, 1978).

As any continuous distribution can be well approximated with a finite Gaussian mixture[4], I assume that

$$\xi_t^i \sim \sum_{j=1}^{m} p_{g_j} N\left(0, g_j^{-1}\right)$$

Let $s_t^i \in \{1,\ldots,m\}$ be an indicator function for the individual i at time t, which shows the normal distribution from which the shock received has been drawn from. Therefore, the variance of η_t^i can be expressed as:

$$\begin{aligned} Var\left(\eta_t^i/s_t^i\right) &= \left(\kappa_i g_{s_t^i}\right)^{-1}\left[\left(1-G\left(\gamma,c,\tau_t^i\right)\right)+\phi G\left(\gamma,c,\tau_t^i\right)\right] \\ Var\left(\eta_t^i/s_t^i\right) &= \left(\kappa_i g_{s_t^i}\right)^{-1} k_t^i\left(\gamma,c,\phi\right). \end{aligned}$$

where $k_t^i(\gamma,c,\phi) \equiv (1-G(\gamma,c,\tau_t^i))+\phi G(\gamma,c,\tau_t^i)$. I assume that the first age-period shocks are given by:

$$\eta_{1,t}^i = \lambda^{-.5}\left(\kappa_i^{-.5}\xi_t^i\sqrt{1-G\left(\gamma,c,\tau_t^i\right)+\phi G\left(\gamma,c,\tau_t^i\right)}\right)$$

[4]For a detailed exposition of different applications of finite mixture models, see Everitt and Hand (1981), Titterington, Smith and Makov (1985), and McLachlan and Peel (2000).

where the factor λ allows me to capture the fact that a considerable size of the lifetime earnings inequality is realized before the individual enters into the labor market (see Keane and Wolpin, 1997; Storesletten, Telmer and Yaron, 2004; and Huggett, Ventura and Yaron, 2011).

As can be seen from 2.3, at any point in time, the resulting process is a moving weighted average of the different regimes, with the weights changing over the life cycle.

2.2.1 RIP versus HIP

In the empirical income macro literature, two types of models are widely used to model the labor income dynamics: the Restricted Income Profile (RIP), and the Heterogeneous Income Profile (HIP) model. The distinction between the two lies in the growth rates of income: a RIP model assumes that the effect on income of an extra year of labor market experience[5] is common to all individuals, while a HIP assumes that the effect is individual-specific. This unobservable individual-specific effect gives a random-effect structure to the model instead of the RIP fixed-effect structure.

This individual heterogeneity can be easily extended to the cubic polynomial of labor market experience, but as pointed out by Baker (1997) and Guvenen (2009), it does not substantially improve the model fit. The consensus is to keep the HIP model with a simple structure and assume that the effects on income of the powers of labor market experience are common to all agents.

[5] Alternatively, age can be used instead of labor market experience.

The result is that a HIP model has lower persistence in the AR(1) process than a RIP model, and the consequences are not trivial for the calibration of a macro model with incomplete markets.

A RIP model can be summarized as follows:

$$\begin{aligned} y^i_{h(t),t} &= x^{i\prime}_{h(t),t}\beta + \varepsilon^i_{h(t),t} + \omega^i_{h(t),t} \\ y^i_{h(t),t} &= \alpha + h\left(t\right)\beta_1 + \widetilde{x}^{i\prime}_{h(t),t}\beta_2 + \varepsilon^i_{h(t),t} + \omega^i_{h(t),t} \end{aligned}$$

where $x^i_{h(t),t} = \left(1, h_i, \widetilde{x}^{i\prime}_{h(t),t}\right)'$, and $\beta = (\alpha, \beta_1, \beta_2)'$.

While a HIP is defined as:

$$\begin{aligned} y^i_{h(t),t} &= \alpha^i + h\left(t\right)\beta^i_1 + \widetilde{x}^{i\prime}_{h(t),t}\beta_2 + \varepsilon^i_{h(t),t} + \omega^i_{h(t),t} \\ y^i_{h(t),t} &= x^{i\prime}_{h(t),t}\beta^i + \varepsilon^i_{h(t),t} + \omega^i_{h(t),t} \end{aligned}$$

where $\beta^i = (\alpha^i, \beta^i_1, \beta_2)'$.

Even though the economic rationale favors a HIP model[6], the verdict is not conclusive from a statistical point of view (see Abowd and Card, 1989; and Hryshko, 2012). In this chapter, I focus my analysis on a RIP model with age-dependent shocks.

2.3 Estimation

I study three models in this chapter: a conventional RIP model, a RIP model with constant persistence and heterogeneity in the innovations, and a RIP model with heterogeneity in both the persistence and the innovations.

[6]In a model with human capital accumulation, the differences in ability translate into individual-specific growth rates of income.

In each model, for the treatment effects of the logarithm of annual earnings I include a cubic polynomial of age to capture the hump shape of mean earnings in the life cycle, the logarithm of GDP per capita to control for the effects of any aggregate shock in the economy[7], and six educational categories:

- Elementary school (less than 8 years of education and Grades K-8)
- Some high school (9-11 years of education and Grades 9-11)
- High school graduate (12 years of education and Grade 12)
- Some college (13-15 years of education)
- College graduate (16 years of education)
- Graduate school (more than 16 years of education)

In the case of the conventional RIP, after controlling for the effects of the observable characteristics, I identify the relevant moment conditions in the autocovariance matrix and proceed with the GMM estimation in the usual manner. For the last two models, I follow a Bayesian approach and sample their posterior distributions with Markov Chain Monte Carlo techniques.

2.3.1 Priors

My choices for prior distributions are centered in conjugate and uninformative priors, depending on the information that I have available. The selection is standard. Appendix A includes a complete description and references to the values of the

[7]Instead of choosing the GDP per capita as a regressor to control for the effects of the aggregate shocks in the economy, I could have controlled for time or cohort effects. These last two options are common takes in the literature.

hyperparameters. This is the list of the prior distributions:

1. $\omega^i_{h(t),t}$ includes the measurement error and the transitory component of the agent's productivity. I assume that $\omega^i_{h(t),t}/\sigma^2_t \sim N(0, \sigma^2_t)$, and $\underline{s}^2_\omega (\sigma^2_t)^{-1} \sim \chi^2(\nu_\omega)$.

2. For the threshold variable c and the factor of proportionality ϕ, I choose an uninformative prior distribution: $p(c) \propto 1$ and $p(\phi) \propto \frac{1}{\phi}$.

3. The smoothing parameter γ is distributed as a truncated Cauchy distribution:

$$p(\gamma) \propto \begin{cases} (1+\gamma^2)^{-1} & , \gamma > 0 \\ 0 & , \ o.w. \end{cases} .$$

4. I choose conjugate priors for s^i_t and $\underline{\mathbf{p}}$: $s^i_t/\underline{\mathbf{p}} \sim Multinomial(\underline{\mathbf{p}})$, and $\underline{\mathbf{p}} \sim Dir(\alpha)$, where $\alpha{=}(\alpha_1, \ldots, \alpha_m)'$. Along the same lines, $\xi^i_t/s^i_t = k \sim N(0, g_k^{-1})$, with $1 \leq k \leq m$. $\underline{s}^2 g_j \sim \chi^2(\underline{\nu}_g)$.

5. The persistence $\widetilde{\rho}$ has a normal distribution with mean $\underline{\mu}_{\widetilde{\rho}}$ and precision $h_\rho I$, i.e. $\widetilde{\rho} \sim N\left(\underline{\mu}_{\widetilde{\rho}}, h_\rho^{-1} I\right)$.

6. $\beta/(\mu, H) \sim N(\mu, H^{-1})$, where $\mu \sim N\left(\underline{\mu}_\mu, \underline{H}_\mu^{-1}\right)$and $H \sim Wishart(\underline{\nu}_H, \underline{S}_H)$

7. $s^2_\kappa \kappa_i \sim \chi^2(\underline{\nu}_\kappa)$ and $s^2_\lambda \lambda \sim \chi^2(\nu_\lambda)$.

2.3.2 Posterior Distribution

The full posterior distribution of this model is the product of the LSTAR likelihood and priors 1-7:

$$p\left\{\left(y^*, \{\varepsilon^i_{h(t),t}\}^{I,T_i}_{i=1,t=t_i}, \beta, \{\sigma^2_t\}^T_{t=1}, \rho_1, \rho_2, \gamma, c, \{\kappa_i\}^I_{i=1}, \lambda, \mu, H, \mathbf{p}, \{g_j\}^m_{j=1}, \{s^i_t\}^{I,T_i}_{i=1,t=t_i}\right)/y\right\} \tag{2.4}$$

where $y_i = \left(y^i_{h(t_i)t_i}, \ldots, y^i_{h(T_i)T_i}\right)'$ and $y = (y'_1, \ldots, y'_I)'$, as well as $y^*_i = \left(y^{*,i}_{h(t_i)t_i}, \ldots, y^{*,i}_{h(T_i)T_i}\right)'$ and $y^* = (y'^*_1, \ldots, y'^*_I)'$, with $y^{*,i}_{h(t),t}$ being an imputed value, in case the value is missing or top-coded, i.e.

$$y^i_{h(t),t} = \begin{cases} y^i_{h(t),t}, & \text{when the observation is present in the sample;} \\ y^{*,i}_{h(t),t}, & \text{when the observation is not present in the sample;} \end{cases}$$

The object 2.4 has all the necessary information to infer the income dynamics in the sample.

2.3.3 Markov Chain Monte Carlo Methods

The Markov Chain Monte Carlo Methods (MCMC) are simulation-based techniques that consist of generating random draws that are neither independent nor identically distributed, but rather form a Markov Chain that converges to an invariant distribution that is the one under investigation (see Robert and Casella, 2010). Two widely used MCMC methods to analyze a posterior distribution are the Gibbs Sampler and the Metropolis-Hastings algorithm.

2.3.3.1 The Gibbs Sampler

The Gibbs Sampler (Geman and Geman, 1984) consists of partitioning 2.4, and sampling each partition conditional on the other ones. A simple example will illustrate the concept.

Suppose that the object of interest is $p(\theta \mid y)$, the posterior distribution of θ, with $\theta \in \Re^k$ the parameter vector, and $y \in \Re^n$ the vector of observations. The vector θ is split into the partitions θ_0 and θ_1. Let $\theta^{(0)} = \left(\theta_0^{(0)}, \theta_1^{(0)}\right)'$ be the starting point of the algorithm, and for each iteration $m \in \{1, \ldots, B\}$ of the Gibbs sampler, new values are drawn and the partitions are updated iteratively in the following way: $\theta_0^{(m)} \sim p\left(\theta_0 \mid \theta_1^{(m-1)}, y\right)$ and $\theta_1^{(m)} \sim p\left(\theta_1 \mid \theta_0^{(m)}, y\right)$. Then,

$$p\left(\theta^{(m)} \mid \theta^{(m-1)}, y\right) = p\left(\theta_0^{(m)} \mid \theta_1^{(m-1)}, y\right) p\left(\theta_1^{(m)} \mid \theta_0^{(m)}, y\right)$$

For a sufficiently large B, this Markov Chain converges to an invariant distribution that coincides with the distribution of interest $p(\theta \mid y)$.

Given the parameters of this model, I find it convenient to consider 15 blocks. Of these, 12 are amenable to Gibbs sampling because the conditional distributions are simple and follow known distributions. However, the other three blocks have distributions from which random samples cannot be generated directly, and a Metropolis-Hastings step within the Gibbs sampler is required. For complete algebraic derivations of the posterior distributions of each partition, see Appendix B.

The 12 Gibbs Sampling steps are:

1. Sample y^* conditional on $\beta, \{\sigma_t^2\}_{t=1}^T, \left\{x_{h(t),t}^i\right\}_{i,t}, \{s_t^i\}_{i=1,t=t_i}^{I,T_i}, \{g_j\}_{j=1}^m, \rho_1, \rho_2, \gamma,$ c, ϕ, and y.
2. Sample $\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}$ conditional on $y, y^*, \beta, \left\{x_{h(t),t}^i\right\}_{i,t}, \{\sigma_t^2\}_{t=1}^T, \rho_1, \rho_2, \gamma,$ $c, \{\kappa_i\}_{i=1}^I, \lambda, \phi, \{s_t^i\}_{i=1,t=t_i}^{I,T_i}$, and $\{g_j\}_{j=1}^m$.
3. Sample β conditional on $y, y^*, \left\{x_{h(t),t}^i\right\}_{i,t}, \left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}, \mu, H$, and $\{\sigma_t^2\}_{t=1}^T$.
4. Sample $\{\sigma_t^2\}_{t=1}^T$ conditional on y, y^*, $\left\{x_{h(t),t}^i\right\}_{i,t}, \left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}$, and β.
5. Sample $\widetilde{\rho}$ conditional on $\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}, \gamma, c, \lambda, \{\kappa_i\}_{i=1}^I, \{s_t^i\}_{i=1,t=t_i}^{I,T_i}, \phi,$ and $\{g_j\}_{j=1}^m$.

6. Sample $\{\kappa_i\}_{i=1}^{I}$ conditional on $\left\{\varepsilon_{h(t),t}^{i}\right\}_{i=1,t=t_i}^{I,T_i}, (\rho_1, \rho_2), \lambda, (\gamma, c), \phi, \{s_t^i\}_{i=1,t=t_i}^{I,T_i}$, and $\{g_j\}_{j=1}^{m}$.

7. Sample λ conditional on $\left\{\varepsilon_{1,t_i}^{i}\right\}_{i=1}^{I}, \{\kappa_i\}_{i=1}^{I}, \left\{s_{t_i}^{i}\right\}_{i=1}^{I}$, and $\{g_j\}_{j=1}^{m}$.

8. Sample μ conditional on β, and H.

9. Sample H conditional on β, and μ.

10. Sample $\mathbf{p}$ conditional on $\{s_t^i\}_{i=1,t=t_i}^{I,T_i}$.

11. Sample $\{s_t^i\}_{i=1,t=t_i}^{I,T_i}$ conditional on $\mathbf{p}$, $\left\{\varepsilon_{h(t),t}^{i}\right\}_{i=1,t=t_i}^{I,T_i}, \{\kappa_i\}_{i=1}^{I}, \lambda, \rho_1, \rho_2, \gamma, c, \phi$, and $\{g_j\}_{j=1}^{m}$.

12. Sample $\{g_j\}_{j=1}^{m}$ conditional on $\{s_t^i\}_{i=1,t=t_i}^{I,T_i}, \left\{\varepsilon_{h(t),t}^{i}\right\}_{i=1,t=t_i}^{I,T_i}, \{\kappa_i\}_{i=1}^{I}, \rho_1, \rho_1, \lambda, \gamma, c$, and ϕ.

2.3.3.2 Metropolis-Hastings

There are three posterior distributions (3 blocks) that still need to be sampled in order to have the full posterior distribution of the model. One way to do it is with a Metropolis-Hastings algorithm (Metropolis et al., 1953; and Hastings, 1970).

The Metropolis-Hastings is a MCMC method that consists of drawing values from an arbitrary density $q(\cdot)$, and accepting only those draws that are more likely to be generated from the density $p(\cdot)$ under study. Under some regularity conditions, the collection of accepted draws approximates well $p(\cdot)$.

Formally, let $D \in \Re^k$ be the support of the target distribution $p(x)$ with $x \in D$, and let $q(x^*/x)$ with $x^* \in D$ be the arbitrary proposal density. At the iteration $m \in \{1, \ldots, N\}$ of the algorithm and given $x = x^{(m-1)}$, the probability of accepting a candidate draw x^* is

$$\alpha\left(x^*, x\right) = \min\left\{\frac{p\left(x^*\right)}{p\left(x\right)}\frac{q\left(x/x^*\right)}{q\left(x^*/x\right)}, 1\right\}. \tag{2.5}$$

If x^* is accepted, $x^{(m)} = x^*$, otherwise $x^{(m)} = x^{(m-1)}$. A closer inspection of 2.5 shows that a draw x^* is more likely to be accepted if the chances to move from x to x^* are lower, and vice versa.

The Metropolis-Hastings acceptance criterion 2.5 ensures that a reversibility condition is satisfied, and this is a sufficient condition to prove the convergence of the Markov Chain, given by the collection of $x^{(m)}$ with $m \in \{1, \ldots, N\}$. For a textbook treatment of this MCMC method, see Chib and Greenberg (1995), and Geweke (2005).

In the solution of this model, I sample these three posterior distributions with proposal densities that have the common property that $q\left(x^*/x\right)$ is independent from x[8]. This means that $q\left(x^*/x\right) = q\left(x^*\right)$, and the acceptance criterion 2.5 is reduced to

$$\alpha\left(x^*, x\right) = \min\left\{\frac{p\left(x^*\right)}{p\left(x\right)}\frac{q\left(x\right)}{q\left(x^*\right)}, 1\right\}.$$

The Independent Metropolis-Hastings algorithms for the objects

$$\left(\gamma/\left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i}, c, \left(\rho_1, \rho_2\right), \lambda, \phi, \left\{\kappa_i\right\}^I_{i=1}, \left\{s^i_t\right\}^{I,T_i}_{i=1,t=t_i}, \left\{g_j\right\}^m_{j=1}\right)'$$

and

$$\left(\phi/\left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i}, \rho_1, \rho_2, \gamma, c, \lambda, \left\{\kappa_i\right\}^I_{i=1}, \left\{s^i_t\right\}^{I,T_i}_{i=1,t=t_i}, \left\{g_j\right\}^m_{j=1}\right)'$$

---

[8] A Metropolis-Hastings algorithm with this feature is also referred to as "Independent Metropolis-Hastings"

consists of draws taken from truncated normal distributions, with $\gamma^* \sim N\left(\mu_\gamma, \sigma_\gamma^2\right) I\left(a_\gamma, b_\gamma\right)$ and $\phi^* \sim N\left(\mu_\phi, \sigma_\phi^2\right) I\left(a_\phi, b_\phi\right)$.[9]

In the case of

$$\left(c / \left\{\varepsilon_{h(t),t}^{i}\right\}_{i=1,t=t_i}^{I,T_i}, \rho_1, \rho_2, \gamma, \phi, \lambda, \left\{\kappa_i\right\}_{i=1}^{I}, \left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i}, \left\{g_j\right\}_{j=1}^{m}\right)',$$

the variable c^* is sampled from a discrete distribution in the support $[a_c, b_c]$, with the vector of probabilities $\mathbf{p_c}$. The parameters of the proposal densities are chosen in such way that the acceptance probabilities are close to 30%. This ensures a good mixing of the Markov Chain (see Müller, 1991; and Canova, 2007).

2.4 Panel Study of Income Dynamics (PSID)

The PSID is a rich panel dataset that has information for more than 9,000 families and 70,000 individuals, who have been followed since 1968. It is the largest and longest household-based survey in the world and an excellent starting point to study income dynamics in the United States.

My sample consists of male heads of households in the time period between 1968 and 1996, including those individuals satisfying the following conditions:

1. Has at least two consecutive earning observations;
2. Is between 20 and 64 years old at the time of the survey;
3. Does not belong to the Survey of Economic Opportunities sample (SEO);

---

[9] There are other ways to sample these objects. For instance, Lopes and Salazar (2006) use a random-walk Metropolis chain to update (γ, ϕ, c), with $\gamma^* \sim \Gamma\left(\frac{\left(\gamma^{(i)}\right)^2}{\Delta_\gamma}, \frac{\gamma^{(i)}}{\Delta_\gamma}\right)$, $\phi^* \sim \Gamma\left(\frac{\left(\phi^{(i)}\right)^2}{\Delta_\phi}, \frac{\phi^{(i)}}{\Delta_\phi}\right)$, and $c^* \sim N\left(c^{(i)}, \triangle_c\right) I\left(a_c, b_c\right)$.

4. Has positive values for hours and labor income[10];

5. Has hourly labor income between $2 and $400 in 1993 dollars[11];

6. Has worked between 520 and 5,110 hours every year[12];

7. Has no missing values for years of education, nor any inconsistencies;

One particular feature of the PSID that needs to be taken into account: the income reported is the income earned in the year before the survey takes place, while all other variables refer to the same year. This mismatch makes things more complicated in the first wave in 1968, where the income for 1967 is known but not the values taken in that year for all the other variables. To a lesser extent, a similar situation happens after 1997, when the PSID starts to have a biannual periodicity. Therefore, I exclude the 1968, 1999-2011 waves from the sample.[13]

I start with a sample of 4,887 individuals and, after checking for any inconsistency, my final sample is reduced to 4,548 individuals with 54,622 individual-year observations. On average, an individual is followed for 12 years. As shown in Table 2.1,

---

[10] The variables taken from the PSID for labor income are: V74, V514, V1196, V1897, V2498, V3051, V3463, V3863, V5031, V5627, V6174, V6767, V7413, V8066, V9376, V8690, V11023, V12372, V13624, V14671, V16145, V17534, V18878, V20178, V21484, V23323, (ER4140+ER4117+ER4119), (ER6980+ER6957+ER6959), (ER9231+ER9208+ER9210), and (ER12080+ER12065+ER12193).

[11] To deflate the labor income series, I use the CPI from the US Department Of Labor, Bureau of Labor Statistics available at ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt

[12] The purpose of having an extreme value such as 5,110 hours is to discard those observations that are certainly misreported. To work 5,110 in a year implies working 16 hours a day, 6 days a week. It is unlikely that anyone can work that much.

[13] The situation can be improved for the 1999-2011 waves by imputing the missing values for income. A Bayesian approach is more convenient for dealing with situations of this sort.

Table 2.1: Summary Statistics

5^{th} *Percentile* :	\$8,999
25^{th} *Percentile* :	\$19,631.52
Median:	\$30,072.29
Mean:	\$35,750
75^{th} *Percentile* :	\$42,756.95
95^{th} *Percentile* :	\$77,935.34
Std. deviation:	\$33,448.60
Interquantile Range:	\$23,125.43
Number of individuals:	4,548
Number of observations:	54,622

the mean and median income in the sample are \$35,750 and \$30,072, with a standard deviation and interquantile range of \$33,448 and \$23,125 in 1993 dollars .

The earnings distribution has the typical right asymmetry for an income variable, and a closer inspection of the medians in the boxplots, depicted in red in Figure 2.1, shows the typical hump shape for earnings in the life cycle. The interquantile range, given by the size of the boxplots, is a robust measure of the income variability along the life cycle. It grows until individuals are in their mid-30s, remains approximately constant until they reach their 40s, and starts to shrink thereafter.

A plausible economic interpretation is that these differences in the interquantile ranges are a consequence of the individuals' decisions to accumulate human capital in the life cycle (see Ben-Porath, 1967). Another hypothesis is that this variability in the interquantile ranges is due to occupational mobility in the life cycle, with occupation-specific human capital (see Kambourov and Manovskii, 2009) and

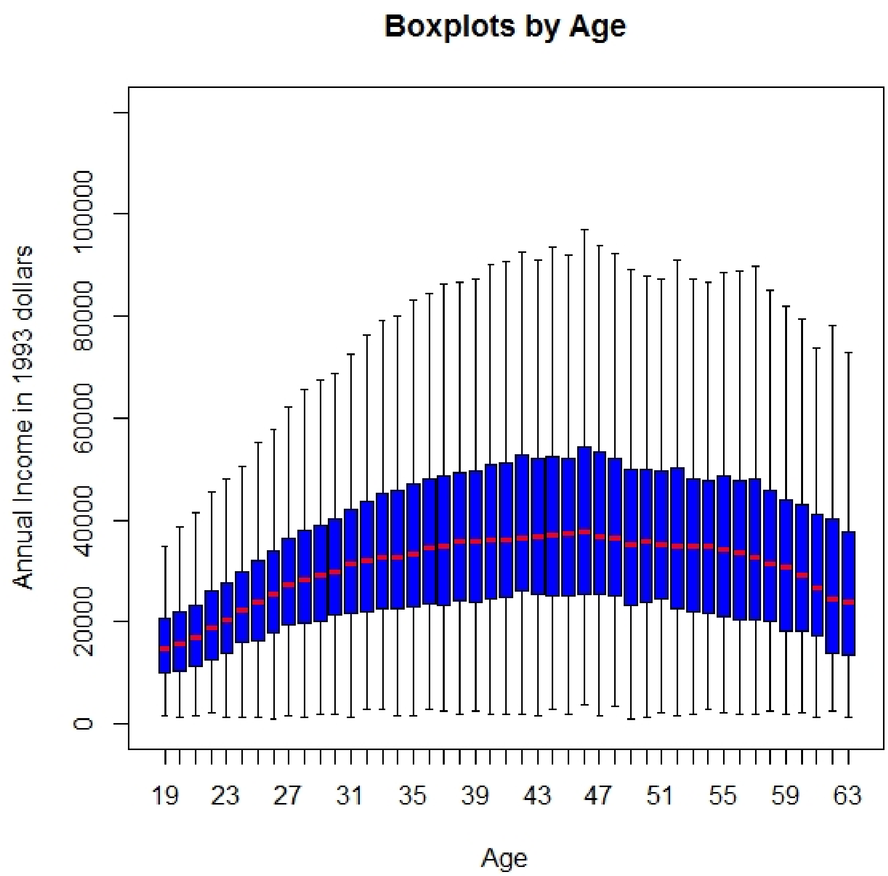

Figure 2.1: Boxplots by Age.

occupations having a direct relationship between risk and return (see Cubas and Silos, 2012). Changes in the intra-household allocation (see Greenwood, Guner and Knowles, 2003; and Knowles, 2012) or the "boomerang" of moving back to one's parent house as means of insurance against bad shocks (Kaplan, 2012) could be other possible explanations.

Whether acquiring human capital, modulating risk, or intra-family dynamics explains Figure 2.1, it is clear that to model income dynamics, the forces operating in the life cycle must be taken into account.

2.5 Results

The size of the MCMC sample for the LSTAR(1) model is 20000 and seems to converge in 5000 iterations. After discarding the initial 5000, a visual inspection of the different trace plots shows a good mixing (see Figures 2.2 and 2.3). In the case of the RIP model with constant persistence and heterogeneity in the innovations, the MCMC sample is 14000 and seems to converge after 4000 iterations. These initial 4000 iterations are the burn-in phase of the algorithm and are thrown away. I repeat the sampling from the posterior distributions several times, at different dispersed starting points (Gelman and Rubin, 1992) to test for convergence, and the results are favorable.

2.5.1 Bayesian RIP model

The benchmark and starting point is a conventional RIP model estimated with GMM techniques. This is a standard model in the empirical macro literature and is

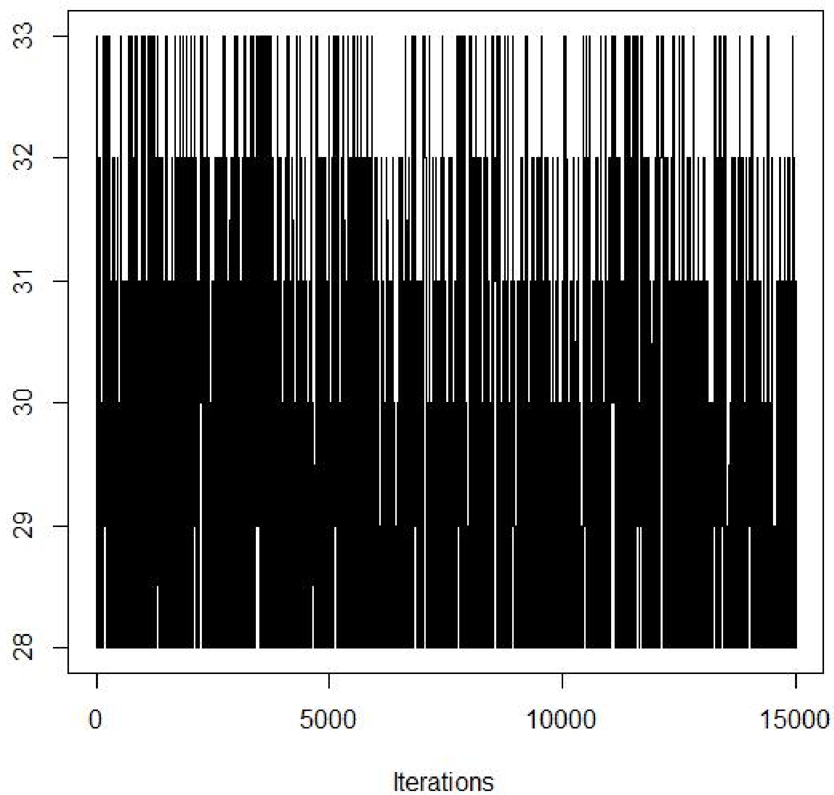

Figure 2.2: Trace plot for the threshold variable c.

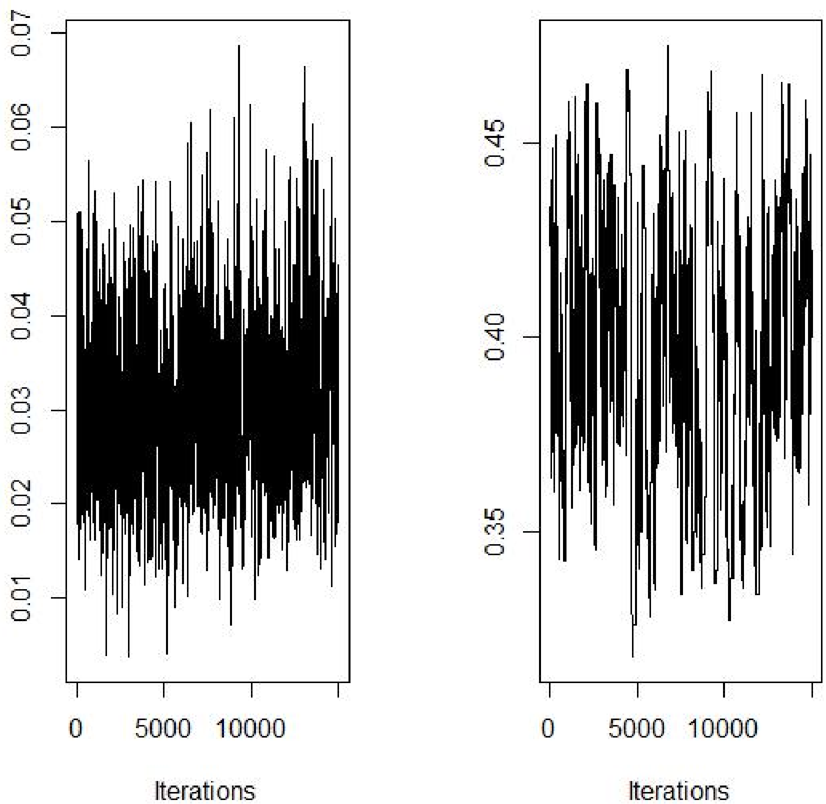

Figure 2.3: Trace plots for γ (left) and ϕ (right).

Table 2.2: GMM

Parameters	**Values**
ρ	0.96
σ_α^2	0.043
σ_η^2	0.11

generally used for the calibration of dynamic general equilibrium models.

Table 2.2 shows that the idiosyncratic earnings shocks in the life cycle are persistent with a persistence parameter ρ equal to 0.96. This stationary AR(1) process is close to a unit root process, and this high persistence has the practical implication that an arbitrary shock is halved after 17 periods. This means that, assuming a working life of 40 years, a worker spends more than 40% of his working life with the effects of a shock that has not yet reduced in half.

Even though this model assumes that the income growth rate is common to all agents, the starting point in the level of earnings is different for all workers. This is represented by an intercept α which has a variance σ_α^2; according to my estimations, it is equal to 0.043. The innovations in the residual earnings have a variance σ_η^2 equal to 0.11. The sum of σ_α^2 and σ_η^2 gives the variance of the idiosyncratic earnings shocks for the first-age period without the transitory component of the variance. In this case, this value is equal to 0.15 and means that a worker that suffers a one-standard-deviation shock has a 37% increase in their earnings with respect to the mean.[14]

[14]This increase can be explained in the following way: labor earnings in efficiency units are modeled as $e^x w\theta$, where w is the wage rate, θ is the hours worked, and $x \sim N\left(\mu, \sigma^2\right)$

Table 2.3: Persistence Parameters

Parameters	**Mean**	**Std. Dev.**
1 Regime: ρ	0.89	0.007
2 Regimes: ρ_1	0.84	0.047
2 Regimes: ρ_2	0.83	0.02

The first contender for this standard RIP is a model similar to the LSTAR(1) previously presented but with just one regime. This means that this model is a RIP with constant persistence in the life cycle and a rich structure of heterogeneity in the innovations. Table 2.3 indicates that the posterior distribution of the persistence parameter ρ has a mean of 0.89 and a standard deviation of 0.007. Moreover, as can be seen in Figure 2.4, there is a 99% probability that the posterior value of ρ is between 0.87 and 0.91. This means that the GMM value for ρ of 0.96 is not even possible in this model.

In order to put into perspective this reduction in the value of ρ, a persistence of 0.89 implies that a shock is halved in 6 periods: in contrast, it takes 17 periods for a shock to be halved in the conventional model. Naturally, this significant reduction in the persistence of the shocks has nontrivial economic implications and points out the importance of the model specification.

Why does this reduction in the value of ρ occur? Even though the purpose of this chapter is not to explain the reasons for this difference, it is interesting to point

---

is the earnings shocks. Therefore, the efficiency units e^x is a log-normal random variable with mean $e^{\mu+\sigma^2/2}$. If a worker suffers a shock with a value of one-standard-deviation from the mean μ, the increase in labor earnings is $e^{\sigma-\sigma^2/2}$.

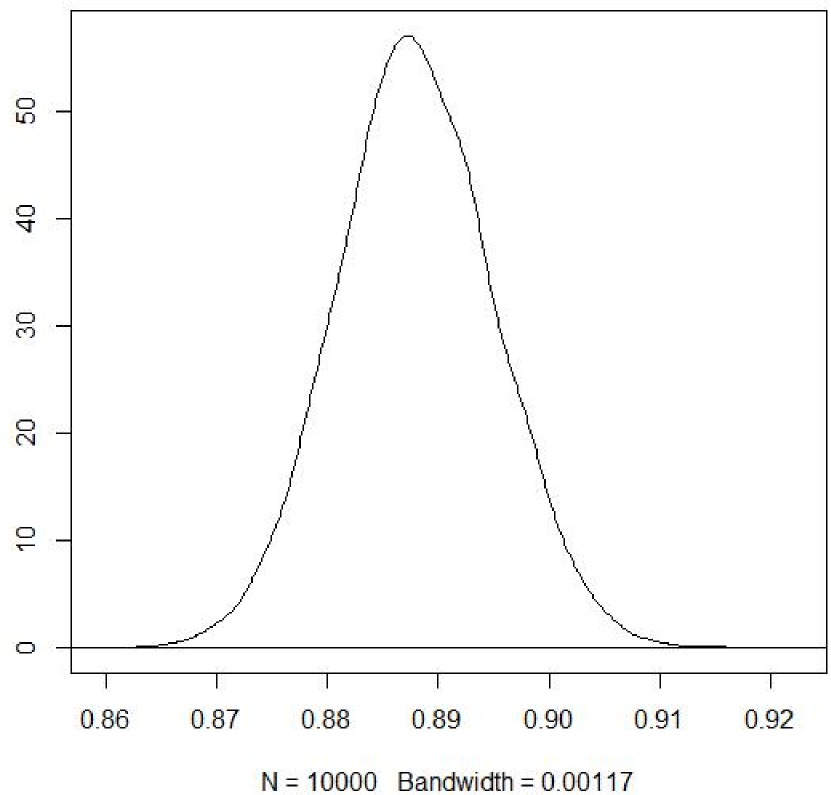

Figure 2.4: Posterior density estimation for ρ in a model with 1 Regime.

out some possible explanations. First, one might be tempted to think it is due to the innovations' structure, but unless the innovations are not orthogonal to the $\epsilon' s$, this conditional heteroscedasticity only affects the efficiency of the GMM estimator, not its unbiasedness.

Another potential explanation is related to the asymptotic distribution of the estimate of ρ, $\widehat{\rho}$. It is well known that $\widehat{\rho}$ is not asymptotically normal in an AR(1) process when $\rho \to 1$. This is one of the few cases, as pointed by Sims and Uhlig (1991), in which the Bayesian posterior probability statements do not coincide with the frequentist statements that emerge from the p-values. Sims and Uhlig (1991) simulate the joint distribution of $\widehat{\rho}$ and ρ, and they show how the distribution of $\widehat{\rho}$ conditional to ρ becomes asymmetric, with high probability on values higher than 1, as ρ is close to 1. They show that, on the contrary, the distribution of ρ conditional to $\widehat{\rho}$ remains normal and well-behaved. Then, as an exemplar for selected values of ρ, they find the prior distributions of ρ that will make the Bayesian posterior probability statements equivalent to the corresponding statements from the p-values. All of these priors puts a high probability on values of ρ higher than 1.

Therefore, in order to reach the same conclusions in the estimation of ρ in the GMM and the Bayesian approach, one needs to assume a priori that it is highly likely for ρ to be close to 1. This would be why the value of the GMM estimate is higher than the posterior mean of ρ.

The parameters g^{-1}, κ_i^{-1} and λ^{-1} are the components of the variance of the innovations. Figures 2.5 and 2.6 show the posterior distribution of λ^{-1} and g, while

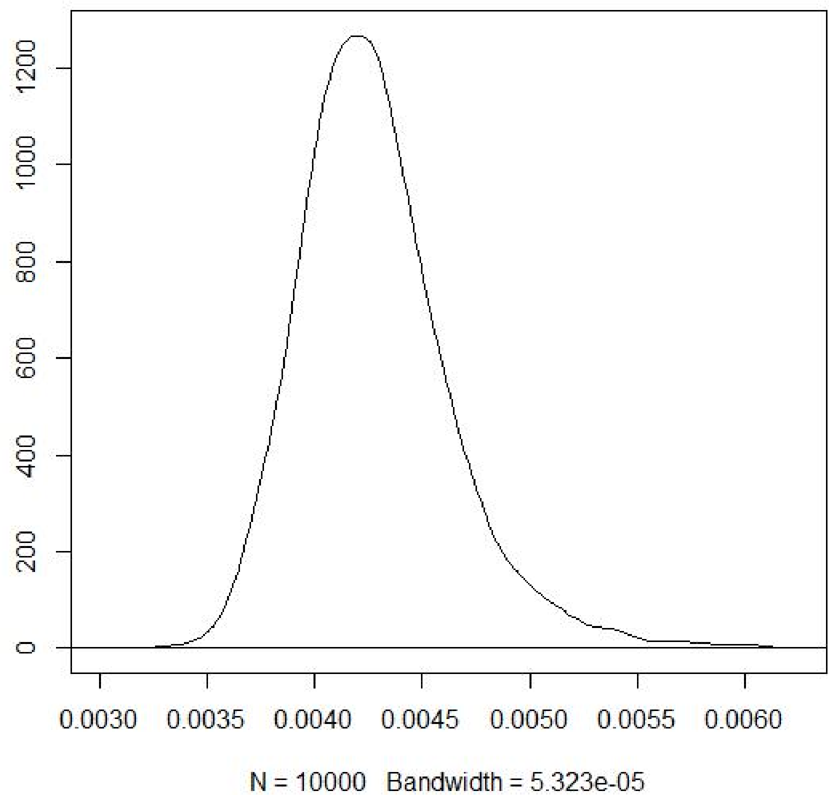

Figure 2.5: Posterior density estimation for λ^{-1} in a model with 1 Regime.

Table 2.4: Variance components (1 regime)

Parameters	**Mean**	**25%**	**50%**	**75%**
λ^{-1}	0.0043	0.004	0.00424	0.0045
κ_i^{-1}	0.517	0.372	0.441	0.608
g	0.0107	0.0103	0.0106	0.0111

Figure 2.7 refers to the boxplot of κ_i^{-1}. It is interesting to notice that 75% of the individuals in the sample have κ_i^{-1} with values less than 0.608 (see Table 2.4), but there is a proportion that has κ_i^{-1} greater than 1. This means that in the sample there is a group of individuals with a higher conditional variance than the rest. Because, it was not possible to identify this group in the conventional RIP model, there might be an upward bias in the variance estimation of the conventional RIP model.

The mean of the initial variance of the shocks in the first age-period is the product of the mean of g^{-1}, κ_i^{-1} and λ^{-1} and is equal to 0.20. In economic terms, this means that a positive one-standard-deviation shock represents 42% higher earnings with respect to the mean earnings.

Both models deliver similar stories in the dispersion of the initial shocks. However, the wide dispersion of values for κ_i^{-1}, which is not taken into account by the conventional RIP, and the differences in the value of ρ suggest some potential drawbacks in the standard specification.

2.5.2 LSTAR(1) model

In the previous subsection, the standard model used in the empirical macro literature was compared against a RIP model with constant persistence and hetero-

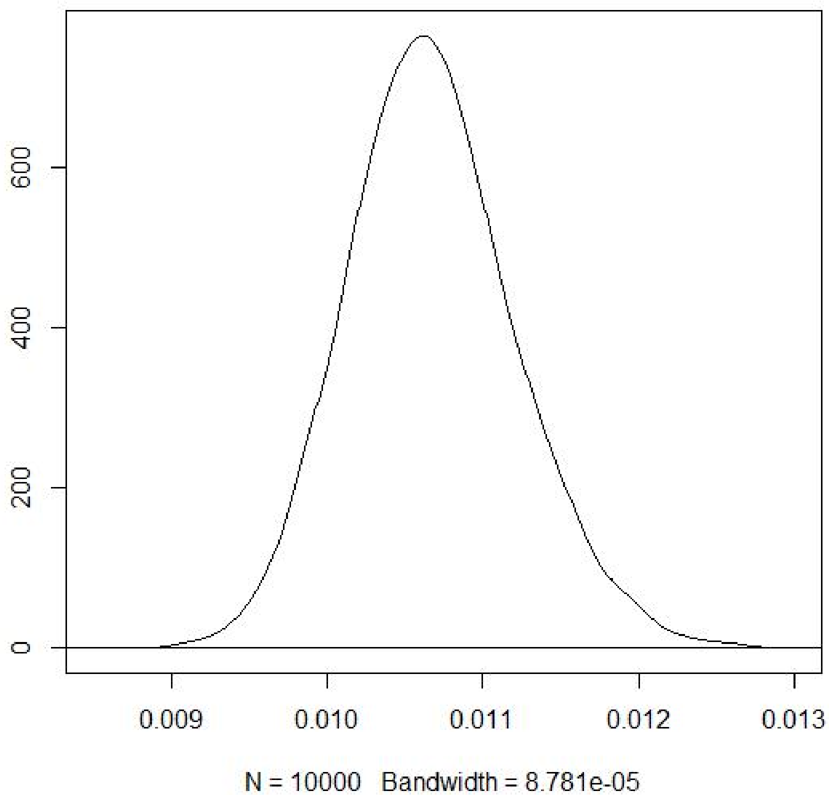

Figure 2.6: Posterior density estimation for the precision g in a model with 1 Regime.

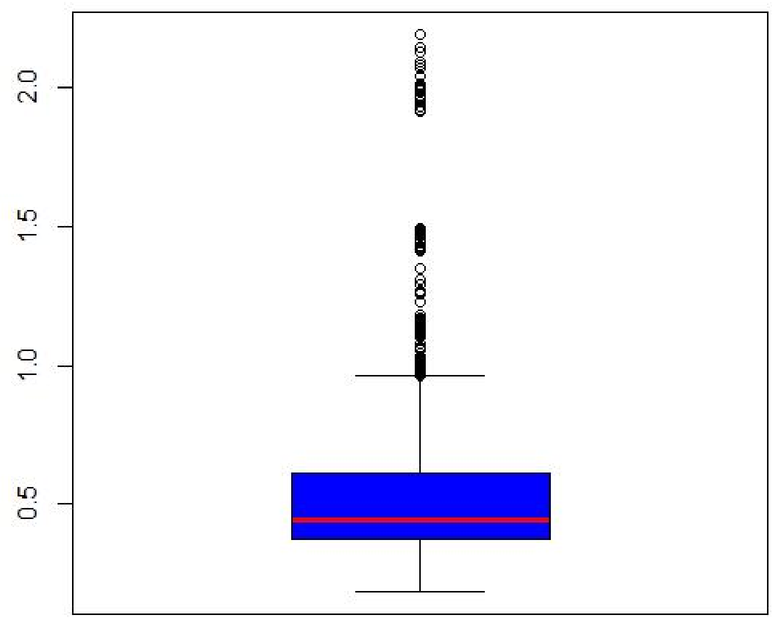

Figure 2.7: Posterior density estimation for κ^{-1} in a model with 1 Regime.

Table 2.5: Parameters of the weight function.

Parameters	Mean	Std. Dev	25%	50%	75%
c	29.73	1.609	28	29	31
γ	0.0305	0.008	0.025	0.03	0.036

geneity in the innovations. Now, the assumption of a constant persistence is left and the standard RIP is compared to a LSTAR(1) model which, by definition, exhibits heterogeneity in the persistence and the innovations.

In a LSTAR(1) model two regimes are connected with a smooth transition function. Depending on how close the variable age is from a threshold, the weights assigned to each regime change and the nature of the idiosyncratic shocks changes in every point of the life cycle.

Figure 2.8 depicts the histogram for the posterior distribution of the threshold variable c, and Table 2.5 shows the summary statistics. The histogram for c reveals that there is a probability higher than 50% that the threshold is between 28 and 29 years old, with 28 being the mode of the distribution and the most likely value.

This result means that a worker in his twenties experiences different shocks than an older worker. If we want to know the economic reasons for this, we need to inquire on the changes that happen during this time in the life of a worker. It might be the case that questions related with job mobility, labor experience or changes in the composition of the household are relevant.

The weight function that connects both regimes plays an important role in the estimation results. The parameter γ controls the smoothness of this function and, in

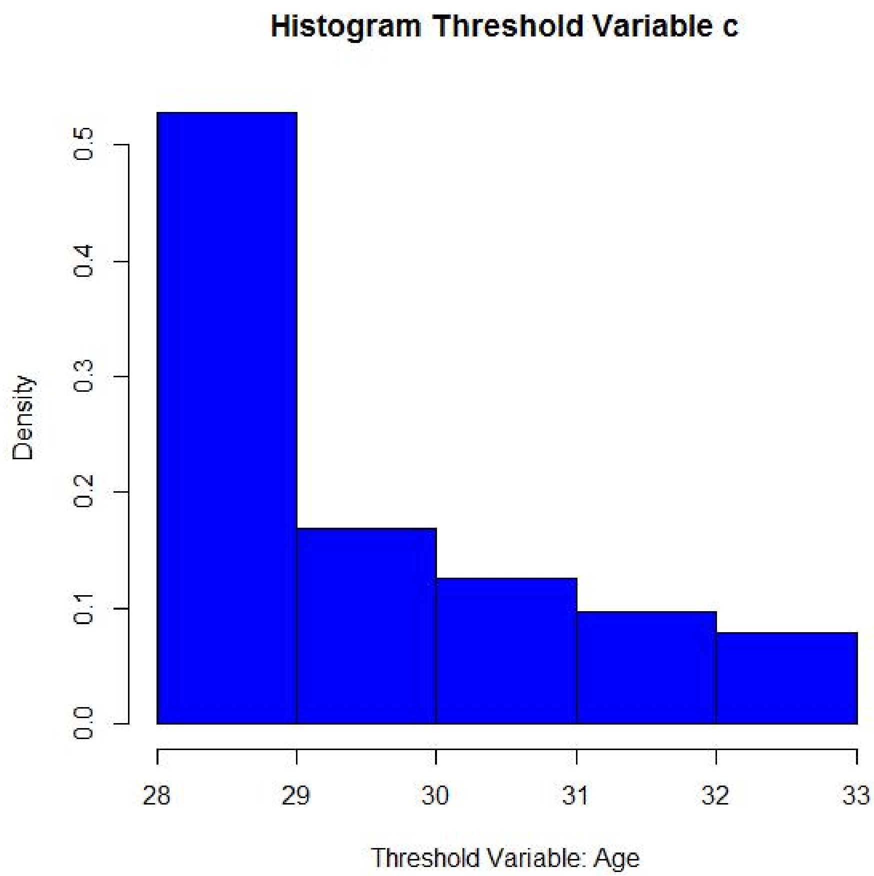

Figure 2.8: Histogram for the threshold variable age.

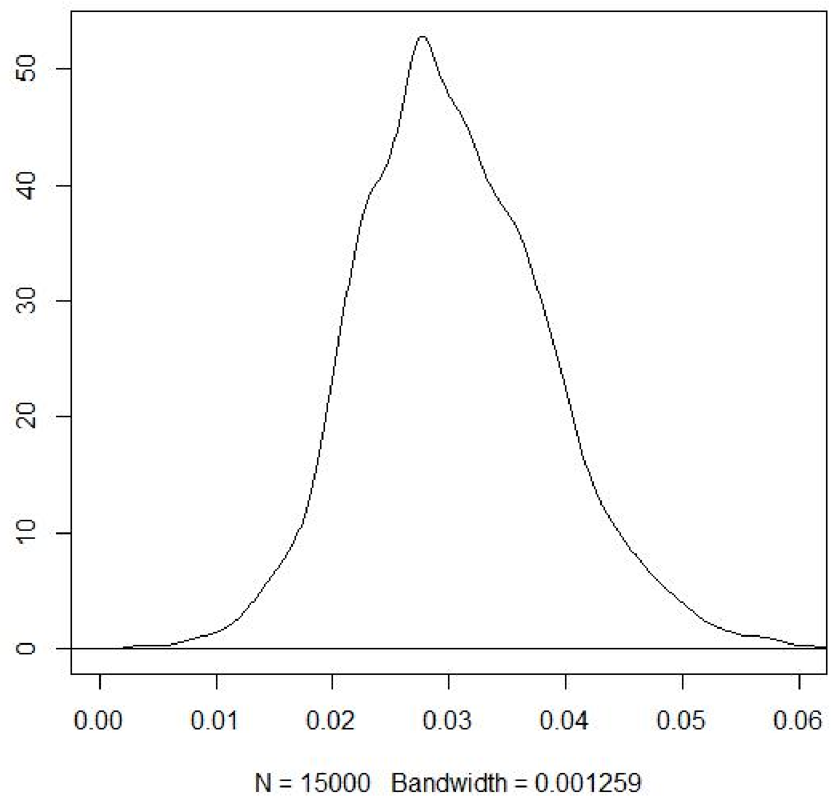

Figure 2.9: Posterior density estimation for γ.

this case, has a posterior mean of 0.03 and a standard deviation of 0.008 (see Table 2.5). An inspection of Figure 2.9 shows that there is a 50% posterior probability that γ is between 0.025 and 0.036. The parameters γ and c complete the description of the weight function. This will matter for the analysis of the persistence parameters $\rho's$ and the variance of the innovations.

It is quite interesting to notice what happens with ρ_1 and ρ_2. The posterior means of ρ_1 and ρ_2 are 0.84 and 0.83, with a standard deviation of 0.047 and 0.02

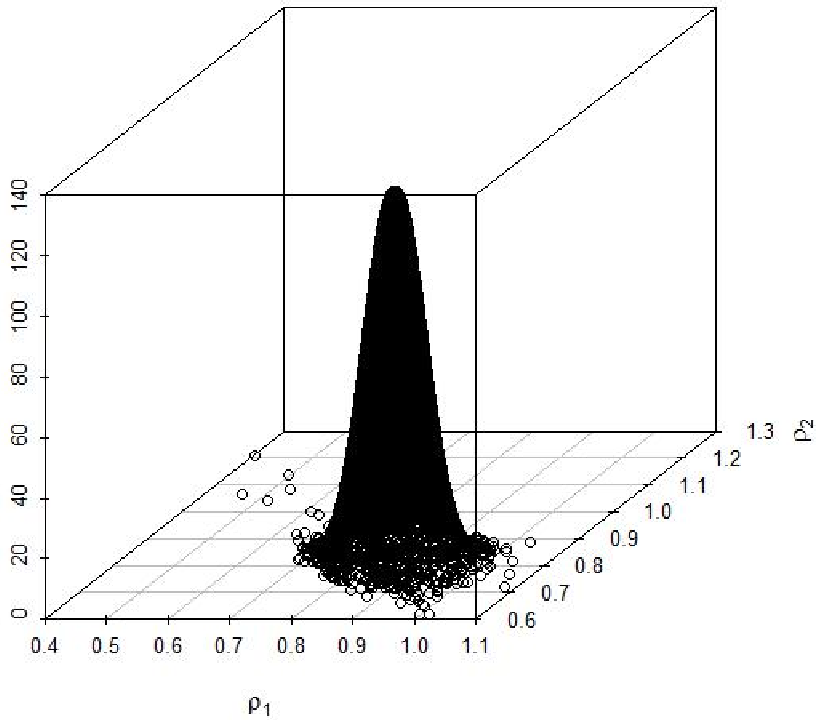

Figure 2.10: Joint posterior distribution of ρ_1 and ρ_2.

respectively and a correlation coefficient of -0.28 (see Table 2.3). Their joint posterior distribution in Figure 2.10 has an important mass of probability in the region $[0.75, 0.95] \times [0.80, 0.85]$.

It seems most likely that ρ_1 should be higher than ρ_2 and is true that there is a positive probability for the persistence parameter ρ_1 to be lower than the persistence ρ_2. However, the actual ρ in the life cycle is a convex combination of ρ_1 and ρ_2. This weighted ρ does not show considerable variation in the life cycle, having a minimum

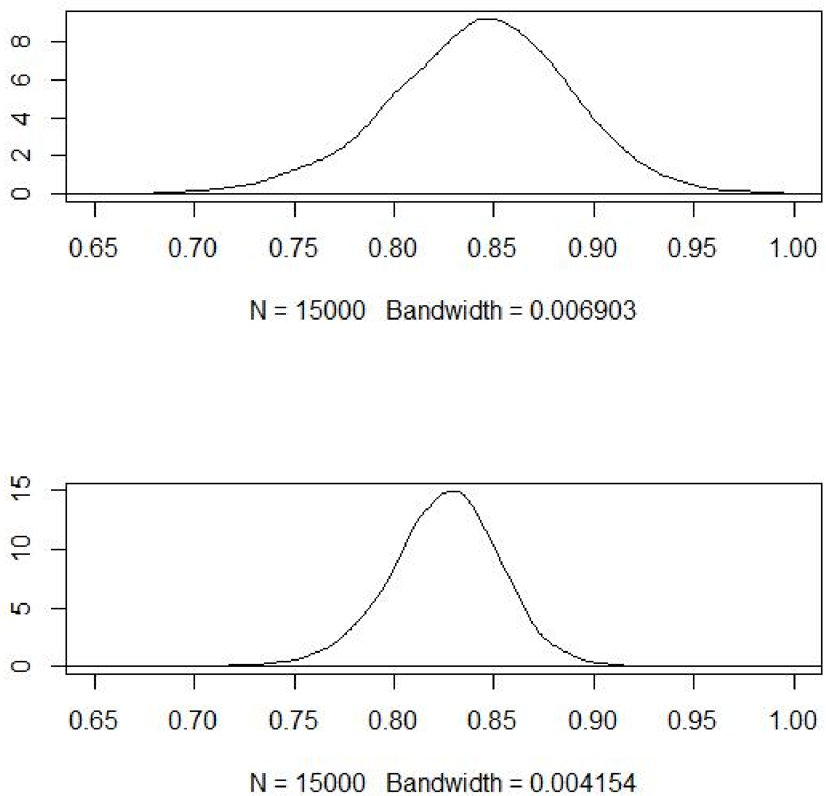

Figure 2.11: Posterior density estimation for ρ_1 (above) and ρ_2 (below).

value of 0.829 and maximum value of 0.835.

This actual ρ is approximately equal to 0.83 and implies that a shock is halved in 3.7 periods, significantly less than the 17 periods needed in the conventional RIP model. The marginal densities for ρ are in Figure 2.11. The wide posterior dispersion of ρ_1 suggests that there might be forces in action that a threshold like age is unable to detect.

The effect of the nonlinearities on the persistence parameters translates into

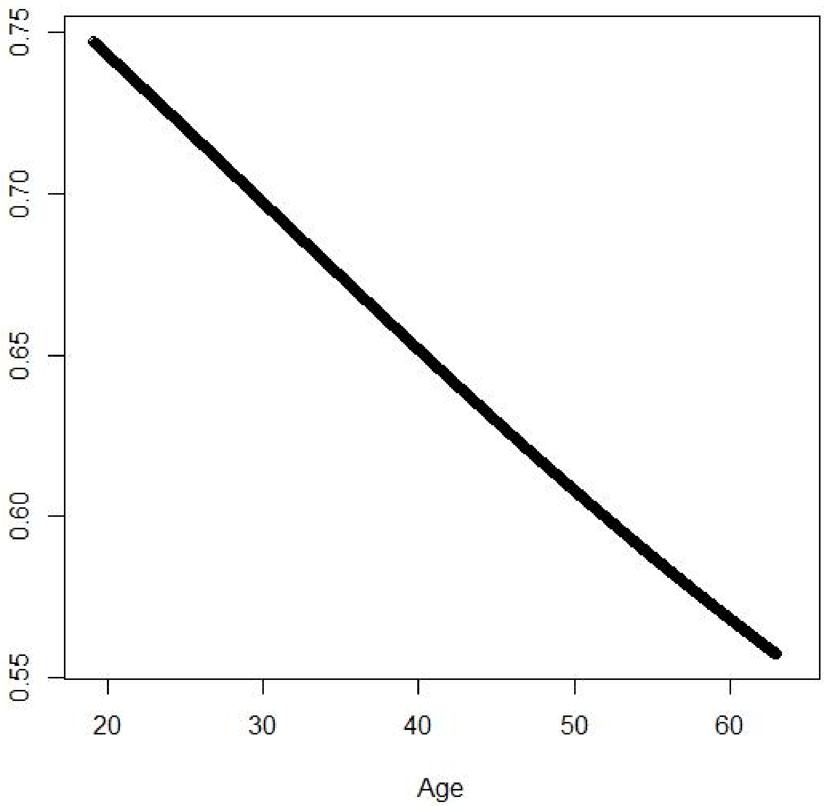

Figure 2.12: Posterior mean of $k_t^i(\gamma, c, \phi)$.

a reduction of the weighted ρ's value and seems not to affect the difference between ρ_1 and ρ_2. The situation is different for the variance of the innovations.

There are two parameters that show the effect of these nonlinearities on the variance of the idiosyncratic shocks. These parameters are ϕ and $k_t^i(\gamma, c, \phi)$, where ϕ is the ratio of the variance of the second regime with respect to the first regime, and $k_t^i(\gamma, c, \phi)$ is the component of the variance that captures the nonlinearity effects.

Figure 2.12 shows the behavior of $k_t^i(\gamma, c, \phi)$ for the average ϕ, c and γ, and it

Table 2.6: Variance components (2 regimes).

Parameters	**Mean**	**25%**	**50%**	**75%**
λ^{-1}	0.0042	0.004	0.0042	0.0044
κ_i^{-1}	0.477	0.357	0.408	0.546
g	0.005	0.0045	0.0049	0.0054
ϕ	0.397	0.37	0.398	0.426

is clear how the variance of the innovations is reduced as agents get older. Table 2.6 and Figure 2.13 indicate that there is a 50% posterior probability that ϕ is between 0.37 and 0.42, and the average of the ratio of these two variances is almost 40%. The stationary variance of the regime for the young workers is 2.5 times the stationary variance of regime for the old workers.[15] While the actual variance in the life cycle of a worker of 64 years is 74% of the corresponding variance for a 19-year-old worker.

The individual-specific component of the variance κ_i^{-1} shows a similar behavior as in the previous subsection. It has a posterior mean of 0.48 and 75% of the individuals in the sample have values of κ_i^{-1} below 0.55. However, there is nonnegligible proportion of individuals with values higher than 1 (see Figure 2.14). The posterior distribution of the rest of the components of the variance, λ and g, can be seen in Figures in 2.15 and 2.16.

The initial variance in the first-age period in this model is 0.30, which means that a one-standard-deviation shock represents a 49% increase in labor earnings with respect to the mean.

[15]The stationary variance of an AR(1) process with correlation coefficient ρ and variance for the error term σ^2 is equal to $\frac{\sigma^2}{1-\rho^2}$.

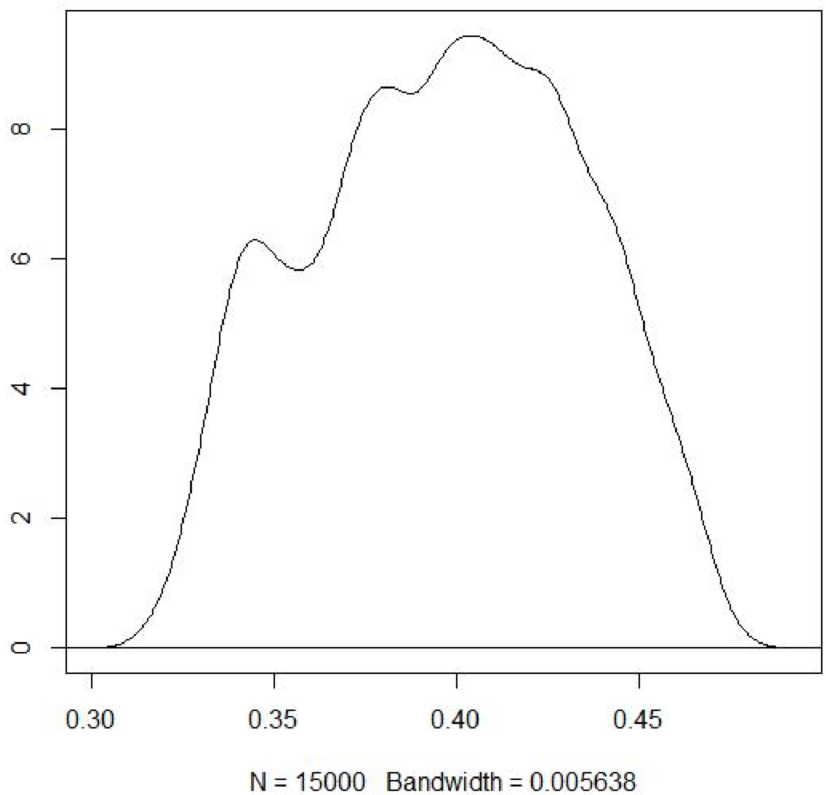

Figure 2.13: Posterior density estimation for ϕ.

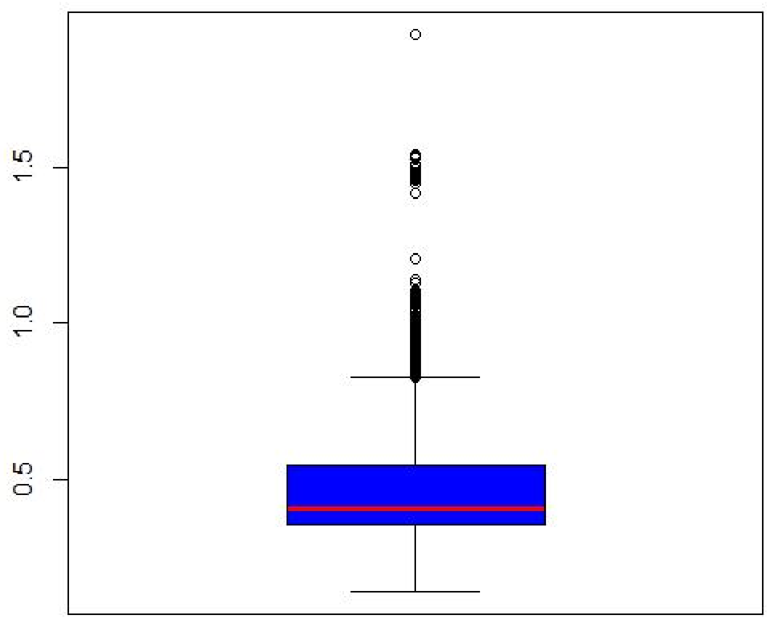

Figure 2.14: Boxplot for the posterior distribution of κ_i^{-1} in a LSTAR(1).

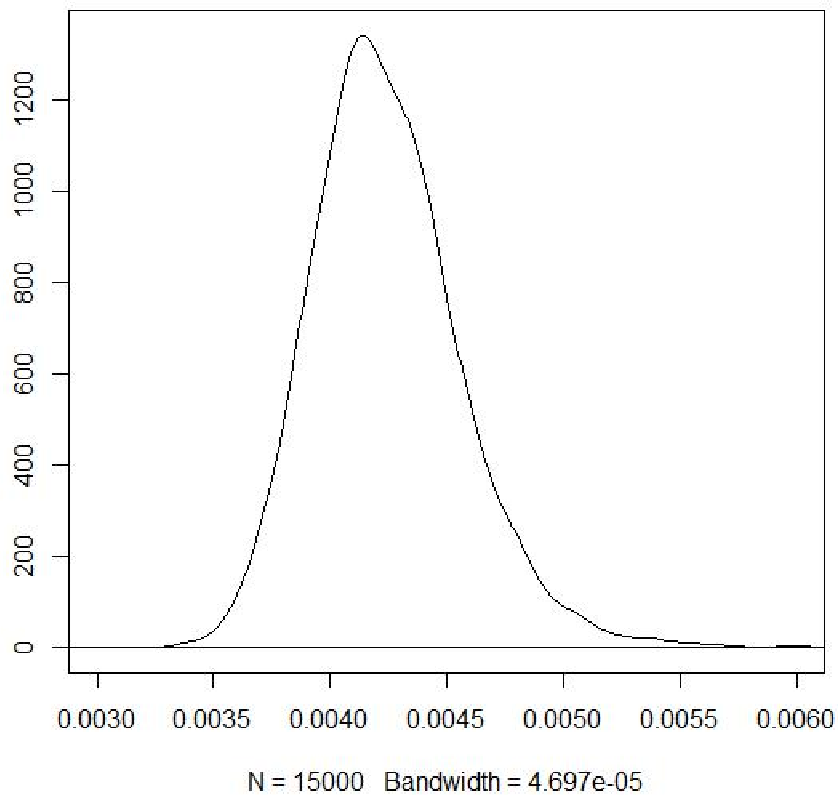

Figure 2.15: Posterior distribution of λ^{-1} in a LSTAR(1).

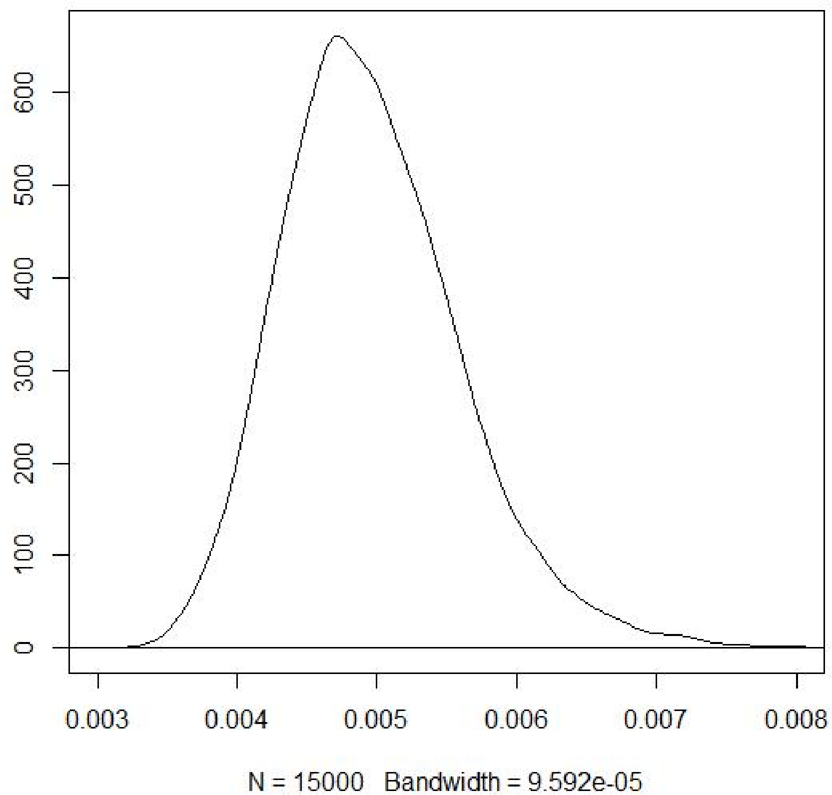

Figure 2.16: Posterior distribution of the precision g in a LSTAR(1).

2.6 Conclusion

In this chapter, I have shown an age profile in the idiosyncratic earnings shocks and introduced for the first time a Smoothed Transition Autoregressive model for the persistent component of the residual earnings. This nonlinear model has the main advantage of disentangling the economic forces acting in the life cycle and a simple structure capable of capturing the intrinsic complexity of an income process.

I have departed from the standard GMM framework used in the literature by following a Bayesian approach. From a methodological point of view, Bayesian methods are capable of dealing with nonlinearities and levels of heterogeneity in the income process which are not easy to estimate with GMM techniques. They are better suited to deal with a data set like the PSID, which has top-coded observations and a biannual periodicity since 1997, where missing values need to be imputed. Moreover, they simplify the estimation of income processes where the persistent component is not directly observable and filtering techniques need to be applied.

My results show that workers younger than 29 experience income shocks with higher variance and a positive probability of having a lower persistence than older workers. These variances have significant differences in magnitude. On the one hand, shocks seem to be riskier; however, they last less time when the worker is younger. The opposite is true when the worker is older. The high dispersion in the posterior distribution of the persistence when a worker is young suggests that using age as a threshold variable might not give a complete picture of what is seen in the data.

A comparison of a RIP model with constant persistent and heterogeneity with respect to a standard RIP shows the unexpected result that the mean of ρ is lower in the first model. Even though the purpose of this chapter is not to explain this difference, one possible explanation could be the unit root problem pointed out by Sims and Uhlig (1991). If that happens to be the case, Bayesian methods should be the standard in the literature.

Possible venues for further research are the extension of this model to a HIP setting and the introduction of new threshold variables, such as level of income, whether the economy is in recession or not, occupational mobility, and the sectors where the head of household is employed. After these extensions are studied, one can apply a model selection criterion and with the best model simulate earning paths in the life cycle. These simulations should be contrasted with the ones obtained with a standard model at different age bins and compared with the actual earning paths seen in the data. This comparison will shed light on the goodness of fit of different models and their economic implications.

Perhaps because of the difficulties of modeling income processes with nonlinearities, few papers consider an age profile in the earnings shocks. This chapter fills this gap and shows that these difficulties can be easily overcome and can substantially improve our understanding of different economic phenomena.

APPENDIX A
HYPERPARAMETERS

The values of the hyperparameters are set in such way that the prior distributions are flexible enough to cover the range of all possible and reasonable values of the parameters under investigation. My choices are in line with previous estimations made in the literature. A good reference in this matter is Heathcote, Storesletten and Violante (2010).

The nuisance term $\omega^i_{h(t),t}$ includes the transitory component in the agent's productivity together with the measurement error present in the survey. As was described in the main body of the paper, the variance $\omega^i_{h(t),t}$ is distributed as a chi-squared. My choice of hyperparameters is $\nu_\omega = 200$ and $\underline{s}^2_\omega = 15$, implying that $P\left(0.064 \leq \sigma^2_t \leq 0.089\right) = 0.9$. Moreover, there is a 99% probability that σ^2_t is between 0.02 and 0.11.

For g_j^{-1}, I choose $\underline{s}^2 = 0.2$ and $\underline{\nu}_g = 2$ which means that the variance of η^i_t in the autoregressive component of the income process is between 0.033 and 1.95 with a 90% probability. Formally, $P\left(0.033 \leq g_j^{-1} \leq 1.95\right) = 0.9 \; \forall j$.

The vector $\underline{\mathbf{p}}$ is distributed as a Dirichlet distribution with $\alpha = 10 \times \mathbf{1}_M$, making it uniform over the $M - 1$ simplex, with $p_j' s$ similar to each other and away from zero.

With respect to the persistence parameters, the mean of the random vector $\widetilde{\rho}$ is $\underline{\mu}_{\widetilde{\rho}} = (0.5, 0)'$ and the precision is $h_\rho I = 0.5I$. The parameter λ that multiplies the variance of the first age-period shocks is distributed as $s^2_\lambda \lambda \sim \chi^2\left(\nu_\lambda\right)$ with $s^2_\lambda = 2$

and $\nu_\lambda = 2$, which gives a $P\left(0.33 \leq \lambda^{-1} \leq 19.5\right) = 0.9$. As can be seen, this is not an informative prior. However, it seems reasonable to allow for a λ that could reduce the variance in the first age-period by a third or increase it by a factor of twenty.

I choose μ to have mean $\mathbf{0}$ and considerable dispersion, making the prior uninformative. The values chosen are $\underline{\mu}_\mu = \mathbf{0}$ and $\underline{H}_\mu = 0.01\mathbf{I}$. In the same line, $H \sim Wishart\left(\underline{\nu}_H, \underline{S}_H\right)$ with $\underline{\nu}_H = 10$ and $\underline{S}_H = 1,000\mathbf{I}$.

Finally, the agent-specific component κ_i of the variance of η_t^i is distributed as a chi-squared random variable, where the degrees of freedom $\underline{\nu}_\kappa$ and the multiplying constant s_κ^2 are chosen in such way that the individual variance can be multiplied by values that could potentially augment it by a factor of 10 or diminish it by 10% with a high probability, i.e., $s_\kappa^2 \kappa_i \sim \chi^2\left(\underline{\nu}_\kappa\right)$ with $s_\kappa^2 = 1$ and $\underline{\nu}_\kappa = 2$, which implies $P\left(0.17 \leq \kappa_i^{-1} \leq 9.75\right) = 0.9$.

APPENDIX B
POSTERIOR DISTRIBUTIONS

In this appendix, I derive the posterior distributions of the parameters of interest in each of the Gibbs sampler steps:

- $\underline{\left(y^{*}/\left\{x^{i}_{h(t),t}\right\}_{i,t},\beta,\{\sigma^2_t\}^{T}_{t=1},\{s^i_t\}^{I,T_i}_{i=1,t=t_i},\{g_j\}^{m}_{j=1},\rho_1,\rho_2,\gamma,c,\phi,y\right)'}$

  Our starting point is:

$$\begin{cases} y^i_{h(t),t} = x^{i\prime}_{h(t),t}\beta + \varepsilon^i_{h(t),t} + \omega^i_{h(t),t} \\ \varepsilon^i_{h(t),t} = \underbrace{\left[\left(1 - G\left(\gamma, c, \tau^i_t\right)\right)\rho_1 + G\left(\gamma, c, \tau^i_t\right)\rho_2\right]}_{m^i_t(\gamma,c,\rho_1,\rho_2)}\varepsilon^i_{h(t)-1,t-1} + \eta^i_t \end{cases}$$

  Then,

  $\varepsilon^i_{h(t),t} = y^i_{h(t),t} - x^{i\prime}_{h(t),t}\beta - \omega^i_{h(t),t}$ and $y^i_{h(t),t} = x^{i\prime}_{h(t),t}\beta + m^i_t\left(\gamma, c, \rho_1, \rho_2\right)\varepsilon^i_{h(t)-1,t-1} + \eta^i_t + \omega^i_{h(t),t}$

  Combining both equations, we have:

$$y^i_{h(t),t} = x^{i\prime}_{h(t),t}\beta + m^i_t\left(\gamma, c, \rho_1, \rho_2\right)\left(y^i_{h(t)-1,t-1} - x^{i\prime}_{h(t)-1,t-1}\beta - \omega^i_{h-1,t-1}\right) + \eta^i_t + +\omega^i_{h(t),t}$$

$$y^i_{h(t),t} = \left(x^i_{h(t),t} - m^i_t\left(\gamma, c, \rho_1, \rho_2\right)x^i_{h(t)-1,t-1}\right)'\beta + m^i_t\left(\gamma, c, \rho_1, \rho_2\right)y^i_{h(t)-1,t-1} + \omega^{i\prime}_{h(t),t} - -m^i_t\left(\gamma, c, \rho_1, \rho_2\right)\omega^i_{h-1,t-1} + \eta^i_t$$

  Using the above, we can find the posterior distribution:

$$p\left(y^{*,i}_{h(t),t}/\left\{x^i_{h(t),t}\right\}_{i,t},\beta,\{\sigma^2_t\}^{T}_{t=1},\{s^i_t\}^{I,T_i}_{i=1,t=t_i},\{g_j\}^{m}_{j=1},\rho_1,\rho_2,\gamma,c,\phi,y\right) \propto$$
$$\propto p\left(y^i_{h(t)+1,t+1}/y^{*,i}_{h(t),t},\cdots\right)p\left(y^{*,i}_{h(t),t}/y^i_{h(t)-1,t-1},\cdots\right)$$

$$\propto \exp\left\{ -\frac{1}{2} \left\{ \begin{array}{c} \left(\underbrace{\sigma_{t+1}^2 + m_{t+1}^i(\gamma, c, \rho_1, \rho_2)^2 \sigma_t^2 + Var\left(\eta_{t+1}^i\right)}_{R_{t+1}} \right)^{-1} \times \\ \times \left[\begin{array}{c} y_{h(t)+1,t+1}^i - \left(x_{h+1,t+1}^i - m_{t+1}^i(\gamma, c, \rho_1, \rho_2) x_{h(t),t}^i \right)' \beta - \\ -m_{t+1}^i(\gamma, c, \rho_1, \rho_2) y_{h(t),t}^{*,i} \end{array} \right]^2 + \\ +R_t^{-1} \left[\begin{array}{c} y_{h(t),t}^{*,i} - \left(x_{h(t),t}^i - m_t^i(\gamma, c, \rho_1, \rho_2) x_{h(t)-1,t-1}^i \right)' \beta - \\ -m_t^i(\gamma, c, \rho_1, \rho_2) y_{h(t)-1,t-1}^i \end{array} \right]^2 \end{array} \right\} \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left\{ \begin{array}{c} \left(y_{h(t),t}^{*,i}\right)^2 \left[\left(m_{t+1}^i(\gamma, c, \rho_1, \rho_2)\right)^2 R_{t+1}^{-1} + R_t^{-1} \right] - \\ -2y_{h(t),t}^{*,i} \left[\begin{array}{c} R_{t+1}^{-1} m_{t+1}^i(\gamma, c, \rho_1, \rho_2) \times \\ \times \left(\begin{array}{c} y_{h(t)+1,t+1}^i - \\ \left(x_{h+1,t+1}^i - m_{t+1}^i(\gamma, c, \rho_1, \rho_2) x_{h(t),t}^i \right)' \beta \end{array} \right) \\ + R_t^{-1} \left(\left(\begin{array}{c} x_{h(t),t}^i - \\ m_t^i(\gamma, c, \rho_1, \rho_2) x_{h(t)-1,t-1}^i \end{array} \right)' \beta + \\ +m_t^i(\gamma, c, \rho_1, \rho_2) y_{h(t)-1,t-1}^i \right) \end{array} \right] \end{array} \right\} \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \overline{h}_y \left(y_{h(t),t}^{*,i} - \overline{\mu}_y \right)^2 \right\}$$

where

$$\overline{h}_y = \left(m_{t+1}^i(\gamma, c, \rho_1, \rho_2)\right)^2 R_{t+1}^{-1} + R_t^{-1}$$

and

$$\overline{\mu}_y = \overline{h}_y^{-1} \left\{ \begin{array}{c} R_{t+1}^{-1} m_{t+1}^i(\gamma, c, \rho_1, \rho_2) \times \\ \times \left(y_{h(t)+1,t+1}^i - \left(x_{h+1,t+1}^i - m_{t+1}^i(\gamma, c, \rho_1, \rho_2) x_{h(t),t}^i \right)' \beta \right) + \\ +R_t^{-1} \left[\begin{array}{c} \left(x_{h(t),t}^i - m_t^i(\gamma, c, \rho_1, \rho_2) x_{h(t)-1,t-1}^i \right)' \beta - \\ m_t^i(\gamma, c, \rho_1, \rho_2) y_{h(t)-1,t-1}^i \end{array} \right] \end{array} \right\}$$

Therefore,

$$y_{h(t),t}^{*,i} / \cdots \sim N\left(\overline{\mu}_y, \overline{h}_y^{-1}\right).$$

In the case that $y_{h(t),t}^{*,i}$ is top-coded, I impute a value from the distribution

$y_{h(t),t}^{*,i} / \cdots \sim N\left(\overline{\mu}_y, \overline{h}_y^{-1}\right) I\left(y_{h(t),t}^{*,i}\right)_{[\overline{y},+\infty)}$, where $\overline{y}$ is the top-coded value given in the sample.

- $\underline{\left(\left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i} / \left\{x^i_{h(t),t}\right\}, \beta, \{\sigma^2_t\}^T_{t=1}, \{\rho_i\}^2_{i=1}, \gamma, c, \lambda, \phi, \{\kappa_i\}^I_{i=1}, \{s^i_t\}^{I,T_i}_{i=1,t=t_i},\right.}$ $\underline{\left.\{g_j\}_j, y, y^*\right)'}$

Following Durbin and Koopman (2002), we have for each individual i, conditional on $\{s^i_t\}^{T_i}_{t_i}$, a conditionally linear Gaussian state-space model given by:

$$\begin{cases} \widetilde{y}^i_{h(t),t} = y^i_{h(t),t} - x^{i\prime}_{h(t),t}\beta = \varepsilon^i_{h(t),t} + \omega^i_{h(t),t} \\ \varepsilon^i_{h(t),t} = m^i_t(\gamma, c, \rho_1, \rho_2)\,\varepsilon^i_{h(t)-1,t-1} + \eta^i_t \end{cases} \quad (A.1)$$

where $\omega^i_{h(t),t} \sim N(0, \sigma^2_t)$ and $\eta^i_t / s^i_t \sim N\left(0, g^{-1}_{s^i_t}\right)$.

Let $w^i = \left(\omega^i_{h(t_i),t_i}, \eta^i_{t_i}, \ldots, \omega^i_{h(T_i),T_i}, \eta^i_{T_i}\right)'$, $\widetilde{y}^i = \left(\widetilde{y}^i_{h(t_i),t_i}, \ldots, \widetilde{y}^i_{h(T_i),T_i}\right)'$ and $\varepsilon^i = \left(\varepsilon^i_{h(t_i),t_i}, \ldots, \varepsilon^i_{h(T_i),T_i}\right)'$,

where $w^i \sim N(0, \Omega)$, with $\Omega = diag\left(\sigma^2_{t_i}, g^{-1}_{s^i_{t_i}}, \ldots, \sigma^2_{T_i}, g^{-1}_{s^i_{T_i}}\right)$.

Step 1: Take a random draw from $w^+ \sim N(0, \Omega)$ and use it to generate $\widetilde{y}^{i+}$ and ε^{i+} from the recursion $(A.1)$.

Step 2: Apply the Kalman filter (Kalman, 1960) and the disturbance smoother to $\widetilde{y}^{i+}$ and $\widetilde{y}^i$:

Let

$$\begin{aligned}
\widehat{\varepsilon}^i_{h(t),t/t-1} &= E\left(\varepsilon^i_{h(t),t} / \widetilde{y}^i_{h(t)-1,t-1}, s^i_t\right) \\
e^i_t &= \widetilde{y}^i_{h(t),t} - \widehat{\varepsilon}^i_{h(t),t/t} \\
R_{t/t-1} &= Var\left(\varepsilon^i_{h(t),t} / \widetilde{y}^i_{h(t)-1,t-1}, s^i_t\right) \\
&= \left(m^i_t(\gamma, c, \rho_1, \rho_2)\right)^2 R_{t-1/t-1} + g^{-1}_{s^i_t} \\
P_{t/t-1} &= Var\left(\widetilde{y}^i_{h(t),t} / \widetilde{y}^i_{h(t)-1,t-1}, s^i_t\right) \\
&= R_{t/t-1} + \sigma^2_t
\end{aligned}$$

As,

$$\left[\begin{pmatrix} \widetilde{y}^i_{h(t),t} \\ \varepsilon^i_{h(t),t} \end{pmatrix} / \widetilde{y}^i_{h(t)-1,t-1}, s^i_t\right] \sim N\left(\begin{pmatrix} \widehat{\varepsilon}^i_{h(t),t/t-1} \\ \widehat{\varepsilon}^i_{h(t),t/t-1} \end{pmatrix}, \begin{pmatrix} P_{t/t-1} & R_{t/t-1} \\ R'_{t/t-1} & R_{t/t-1} \end{pmatrix}\right)$$

Then,

$$\begin{aligned} \widehat{\varepsilon}^i_{h(t),t/t} &= \widehat{\varepsilon}^i_{h(t),t/t-1} + \underbrace{R_{t/t-1}P^{-1}_{t/t-1}}_{K_t} e^i_t \\ R_{t/t} &= R_{t/t-1} - R_{t/t-1}P^{-1}_{t/t-1}R_{t/t-1} \\ &= (I - K_t)\, R_{t/t-1} \end{aligned}$$

The disturbance smoother is given by:

$$\widehat{w}_t = E\left(w_t/\widetilde{y}^i\right) = \begin{pmatrix} \sigma^2_t P^{-1}_{t/t-1} & -\sigma^2_t K'_t \\ 0 & g^{-1}_{s^i_t} \end{pmatrix} \begin{pmatrix} e^i_t \\ r_t \end{pmatrix}$$

where

$$r_{t-1} = P^{-1}_{t/t-1} e^i_t + \left(m^i_t\left(\gamma, c, \rho_1, \rho_2\right) - K_t\right) r_t$$

with $r_{T_i} \equiv 0$.

Step 3: Compute $\widehat{\varepsilon}^{i+} = E\left(\varepsilon^{i+}/\widetilde{y}^{i+}, \{s^i_t\}^{T_i}_{t=t_i}\right)$ and $\widehat{\varepsilon}^i = E\left(\varepsilon^i/\widetilde{y}^i, \{s^i_t\}^{T_i}_{t=t_i}\right)$ with the forwards recursion:

$$\widehat{\varepsilon}^i_{h(t)+1,t+1} = m^i_t\left(\gamma, c, \rho_1, \rho_2\right) \widehat{\varepsilon}^i_{h(t),t} + g^{-1}_{s^i_t} r_t$$

Step 4: Keep $\widetilde{\varepsilon}^i = \varepsilon^{i+} - \widehat{\varepsilon}^{i+} + \widehat{\varepsilon}^i$.

Finally,

$$\left\{\widetilde{\varepsilon}^i\right\}^I_{i=1} \sim p\left(\left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i} / \ldots\right)$$

- $$\frac{\left(\beta/\left\{x^i_{h(t),t}\right\}_{i,t},\{\sigma^2_t\}^T_{t=1},\left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i},\mu,H,y,y^*\right)'}{p\left(\beta/\left\{x^i_{h(t),t}\right\}_{i,t},\{\sigma^2_t\}^T_{t=1},\left\{\varepsilon^i_{h(t),t}\right\}^{I,\,T_i}_{i=1,t=t_i},\mu,H,y,y^*\right)\propto}$$
$$\propto\left[\prod_{i=1}^{I}\prod_{t=t_i}^{T_i}p\left(y^i_{h(t),t}/x^i_{h(t),t},\beta,\sigma^2_t,\varepsilon^i_{h(t),t}\right)\right]p\left(\beta/\mu,H\right)$$
$$\propto\left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i}\exp\left[-\tfrac{1}{2\sigma^2_t}\left(y^i_{h(t),t}-x^i_{h(t),t}\beta-\varepsilon^i_{h(t),t}\right)^2\right]\right\}\exp\left\{-\tfrac12\left(\beta-\mu\right)'H\left(\beta-\mu\right)\right\}$$
$$\propto\exp\left[-\tfrac12\sum_{i=1}^{I}\sum_{t=t_i}^{T_i}\left(\frac{y^i_{h(t),t}-x^{i\prime}_{h(t),t}\beta-\varepsilon^i_{h(t),t}}{\sigma_t}\right)^2\right]\exp\left\{-\tfrac12\left(\beta-\mu\right)'H\left(\beta-\mu\right)\right\}$$
$$\propto\exp\left\{-\tfrac12\left[\beta'\sum_{i=1}^{I}\sum_{t=t_i}^{T_i}\frac{x^i_{h(t),t}x^{i\prime}_{h,t}}{\sigma^2_t}\beta-2\beta'\sum_{i=1}^{I}\sum_{t=t_i}^{T_i}\frac{x^i_{h(t),t}\left(y^i_{h(t),t}-\varepsilon^i_{h(t),t}\right)}{\sigma^2_t}+\beta'H\beta-2\beta'H\mu\right]\right\}$$
$$\propto\exp\left\{-\tfrac12\left[\beta'\left(\sum_{i=1}^{I}\sum_{t=t_i}^{T_i}\frac{x^i_{h(t),t}x^{i\prime}_{h(t),t}}{\sigma^2_t}+H\right)\beta-2\beta'\left(\sum_{i=1}^{I}\sum_{t=t_i}^{T_i}\frac{x^i_{h(t),t}\left(y^i_{h(t),t}-\varepsilon^i_{h(t),t}\right)}{\sigma^2_t}+H\mu\right)\right]\right\}$$
$$\propto\exp\left\{-\tfrac12\left[\left(\beta-\overline{\mu}_\beta\right)'\overline{H}_\beta\left(\beta-\overline{\mu}_\beta\right)\right]\right\}$$
where $\overline{H}_\beta=\sum_{i=1}^{I}\sum_{t=t_i}^{T_i}\frac{x^i_{h(t),t}x^{i\prime}_{h(t),t}}{\sigma^2_t}+H$ and $\overline{\mu}_\beta=\overline{H}_\beta^{-1}\left(\sum_{i=1}^{I}\sum_{t=t_i}^{T_i}\frac{x^i_{h(t),t}\left(y^i_{h(t),t}-\varepsilon^i_{h(t),t}\right)}{\sigma^2_t}+H\mu\right)$.

Thus,

$$\beta/\left(\left\{x^i_{h(t),t}\right\}_{i,t},\{\sigma^2_t\}^T_{t=1},\left\{\varepsilon^i_{h(t),t}\right\}^{I,\,T_i}_{i=1,\,t=t_i},\mu,H,y,y^*\right)\sim N\left(\overline{\mu}_\beta,\overline{H}_\beta^{-1}\right).$$

- $$\frac{\left(\{\sigma^2_t\}^T_{t=1}/\left\{x^i_{h(t),t}\right\}_{i,t},\beta,\left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i},y,y^*\right)'}{p\left(\sigma^2_t/\left\{x^i_{h(t),t}\right\}_i,\beta,\left\{\varepsilon^i_{h(t),t}\right\}_i,y,y^*\right)\propto}$$
$$\propto\left[\prod_{i\in B_t}p\left(y^i_{h(t),t}/x^i_{h(t),t},\beta,\sigma^2_t,\kappa_i,\varepsilon^i_{h(t),t}\right)\right]p\left(\sigma^2_t\right)$$
where $B_t=\left\{i\in I\ s.t.\ y^i_{h(t),t}\ is\text{ in the sample}\right\}$, and $n^t_I=\#B_t$
$$\propto\left(\sigma^2_t\right)^{-\frac{n^t_I}{2}}\exp\left\{-\tfrac{1}{2\sigma^2_t}\sum_{i\in B_t}\left(y^i_{h(t),t}-x^{i\prime}_{h(t),t}\beta-\varepsilon^i_{h(t),t}\right)^2\right\}\left(\sigma^2_t\right)^{-\left(\frac{\nu_\omega}{2}-1\right)}\exp\left\{-\tfrac{1}{2\sigma^2_t}\underline{s}^2_\omega\right\}$$
$$\propto\left(\sigma^2_t\right)^{-\left(\frac{n^t_I}{2}+\frac{\nu_\omega}{2}-1\right)}\exp\left\{-\frac{1}{\sigma^2_t}\left[\frac{\sum_{i\in B_t}\left(y^i_{h(t),t}-x^{i\prime}_{h(t),t}\beta-\varepsilon^i_{h(t),t}\right)^2+\underline{s}^2_\omega}{2}\right]\right\}$$
$$\propto\left(\sigma^2_t\right)^{-\left(\frac{n^t_I}{2}+\frac{\nu_\omega}{2}-1\right)}\exp\left\{-\frac{1}{\sigma^2_t}\left[\frac{n^t_I\overline{s}^2_{\omega,t}+\underline{s}^2_\omega}{2}\right]\right\}$$

$$\therefore \left(\sigma_t^2\right)^{-1}/\cdots \sim \Gamma\left(\frac{n_I^t}{2}+\frac{\nu_\omega}{2}, \frac{\sum\limits_{i\in B_t}\left(y_{h(t),t}^i - x_{h(t),t}^{i\prime}\beta - \varepsilon_{h(t),t}^i\right)^2 + \underline{s}_\omega^2}{2}\right)$$

- $\underline{\left(\widetilde{\rho}/\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}, \gamma, c, \lambda, \phi, \{\kappa_i\}_{i=1}^I, \{s_t^i\}_{i=1,t=t_i}^{I,T_i}, \{g_j\}_{j=1}^m\right)'}$

$$p\left(\widetilde{\rho}/\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}, \gamma, c, \lambda, \phi, \{\kappa_i\}_{i=1}^I, \{s_t^i\}_{i=1,t=t_i}^{I,T_i}, \{g_j\}_{j=1}^m\right) \propto$$

$$\propto \left\{\prod_{i=1}^I \prod_{t=t_i+1}^{T_i} p\left(\varepsilon_{h(t),t}^i/\left\{\varepsilon_{h(t),t}^i\right\}_{t_i}^{t-1}, \widetilde{\rho}, \gamma, c, \lambda, \phi, \kappa_i, \{s_t^i\}_{t=t_i}^t, \{g_j\}_{j=1}^m\right)\right\} \times p\left(\widetilde{\rho}\right)$$

$$\propto \left\{\prod_{i=1}^I \prod_{t=t_i+1}^{T_i} p\left(\varepsilon_{h(t),t}^i/\varepsilon_{h(t)-1,t-1}^i, \cdots\right)\right\} \times p\left(\widetilde{\rho}\right)$$

$$\propto \left\{\prod_{i=1}^I \prod_{t=t_i+1}^{T_i} \exp\left\{-\frac{1}{2}\left[\frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\left(\varepsilon_{h(t),t}^i - \varepsilon^i(\gamma,c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2\right]\right\}\right\} \times$$

$$\times \exp\left\{-\frac{1}{2}h_\rho\left(\widetilde{\rho}-\mu_{\widetilde{\rho}}\right)'\left(\widetilde{\rho}-\mu_{\widetilde{\rho}}\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left\{\begin{array}{c}\sum\limits_{i=1}^I \sum\limits_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\left(\varepsilon_{h(t),t}^i - \varepsilon^i(\gamma,c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2 + \\ + h_\rho\left(\widetilde{\rho}-\mu_{\widetilde{\rho}}\right)'\left(\widetilde{\rho}-\mu_{\widetilde{\rho}}\right)\end{array}\right\}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left\{\begin{array}{c}\sum\limits_{i=1}^I \sum\limits_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\left(\begin{array}{c}\widetilde{\rho}'\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon^i(\gamma,c)_{h(t)-1,t-1}\widetilde{\rho}- \\ -2\widetilde{\rho}'\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon_{h(t),t}^i\end{array}\right) \\ + h_\rho\left(\widetilde{\rho}'\widetilde{\rho}-2\widetilde{\rho}'\mu_{\widetilde{\rho}}\right)\end{array}\right\}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left\{\begin{array}{c}\widetilde{\rho}'\left(\sum\limits_{i=1}^I \sum\limits_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon^i(\gamma,c)_{h(t)-1,t-1}\right)\widetilde{\rho}- \\ -2\widetilde{\rho}'\left(\sum\limits_{i=1}^I \sum\limits_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon_{h(t),t}^i\right) + h_\rho\left(\widetilde{\rho}'\widetilde{\rho}-2\widetilde{\rho}'\mu_{\widetilde{\rho}}\right)\end{array}\right\}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left\{\begin{array}{c}\widetilde{\rho}'\left(\sum\limits_{i=1}^I \sum\limits_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon^i(\gamma,c)_{h(t)-1,t-1} + h_\rho I\right)\widetilde{\rho}- \\ -2\widetilde{\rho}'\left(\sum\limits_{i=1}^I \sum\limits_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon_{h(t),t}^i + h_\rho\mu_{\widetilde{\rho}}\right)\end{array}\right\}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\widetilde{\rho}-\overline{\mu}_{\widetilde{\rho}}\right)'\overline{H}_{\widetilde{\rho}}\left(\widetilde{\rho}-\overline{\mu}_{\widetilde{\rho}}\right)\right]\right\}$$

where $\overline{H}_{\widetilde{\rho}} = \sum\limits_{i=1}^I \sum\limits_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon^i(\gamma,c)_{h(t)-1,t-1} + h_\rho I$, and

$$\overline{\mu}_{\widetilde{\rho}} = \overline{H}_{\widetilde{\rho}}^{-1}\left[\sum_{i=1}^I \sum_{t=t_i+1}^{T_i} \frac{\kappa_i g_{s_t^i}}{k_t^i(\gamma,c,\phi)}\varepsilon^i(\gamma,c)_{h(t)-1,t-1}'\varepsilon_{h(t),t}^i + h_\rho\mu_{\widetilde{\rho}}\right]$$

$$\therefore \widetilde{\rho}/\ldots \sim N\left(\overline{\mu}_{\widetilde{\rho}}, \overline{H}_{\widetilde{\rho}}^{-1}\right)$$

- $\underline{\left((\gamma,c)/\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}, (\rho_1,\rho_2), \lambda, \phi, \{\kappa_i\}_{i=1}^I, \{s_t^i\}_{i=1,t=t_i}^{I,T_i}, \{g_j\}_{j=1}^m\right)'}$

Let $A = \{i : \text{individual i has a first-age period observation}\}$, and $q = \#A$.

$$p\left((\gamma, c) / \left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i}, (\rho_1, \rho_2), \lambda, \phi, \{h_i\}^I_{i=1}, \{s^i_t\}^{I,T_i}_{i=1,t=t_i}, \{g_j\}^m_{j=1}\right) \propto$$

$$\propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i} p\left(\varepsilon^i_{h(t),t} / \left\{\varepsilon^i_{h(t),t}\right\}^{t-1}_{t_i}, (\rho_1, \rho_2), \gamma, c, \lambda, \phi, h_i, \mu_{\rho_1}, h_{\rho_1}, \{s^i_t\}^{t_i}_{t=t_i}, \{g_j\}^m_{j=1}\right)\right\}$$

$$\times p(\gamma)\, p(c)$$

$$\propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i} p\left(\varepsilon^i_{h(t),t} / \varepsilon^i_{h(t)-1,t-1}, \cdots\right)\right\} p(\gamma)\, p(c)$$

$$\propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i} \left(\frac{h_i g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\lambda^{I(t=t_i \wedge i\in A)} h_i g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\left(\varepsilon^i_{h(t),t} - \varepsilon^i(\gamma, c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2\right]\right\}\right\}$$

$$\times p(\gamma)\, p(c)$$

$$\propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i} \left(\frac{h_i g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\lambda^{I(t=t_i \wedge i\in A)} h_i g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\left(\varepsilon^i_{h(t),t} - \varepsilon^i(\gamma, c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2\right]\right\}\right\}$$

$$\times \left(1+\gamma^2\right)^{-1} I(\gamma > 0)$$

$$\propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i} \left(k^i_t(\gamma, c, \phi)\right)^{-1}\right\}\left(1+\gamma^2\right)^{-1} I(\gamma > 0)$$

$$\times \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{I}\sum_{t=t_i}^{T_i} \frac{\lambda^{I(t=t_i \wedge i\in A)} h_i g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\left(\varepsilon^i_{h(t),t} - \varepsilon^i(\gamma, c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2\right]\right\}$$

- $\underline{\left(\{\kappa_i\}^I_{i=1} / \left\{\varepsilon^i_{h(t),t}\right\}^{I,T_i}_{i=1,t=t_i}, (\rho_1, \rho_2), (\gamma, c), \lambda, \phi, \{s^i_t\}^{I,T_i}_{i=1,t=t_i}, \{g_j\}^m_{j=1}\right)'}$

$$p\left(\kappa_i / \left\{\varepsilon^i_{h(t),t}\right\}^{T_i}_{t=t_i}, (\rho_1, \rho_2), (\gamma, c), \lambda, \phi, \{s^i_t\}^{T_i}_{t=t_i}, \{g_j\}^m_{j=1}\right) \propto$$

$$\propto \left\{\prod_{t=t_i}^{T_i} p\left(\varepsilon^i_{h(t),t} / \left\{\varepsilon^i_{h(t),t}\right\}^{t-1}_{t_i}, (\rho_1, \rho_2), \gamma, c, \lambda, \phi, \kappa_i, \mu_{\rho_1}, h_{\rho_1}, s^i_t, \{g_j\}^m_{j=1}\right)\right\} p(\kappa_i)$$

$$\propto \left\{\prod_{t=t_i}^{T_i} p\left(\varepsilon^i_{h(t),t} / \varepsilon^i_{h(t)-1,t-1}, \cdots\right)\right\} p(\kappa_i)$$

$$\propto (\kappa_i)^{\frac{T_i}{2}} \exp\left\{-\frac{1}{2}\sum_{t=t_i}^{T_i} \frac{\lambda^{I(t=t_i \wedge i\in A)} \kappa_i g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\left(\varepsilon^i_{h(t),t} - \varepsilon^i(\gamma, c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2\right\}(\kappa_i)^{\frac{\nu_\kappa}{2}-1}$$

$$\times \exp\left\{-\frac{1}{2}s^2_\kappa \kappa_i\right\}$$

$$\propto (\kappa_i)^{\frac{T_i+\nu_\kappa}{2}-1} \exp\left\{-\kappa_i \frac{1}{2}\left[\sum_{t=t_i}^{T_i} \frac{\lambda^{I(t=t_i \wedge i\in A)} g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\left(\varepsilon^i_{h(t),t} - \varepsilon^i(\gamma, c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2 + s^2_\kappa\right]\right\}$$

$$\therefore \kappa_i / \ldots \sim \Gamma\left(\frac{T_i+\nu_\kappa}{2}, \frac{1}{2}\left[\sum_{t=t_i}^{T_i} \frac{\lambda^{I(t=t_i \wedge i\in A)} g_{s^i_t}}{k^i_t(\gamma,c,\phi)}\left(\varepsilon^i_{h(t),t} - \varepsilon^i(\gamma, c)_{h(t)-1,t-1}\widetilde{\rho}\right)^2 + s^2_\kappa\right]\right)$$

- $\underline{\left(\lambda / \left\{\varepsilon^i_{1,t_i}\right\}_{i\in A}, \{\kappa_i\}_{i\in A}, \left\{s^i_{t_i}\right\}_{i\in A}, \{g_j\}^m_{j=1}, \gamma, c, \phi\right)'}$

$$p\left(\lambda/\left\{\varepsilon^i_{1,t_i}\right\}_{i\in A},\left\{\kappa_i\right\}_{i\in A},\left\{s^i_{t_i}\right\}_{i\in A},\left\{g_j\right\}^m_{j=1},\gamma,c,\phi\right)\propto$$

$$\propto\prod_{i\in A}p\left(\varepsilon^i_{1,t_i}/\lambda,\kappa_i,s^i_{t_i},\left\{g_j\right\}^m_{j=1},\gamma,c,\phi\right)p\left(\lambda\right)$$

$$\propto\left\{\prod_{i\in A}\left(\lambda\kappa_i g_{s^i_{t_i}}k^{i-1}_t\left(\gamma,c,\phi\right)\right)^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\lambda\kappa_i g_{s^i_{t_i}}k^{i-1}_t\left(\gamma,c,\phi\right)\left(\varepsilon^i_{1,t_i}\right)^2\right\}\right\}$$

$$\times\left(\lambda\right)^{\frac{\nu_\lambda}{2}-1}\exp\left\{-\tfrac{1}{2}s^2_\lambda\lambda\right\}$$

$$\propto\left(\lambda\right)^{\frac{\nu_\lambda+q}{2}-1}\exp\left\{-\tfrac{1}{2}\lambda\left[\sum_{i\in A}\kappa_i g_{s^i_{t_i}}k^{i-1}_t\left(\gamma,c,\phi\right)\left(\varepsilon^i_{1,t_i}\right)^2+s^2_\lambda\right]\right\}$$

$$\therefore\lambda/\ldots\sim\Gamma\left(\tfrac{\nu_\lambda+q}{2},\tfrac{1}{2}\left[\sum_{i\in A}\kappa_i g_{s^i_{t_i}}k^{i-1}_t\left(\gamma,c,\phi\right)\left(\varepsilon^i_{1,t_i}\right)^2+s^2_\lambda\right]\right)$$

- $\underline{\left(\mu/\beta,H\right)'}$

$$p\left(\mu/\beta,H\right)\propto\left\{p\left(\beta/\mu,H\right)\right\}p\left(\mu\right)$$

$$\propto\exp\left\{-\tfrac{1}{2}\left(\beta-\mu\right)'H\left(\beta-\mu\right)\right\}\exp\left\{-\tfrac{1}{2}\left(\mu-\underline{\mu}_\mu\right)'\underline{H}_\mu\left(\mu-\underline{\mu}_\mu\right)\right\}$$

$$\propto\exp\left\{-\tfrac{1}{2}\left[\mu'H\mu-2\mu'H\beta+\mu'\underline{H}_\mu\mu-2\mu'\underline{H}_\mu\underline{\mu}_\mu\right]\right\}$$

$$\propto\exp\left\{-\tfrac{1}{2}\left[\mu'\left(H+\underline{H}_\mu\right)\mu-2\mu'\left(H\beta+\underline{H}_\mu\underline{\mu}_\mu\right)\right]\right\}$$

$$\propto\exp\left\{-\tfrac{1}{2}\left[\left(\mu-\overline{\mu}\right)'\overline{H}\left(\mu-\overline{\mu}\right)\right]\right\}$$

where $\overline{H}=\left(H+\underline{H}_\mu\right)$ and $\overline{\mu}=\overline{H}^{-1}\left(H\beta+\underline{H}_\mu\underline{\mu}_\mu\right)$

$$\therefore\mu/\left(\beta,H\right)\sim N\left(\overline{\mu},\overline{H}^{-1}\right)$$

- $\underline{\left(H/\beta,\mu\right)'}$

$$p\left(H/\beta,\mu\right)\propto p\left(\beta/\mu,H\right)p\left(H\right)$$

$$\propto\left|H\right|^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\left(\beta-\mu\right)'H\left(\beta-\mu\right)\right\}\left|H\right|^{\frac{\nu_H-(k+1)-1}{2}}\exp\left\{-\tfrac{1}{2}tr\left(\underline{S}^{-1}_H H\right)\right\}$$

$$\propto\left|H\right|^{\frac{1+\nu_H-(k+1)-1}{2}}\exp\left\{-\tfrac{1}{2}tr\left[\left(\left(\beta-\mu\right)\left(\beta-\mu\right)'+\underline{S}^{-1}_H\right)H\right]\right\}$$

$$\therefore H/\left(\beta,\mu\right)\sim Wishart\left(1+\underline{\nu}_H,\left[\left(\beta-\mu\right)\left(\beta-\mu\right)'+\underline{S}^{-1}_H\right]^{-1}\right)$$

- $\underline{\left(\mathbf{p}/\left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i}\right)'}$

$$p\left(\mathbf{p}/\left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i}\right) \propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i} p\left(s_t^i/\mathbf{p}\right)\right\} p\left(\mathbf{p}\right)$$

$$\propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i} p\left(s_t^i/\mathbf{p}\right)\right\} p\left(\mathbf{p}\right)$$

$$\propto \left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i}\prod_{j=1}^{m} p_j^{I\left(s_t^i=j\right)}\right\}\prod_{j=1}^{m} p_j^{\alpha_j-1}$$

$$\propto \prod_{j=1}^{m} p_j^{\sum_{i=1}^{I}\sum_{t=t_i}^{T_i} I\left(s_t^i=j\right)+\alpha_j-1}$$

$$\therefore \mathbf{p}/\left(\left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i}\right) \sim Dir\left[\left(\sum_{i=1}^{I}\sum_{t=t_i}^{T_i} I\left(s_t^i=j\right)+\alpha_j\right)_{j=1}^{m}\right]$$

- $\underline{\left(\left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i}/\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}, \rho_1, \rho_2, \gamma, c, \lambda, \phi, \left\{\kappa_i\right\}_{i=1}^{I}, \left\{g_j\right\}_{j=1}^{m}, \mathbf{p}\right)'}$

$$p\left(s_t^i=j/\left\{\varepsilon_{h(t),t}^i\right\}_{t=t_i}^{T_i}, \rho_1, \rho_2, \gamma, c, \lambda, \phi, \kappa_i, \left\{g_j\right\}_{j=1}^{m}, \mathbf{p}\right) \propto$$

$$\propto p\left(\eta_t^i/s_t^i=j,\cdots\right) p\left(s_t^i=j/\mathbf{p}\right)$$

$$\propto \underbrace{\begin{array}{c}\lambda^{I\left(t=t_i\hat{\ }i\in A\right)}\left(\kappa_i g_j\right)^{\frac{1}{2}}\left(k_t^i\left(\gamma,c,\phi\right)\right)^{-\frac{1}{2}} \\ \exp\left\{-\frac{1}{2}g_j\kappa_i\left(\begin{array}{c}\left(\lambda\left(\eta_{t_i}^i\right)^2\right) I\left(t=t_i\hat{\ }i\in A\right)+ \\ +\left(1-I\left(t=t_i\hat{\ }i\in A\right)\right)\frac{\left(\eta_t^i\right)^2}{k_t^i(\gamma,c,\phi).}\end{array}\right)\right\}\end{array}}_{w_j}$$

$$\therefore p\left(s_t^i=j/\left\{\varepsilon_{h(t),t}^i\right\}_{t=t_i}^{T_i}, \rho_1, \rho_2, \gamma, c, \lambda, \phi, \left\{\kappa_i\right\}_{i=1}^{I}, \left\{g_j\right\}_{j=1}^{m}, \mathbf{p}\right) = \frac{w_j}{\sum_{j=1}^{m} w_j}$$

- $\underline{\left(\left\{g_j\right\}_{j=1}^{m}/\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i}, \rho_1, \rho_2, \lambda, \gamma, c, \phi, \left\{\kappa_i\right\}_{i=1}^{I}, \left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i}\right)'}$

Let $C_j = \{(i,t) : s_t^i = j\}$ and $\widetilde{n}_j = \#C_j$.

$$p\left(g_j/\left\{\varepsilon_{h(t),t}^i\right\}_{(i,t)\in C_j}, \rho_1, \rho_2, \lambda, \gamma, c, \phi, \left\{\kappa_i\right\}_{(i,t)\in C_j}, \left\{s_t^i\right\}_{(i,t)\in C_j}\right) \propto$$

$$\propto \left\{\prod_{(i,t)\in C_j} p\left(\eta_t^i/s_t^i=j,\cdots\right)\right\} p\left(g_j\right)$$

$$\propto g_j^{\frac{\tilde{n}_j}{2}} \exp\left\{-\frac{1}{2}g_j\left(\sum_{\substack{\{(i,t)\in C_j\cap A^c\}\\ \cup\left\{(i,t)\in C_j\cap A,\text{ with } t\geq t_i+1\right\}}} \kappa_i\frac{\left(\eta_t^i\right)^2}{k_t^i(\gamma,c,\phi).} + \frac{\lambda\kappa_i}{k_{t_i}^i(\gamma,c,\phi)}\left(\eta_{t_i}^i\right)^2 I\left(i\in A\cap C_j\right)\right)\right\}$$

$$\times\left(g_j\right)^{\frac{\nu_g}{2}-1}\exp\left\{-\frac{1}{2}g_j\underline{s}^2\right\}$$

$$\propto\left(g_j\right)^{\frac{\tilde{n}_j+\nu_g}{2}-1}\times$$

$$\times\exp\left\{-\frac{1}{2}g_j\left[\underline{s}^2+\sum_{\substack{\{(i,t)\in C_j\cap A^c\}\\ \cup\left\{(i,t)\in C_j\cap A,\text{ with } t\geq t_i+1\right\}}} \kappa_i\frac{\left(\eta_t^i\right)^2}{k_t^i(\gamma,c,\phi).} + \frac{\lambda\kappa_i}{k_{t_i}^i(\gamma,c,\phi)}\left(\eta_{t_i}^i\right)^2 I\left(i\in A\cap C_j\right)\right]\right\}$$

$$\therefore g_j/\left(\left\{\varepsilon_{h(t),t}^i\right\}_{(i,t)\in C_j},\ldots\right)\sim\Gamma\left(\frac{\tilde{n}_j+\nu_g}{2},\ \frac{\underline{s}^2}{2}+\frac{\sum_{\substack{\{(i,t)\in C_j\cap A^c\}\\ \cup\left\{(i,t)\in C_j\cap A,\text{ with } t\geq t_i+1\right\}}}\kappa_i\frac{\left(\eta_t^i\right)^2}{k_t^i(\gamma,c,\phi).}}{2}+ +\frac{\lambda\kappa_i}{k_{t_i}^i(\gamma,c,\phi)}\frac{\left(\eta_{t_i}^i\right)^2 I(i\in A\cap C_j)}{2}\right)$$

- $$\frac{\left(\phi/\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i},\rho_1,\rho_2,\gamma,c,\lambda,\left\{\kappa_i\right\}_{i=1}^{I},\left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i},\left\{g_j\right\}_{j=1}^{m}\right)'}{p\left(\phi/\left\{\varepsilon_{h(t),t}^i\right\}_{i=1,t=t_i}^{I,T_i},\rho_1,\rho_2,\gamma,c,\lambda,\left\{\kappa_i\right\}_{i=1}^{I},\left\{s_t^i\right\}_{i=1,t=t_i}^{I,T_i},\left\{g_j\right\}_{j=1}^{m}\right)}\propto$$

$$\propto\left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i}p\left(\eta_t^i/s_t^i,\cdots\right)\right\}p\left(\phi\right)$$

$$\propto\left\{\prod_{i=1}^{I}\prod_{t=t_i}^{T_i}\left(k_t^i\left(\gamma,c,\phi\right)\right)^{-\frac{1}{2}}\right\}\exp\left\{-\frac{1}{2}\sum_{i=1}^{I}\kappa_i\sum_{t=t_i}^{T_i}\lambda^{I(i\in A\hat{}t=t_i)}g_{s_t^i}\frac{\left(\eta_t^i\right)^2}{k_t^i(\gamma,c,\phi).}\right\}\frac{1}{\phi}$$

REFERENCES

[1] J. Abowd and D. Card. On the covariance structure of earnings and hours changes. *Econometrica*, 57:411–445, 1989.

[2] S.R. Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684, 1994.

[3] D. Altig, A. J. Auerbach, L. J. Kotlikoff, K. A. Smetters, and J. Walliser. Simulating fundamental tax reform in the united states. *American Economic Review*, 91(3):574–595, 2001.

[4] Michael Baker. Growth-rate heterogeneity and the covariance structure of life-cycle earnings. *Journal of Labor Economics*, 15(2):338–375, April 1997.

[5] Michael Baker and Gary Solon. Earnings dynamics and inequality among canadian men, 1976-1992: Evidence from longitudinal tax records. *Journal of Labor Economics*, 21(2):289–321, 2003.

[6] Robert B. Barsky, Miles S. Kimball, F. Thomas Juster, and Matthew D. Shapiro. Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics*, 112(2):537–579, 1997.

[7] Luc Bauwens, Michel Lubrano, and Jean-Francois Richard. *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, 2000.

[8] Yoram Ben-Porath. The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):352–265, 1967.

[9] Martin Browning, Mette Ejrnæs, and Javier Alvarez. Modelling income processes with lots of heterogeneity. *Review of Economic Studies*, 771(4):1353–1381, 2010.

[10] Fabio Canova. *Methods for Applied Macroeconomic Research*. Princeton University Press, 2007.

[11] Gary Chamberlain and Keisuke Hirano. Predictive distributions based on longitudinal earnings data. *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, 55-56:211–242, 1999.

[12] Kung Sik Chan and Howell Tong. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7(3):179–190, 1986.

[13] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, November 1995.

[14] J.C. Conesa, S Kitao, and D. Krueger. Taxing capital? not a bad idea after all. *American Economic Review*, 99(1):25–48, 2009.

[15] J.C. Conesa and D. Krueger. On the optimal progressivity of the income tax code. *Journal of Monetary Economics*, 53(7):1425–1450, 2006.

[16] T. Cooley and E. Prescott. *Economic Growth and Business Cycle*, chapter 1. Princeton University Press, Princeton, NJ, 1995.

[17] I. Correia. Consumption taxes and redistribution. *American Economic Review*, 100:1673–1694, 2010.

[18] German Cubas and Pedro Silos. The risk premium in labor markets. *Working paper*, mimeo, 2012.

[19] J. Diaz-Gimenez and J. Pijoan-Mas. Flat tax reforms in the u.s.: A boom for the income poor. *University of Carlos III, Madrid*, mimeo, 2005.

[20] Dick van Dijk, Timo Teräsvirta, and Philip Hans Franses. Smooth transition autoregressive models. a survey of recent developments. *Econometric Reviews*, 21(1):1–47, 2002.

[21] D. Domeij and J. Heathcote. On the distributional effects of reducing capital taxes. *International Economic Review*, 45(2):523–554, 2004.

[22] David Domeij and Martin Floden. The labor-supply elasticity and borrowing constraints: Why estimates are biased. *Review of Economic Dynamics*, 9(2):242–262, 2006.

[23] James Durbin and Siem Jan Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–615, September 2002.

[24] J. Eaton and H. Rosen. Labor supply, uncertainty, and efficient taxation. *Journal of Public Economics*, 14(3):363–374, 1980.

[25] A. Erosa and M. Gervais. Optimal taxation in life-cycle economies. *Journal of Economic Theory*, 105(2):338–369, 2002.

[26] Brian S. Everitt and D.J Hand. *Finite Mixture Distributions*. Chapman and Hall, 1981.

[27] M. Feldstein. On the optimal progressivity of the income tax. *Journal of Public Economics*, 2(4):357–376, 1973.

[28] M. Flodén. A note on the accuracy of markov-chain approximations to highly persistent ar(1)-processes. *Economics Letters*, 99(3):516–520, 2008.

[29] M. Flodén and J. Lindé. Idiosyncratic risk in the u.s. and sweden: Is there a role for government insurance? *Review of Economic Dynamics*, 4(2):406–437, 2001.

[30] Martin Floden. The effectiveness of government debt and transfers as insurance. *Journal of Monetary Economics*, 48:81–108, 2001.

[31] M. Friedman. *Capitalism and Freedom*. University of Chicago Press, 1962.

[32] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

[33] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[34] John Geweke. *Contemporary Bayesian Econometrics and Statistics*. Wiley-Interscience, September 2005.

[35] John Geweke and Michael Keane. An empirical analysis of earnings dynamics among men in the psid: 1968-1989. *Journal of Econometrics*, 96(2):293–356, 2000.

[36] Peter Gottschalk. Earnings mobility: Permanent change or transitory fluctuations? *The Review of Economics and Statistics*, 64(3):450–456, 1982.

[37] Peter Gottschalk and Robert Moffitt. Trends in the transitory variance of earnings in the united states. *Economic Journal*, 112(C):68–73, 2002.

[38] M. Gouveia and R. Strauss. Effective federal individual income tax functions: An exploratory empirical analysis. *National Tax Journal*, 47(2):317–339, 1994.

[39] Jeremy Greenwood, Nezih Guner, and John A. Knowles. More on marriage, fertility, and the distribution of income. *International Economic Review*, 44(3):827–862, 2003.

[40] N. Guner, G. Ventura, and R. Kaygusuz. Taxation, aggregates and the household. page CEPR Discussion Paper 6702 and IZA Discussion Paper 3318, 2008.

[41] Fatih Guvenen. Learning your earning: Are labor income shocks really very persistent? *The American Economic Review*, 97(3):687–712, June 2007.

[42] Fatih Guvenen. An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12(1):58–79, January 2009.

[43] Fatih Guvenen, Serdar Ozkan, and Jae Song. The nature of countercyclical income risk. *University of Minnesota. Economics Department*, mimeo, 2012.

[44] Steven J. Haider. Earnings instability and earnings inequality of males in the united states: 1967-1991. *Journal of Labor Economics*, 19(4):799–836, October 2001.

[45] R. Hall and A. Rabushka. *The Flat Tax*. Hoover Classics, Hoover Institution Press, 1985.

[46] G. Hansen. The cyclical and secular behavior of the labor input: Comparing efficiency units and hours worked. *Journal of Applied Econometrics*, 8(1):71–80, 1993.

[47] W. Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[48] John C. Hause. The fine structure of earnings and the on-the-job training hypothesis. *Econometrica*, 48(4):1013–1029, 1980.

[49] J. Heathcote, K. Storesletten, and G. Violante. The macroeconomic implications of rising wage inequality in the united states. *Journal of Political Economy*, 118(4):681–722, 2010.

[50] Jonathan Heathcote, Kjetil Storesletten, and Gianluca Violante. The macroeconomic implications of rising wage inequality in the united states. *Journal of Political Economy*, 118(4):681–722, 2010.

[51] H. Hopenhayn and E. Prescott. Stochastic monotonicity and stationary distributions for dynamic economies. *Econometrica*, 60(6):1387–1406, 1992.

[52] Laura Hospido. Modelling heterogeneity and dynamics in the volatility of individual wages. *Journal of Applied Econometrics*, 27(3):386–414, 2012.

[53] Dmytro Hryshko. Rip to hip: The data reject heterogeneous labor income profiles. *Quantitative Economics*, forthcoming, 2012.

[54] M. Huggett. The risk free rate in heterogenous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5-6):953–969, 1993.

[55] Mark Huggett, Gustavo Ventura, and Amir Yaron. Sources of lifetime inequality. *American Economic Review*, 101(7):2923–2954, 2011.

[56] Shane T. Jensen and Stephen Shore. Semi-parametric bayesian modeling of income volatility heterogeneity. *Journal of the American Statistical Association*, 106(496):1280–1290, 2012.

[57] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[58] Gueorgui Kambourov and Iourii Manovskii. Occupational mobility and wage inequality. *Review of Economic Studies*, 76(2):731–759, April 2009.

[59] Greg Kaplan. Moving back home:insurance against labor market risk. *Journal of Political Economy*, forthcoming, 2012.

[60] Yasar F. Karahan and Serdar Ozkan. On the persistence of income shocks over the life cycle: Evidence, theory, and implications second version. *PIER Working Paper*, 1-303:Department of Economics, University of Pennsylvania, 2011.

[61] Michael P. Keane and Kenneth I. Wolpin. The career decisions of young men. *Journal of Political Economy*, 105(3):473–522, June 1997.

[62] John A. Knowles. Why are married men working so much? relative wages, labor supply and the decline of marriage. *Working Paper*, mimeo, January 2012.

[63] Oleg Korenok. Bayesian methods in nonlinear time series. *Encyclopedia of Complexity and System Science. Springer*, page Entry 217, 2009.

[64] D. Krueger and F. Perri. Public versus private risk sharing. *Journal of Economic Theory*, 146:920–956, 2010.

[65] Lee A. Lillard and Yoram Weiss. Components of variation in panel earnings data: American scientists 1960-70. *Econometrica*, 47(2):437–454, 1979.

[66] Lee A. Lillard and Robert J. Willis. Dynamic aspects of earning mobility. *Econometrica*, 46(5):985–1012, September 1978.

[67] Hedibert F. Lopes and Esther Salazar. Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis*, 27(1):99–117, 2006.

[68] Ritva Luukkonen, Pentti Saikkonen, and Timo Teräsvirta. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499, 1988.

[69] Thomas E. MaCurdy. The use of time series processes to model the error structure of earnings in alongitudinal data analysis. *Journal of Econometrics*, 18(1):83–114, 1982.

[70] John J. McCall. *Income Mobility, Racial Discrimination, and Economic Growth.* Lexington Books, Lexington, MA, 1973.

[71] Geoffrey McLachlan and David Peel. *Finite Mixture Models.* Wiley Series in Probability and Statistics, 2000.

[72] Costas Meghir and Luigi Pistaferri. Income variance dynamics and heterogeneity. *Econometrica*, 72(1):1–32, 2004.

[73] Costas Meghir and Frank Windmeijer. Moment conditions for dynamic panel data models with multiplicative individual effects in the conditional variance. *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, 55-56:317–330, 1999.

[74] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[75] R. Moffit. The negative income tax and the evolution of u.s. welfare policy. *Journal of Economic Perspectives*, 17(3):119–140, 2003.

[76] Peter Müller. A generic approach to posterior integration and gibbs sampling. *Purdue University, Department of Statistics. Technical report*, pages 91–09, 1991.

[77] Taisuke Nakata and Christopher Tonetti. Small sample properties of bayesian estimators of labor income processes. *NYU. Economics Department*, mimeo, 2012.

[78] S. Nishiyama and K. Smetters. Consumption taxes and economic efficiency with idiosyncratic wage shocks. *Journal of Political Economy*, 113(5):1088–1115, 2005.

[79] Andriy Norets and Sam Schulhofer-Wohl. Heterogeneity in income processes. mimeo, 2009.

[80] L. Ohanian, A. Raffo, and R. Rogerson. Long-term changes in labor supply and taxes: Evidence from oecd countries, 1956-2004. *Journal of Monetary Economics*, 55:1353–1362, 2008.

[81] J. Pijoan-Mas. Precautionary savings or working longer hours? *Review of Economic Dynamics*, 9(2):326–352, 2006.

[82] Luigi Pistaferri. Anticipated and unanticipated wage changes, wage risk and intertemporal labor supply. *Journal of Labor Economics*, 21(3):729–754, 2003.

[83] E. Prescott. Why do americans work so much more than europeans? *Federal Reserve Bank of Minneapolis Quarterly Review*, 28:2–13, 2004.

[84] J.V. Rios-Rull. *Frontiers of Business Cycle Research*, chapter 4. Princeton University Press, Princeton, NJ, 1995.

[85] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, second edition, November 2010.

[86] R. Rogerson and J. Wallenius. Micro and macro elasticities in a life cycle model with taxes. *Journal of Economic Theory*, 144:2277–2292, 2009.

[87] Stephen Shore. The intergenerational transmission of income volatility: Is riskiness inherited? *Journal of Business and Economic Statistics*, 29(3):372–381, 2011.

[88] Anthony F. Shorrocks. Income mobility and the markov assumption. *Economic Journal*, 86(343):566–578, 1976.

[89] Christopher A. Sims and Harald Uhlig. Understanding unit rooters: a helicopter tour. *Econometrica*, 59(6):1591–1599, 1991.

[90] K. Storesletten, C. Telmer, and A. Yaron. Consumption and risk sharing over the life cycle. *Journal of Monetary Economics*, 51(3):609–633, 2004.

[91] Kjetil Storesletten, Christopher I. Telmer, and Amir Yaron. Consumption and risk sharing over the life cycle. *Journal of Monetary Economics*, 51(3):609–633, April 2004.

[92] G. Tauchen and R. Hussey. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, 59(2):371–396, 1991.

[93] Timo Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218, 1994.

[94] D. M. Titterington, Adrian F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section, 1985.

[95] Howell Tong. On a threshold model. *In Pattern recognition and signal processing. NATO ASI Series E: Applied Sc.*, 29:575–586, 1978.

[96] Howell Tong and K.S. Lim. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(3):245–292, 1980.

[97] Towell Tong. *Non-Linear Time Series: A Dynamical System Approach*, volume 6 of *Oxford Statistical Science Series*. Oxford University Press, 1993.

[98] Ruey S. Tsay. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405):231–240, March 1989.

[99] Pierre Vandekerkhove. Consistent and asymptotically normal parameter estimates for hidden markov mixtures of markov models. *Bernoulli*, 11(11):103–129, 2005.

[100] G. Ventura. Flat tax reform: A quantitative exploration. *Journal of Economic Dynamics and Control*, 23(9-10):1425–1458, 1999.