Letting Your Past Define Your Taxes:
Optimal History-Dependent Income Taxation in General Equilibrium

Ross Batzer*

University of Minnesota

November 5, 2021

* Email: batz0025@umn.edu, Website: https://sites.google.com/view/ross-batzer

Introduction

- How should labor income taxation vary with a household's income history?
- Existing literature (Dynamic Mirrlees)
  - Optimal history-dependent (HD) taxation *without* general equilibrium (GE) wages
  - General finding: very small welfare benefits from HD taxes over just age-dependent
  - Open question whether this is robust to environments with GE effects
- Another literature (Quantitative Ramsey)
  - Optimal labor income taxation in rich models of life-cycle labor supply with GE effects
  - But optimal taxes are *restricted* to be simple parametric functions in only current income
  - Computational challenges have prevented more complex tax functions or HD

Motivation

- To study history-dependent taxation in a model with GE wages, I need
  - Taxes that can depend on entire income history
  - Flexibility of parametric (Quantitative Ramsey) approach
- Neural networks (NN) and recent advancements in computing make this possible
  - NN's approximate high-dimensional, nonlinear functions (in finite time)
  - Estimating NN's is fast thanks to new numerical libraries (e.g. Google's Tensorflow)

What I Do

- Study history-dependent taxes in a general equilibrium model of life-cycle labor supply
  - Similar to Heathcote, Storesletten and Violante (2020) (HSV)
    - PSID households: \$15K-\$500K in yearly income
  - CES production: different household skill types are complementary
  - Up to 41 state variables: age, prod. & pref. shocks, savings, skills, 35-year income history
- Consider very general class of tax function
  - Avg. tax rate = *any* continuously differentiable function of entire income history
- Use NN's to approximate optimal labor income tax function in steady state
  - Existing methods: use NN's to solve high-dimensional structural models
  - My method (Nested NN): use NN's to both solve the model *and* optimize income taxes

▸ How it Works

Main Findings

1. Optimal history-dependent (HD) tax function is:
    - For most people: progressive (i.e. avg tax rates increase) in current income and age
    - For high income history (rarest skills): regressive in income history
        - Gov. wants to separate types and reward high output
        - High output $\implies$ high average wages through GE effects
2. Welfare gains from HD taxation are large compared to optimal age-dependent tax (2% lifetime cons.)
    - GE drives benefits: no gains from HD taxation when skills are perfect substitutes
3. Similar welfare achieved with parametric tax: $T($*average* past income, age$)$
    - Higher past income $\implies$ taxes increase more slowly with age ($\frac{d^2T}{d\bar{y}\ dage} < 0$)

Related Literature

- Parametric Optimal Taxation (aka Quantitative Ramsey)
  - Kindermann and Krueger (2021), Heathcote, Storesletten and Violante (2017, 2020), Stantcheva (2020), Krueger and Ludwig (2016), Karabarbounis (2016), Peterman (2016), Gervais (2012), Huggett and Parra (2010), Conesa, Kitao and Krueger (2009), Conesa and Krueger (2006), Erosa and Gervais (2002), and others
  - History-dependence (partial equilibrium): Kapička (2020)
  - **My contribution:** General tax function that allows for history-dependence
- Neural networks to solve structural economic models
  - Azinovic, Gaegauf and Scheidegger (2019), Chen, Joseph, Kumhof and Pan (2021),Duarte (2018), Fernández-Villaverde, Nuño, Sorg-Langhans and Vogler (2020), Maliar, Maliar and Winant (2019), and others
  - **My contribution:** Application to optimal policy (nested neural network solution)
- Nonparametric Optimal Taxation
  - History Dependent (aka Dynamic Mirrlees): Ndiaye (2020), Stantcheva (2020), Stantcheva (2017), Golosov, Troshkin and Tsyvinski (2016), Golosov and Tsyvinski (2015), Farhi and Werning (2013), Fukushima (2011), Weinzierl (2011), Albanesi and Sleet (2006), Golosov, Kocherlakota and Tsyvinski (2003), and others
  - Nonparametric Taxes on Current Income (aka Variational Approach): Chang and Park (2020), Saez and Stantcheva (2018), Findeisen and Sachs (2017), and others
  - Static optimal taxation: Heathcote and Tsujiyama (2020), Sachs, Tsyvinski and Werquin (2020), Saez (2001), and others
- Optimal taxation with differentiated labor
  - Stiglitz (1982, 1987), Rothschild and Scheuer (2013), Scheuer and Werning (2016) and others

Plan

1. Model
   - Optimal taxation problem
2. Parameter Selection
3. Solution Method
4. Results
   - Optimal tax functions w/ and w/o differentiated skills
   - Welfare comparison under different tax functions
   - Simple implementation

Model

Agents

- Households (measure one, indexed by $i \in [0,1]$)
  - Work for A periods, age indexed by $a = 0, \ldots, A-1$
  - Choose one time, permanent investment in skills, s, before entering labor market
  - Each period choose consumption c_a, savings b_{a+1} and hours h_a
- Firm: produces consumption good using labor differentiated by skill s
- Government
  - Collects revenue from nonparametric history-dependent income taxes, $T(.)$
  - Uses tax revenue to fund expenditures, $G = g \times Y$ (constant % of output)
  - Taxes distort labor supply *and* skill investment: limit on optimal progressivity

Technology

- Output is a CES aggregate over continuum of skill types s

$$Y = \left(\int_0^\infty [N(s) f_s(s)]^{\frac{\omega-1}{\omega}} ds \right)^{\frac{\omega}{\omega-1}}$$

  - $N(s)$ is total labor supply and $f_s(s)$ is density for type s
  - ω is elasticity of substitution between skills
  - Note: higher skills not inherently more productive, but more valuable because rarer
    - Price for skills/skill premium: $p(s) = \left[\frac{Y}{N(s) f_s(s)} \right]^{\frac{1}{\omega}}$
    - Higher wage when Y larger $\implies$ spillovers from higher output

- Linear savings technology: households can transform one unit of consumption at a into $1+r$ units at $a+1$ risk-free

- Resource constraint

$$\sum_{a=0}^{A-1} \int (c_{ia} + b_{ia+1}) di + gY = (1+r) \sum_{a=0}^{A-1} \int b_{ia} di + Y$$

Individual Wages and Income

- Hourly productivity:

$$\log\theta_{ia} = x(a) + z_{ia} + \varepsilon_{ia}$$

  - $x(a)$: deterministic age-productivity profile
  - $z_{ia} = z_{ia-1} + \eta_{ia}$, $\eta_{ia} \sim N(0, v_\eta)$: permanent shocks
  - $\varepsilon_{ia} \sim N(0, v_\varepsilon)$: transitory shocks

- Total labor income:

$$y_{ia} = \underbrace{p(s_i)}_{\text{skill price}} \times \underbrace{\theta_{ia}}_{\text{productivity}} \times \underbrace{h_{ia}}_{\text{hours worked}}$$

  - $p(s)$: skill price of type s labor = marginal product of type s

Preferences

- Preferences over consumption c, hours h, and skill-investment s for an individual i

$$U_i = -v_i(s_i) + E_0\left[\left(\frac{1-\beta}{1-\beta^A}\right)\sum_{a=0}^{A-1}\beta^a u_i(c_{ia},h_{ia})\right]$$

where disutility from skill investment is

$$v_i(s_i) = \kappa_i^{-\frac{1}{\psi}}\frac{s_i^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}},\ \kappa_i \sim Exp(1)$$

and utility from consumption and labor is

$$u_i(c_{ia},h_{ia}) = \frac{\left[c_{ia}^{\phi_i}(1-h_{ia})^{1-\phi_i}\right]^{1-\gamma}}{1-\gamma}$$

where $\phi_i = \frac{1}{1+\exp\tilde{\phi}_i}$, $\tilde{\phi}_i \sim N(m_\phi, v_\phi)$

Household Problem

- Denote the vector of individual state variables as

$$S_{ia} \equiv (s_i, \phi_i, b_{ia}, z_{ia}, \varepsilon_{ia}, a, \{y_t\}_{t=0}^{a-1})$$

  - 41 individual state variables
    - 4 exogenous: $(\phi_i, z_{ia}, \varepsilon_{ia}, a)$
    - $2+a-1$ (37 when $a = A-1$) endogenous: $(s_i, b_{ia}, \{y_t\}_{t=0}^{a-1})$

- Individuals enter with zero savings $b_{i0} = 0$ and solve

$$v_{ia}(S_{ia}) = \max_{c_{ia}, h_{ia}, b_{ia+1}} u_i(c_{ia}, h_{ia}) + \beta E_a[v_{ia+1}(S_{ia+1})]$$

subject to

$$c_{ia} + b_{ia+1} = (1+r)b_{ia} + y_{ia} - T_a(y_{ia}; \{y_{ij}\}_{j=0}^{a-1})$$

and

$$c_{ia}, b_{ia+1} \geq 0, \;\; h_{ia} \in [0,1]$$

Equilibrium

- Stationary equilibrium is allocation functions $\left(s, \{c_a, h_a, b_{a+1}\}_{a=0}^{A-1}\right)$ and prices $p(s)$ such that
  - Households solve their problem
  - Skill price $p(s)$ is the marginal product of type s

$$p(s) = \left[\frac{Y}{N(s) f_s(s)}\right]^{\frac{1}{\omega}}$$

  - Densities for skills f_s and savings f_b are consistent with individual choices
  - Government budget is satisfied

$$gY \leq \sum_{a=0}^{A-1} \int T_a\left(y_{ia}; \{y_{it}\}_{t=0}^{a-1}\right) di$$

  - Markets clear

$$\sum_{a=0}^{A-1} \int c_{ia} di + gY = r \sum_{a=0}^{A-1} \int b_{ia} di + Y \text{ and } N(s) = \sum_{a=0}^{A-1} \int h_{ia}(s) \exp\{x(a) + z_{ia} + \varepsilon_{ia}\} di$$

Optimal Tax Problem

- Government's social welfare function is ex-ante expected utility of a household born into a stationary equilibrium

$$W = \int U_i di = \int \left\{ -v_i(s_i) + E_0 \left[\left(\frac{1-\beta}{1-\beta^A} \right) \sum_{a=0}^{A-1} \beta^a u_i(c_{ia}, h_{ia}) \right] \right\} di$$

- Government chooses the tax function $T(.)$ to maximize W subject to its budget

$$gY = \sum_{a=0}^{A-1} \int T_a \left(y_{ia}; \{y_{ij}\}_{j=0}^{a-1} \right) di$$

  and that households solve their problem given the tax function

- Tax function $T(.)$ is of the form

$$T_a \left(y_a; \{y_j\}_{j=0}^{a-1} \right) = \tau_a \left(y_a; \{y_j\}_{j=0}^{a-1} \right) y_a$$

  - τ_a is a continuously differentiable function that depends on age and income history
  - Can prove this form has a unique mapping between tax functions and equilibrium allocations

Parameter Selection

Wage Estimation

- I estimate processes for wage shocks using data from the PSID
  - Regress log wages on a polynomial in age and demographic dummies

$$\log w_{ia} = x_0 + x_1 a + x_2 a^2 + D_i + \epsilon_{ia}$$

- Gives age profile $x(a) = x_0 + x_1 a + x_2 a^2$ and stochastic component of wages ϵ_{ia}
- Assume ϵ_{ia} is composed of permanent component z and transitory component ε

$$\epsilon_{ia} = z_{ia} + \varepsilon_{ia}$$

where

$$z_{ia+1} = z_{ia} + \eta_{ia+1}, \; \eta_{ia} \sim N(0, v_\eta)$$
$$z_{i0} \sim N(0, v_z) \text{ and } \varepsilon_{ia} \sim N(0, v_\varepsilon)$$

- Estimate values of $(v_\eta, v_z, v_\varepsilon)$ to match $var(\epsilon_{ia}), var(\epsilon_{ia+2} - \epsilon_{ia})$ and $var(\epsilon_{ia+4} - \epsilon_{ia})$

Wage Parameters

Parameter	Description	Value
x_1	Linear component of life cycle profile	0.031
x_2	Quadratic component of life cycle profile	-0.0005
v_z	Variance of initial condition z_0	0.120
v_η	Variance of permanent shocks z	0.003
v_ε	Variance of transitory shocks ε	0.135

Table: Summary of Parameters for Wage Process

Fixed Parameters

Parameter	Description	Value	Source/Target
A	Years of working life	36	Heathcote et al. (2020)
ψ	Elasticity of skill investment to return	0.65	
ω	Elasticity of substitution across skills	3.124	
g	Government spending (% of output)	0.19	
β	Discount Factor	0.98	Golosov et al. (2016)
R	Return on savings	1/0.98	
m_ϕ	Mean of leisure disutilty	0.275	$H = 0.33$
v_ϕ	Variance of labor disutilty utility	0.026	$var(\log h_i) = 0.12$

Table: Summary of Fixed Parameters

Solution Method

Solving the Optimal Tax Problem

- Optimizing tax function directly requires keeping track of every previous level of income
  - With $A = 36$ periods of work, optimal tax has up to 36 arguments
  - Every argument of tax function becomes a state variable in HH problem
- Question: How do I approximate a nonlinear function with at least 36 arguments?

Polynomial vs NN Approximation

- Suppose I tried to approximate the same tax function with polynomials and NN's
- With polynomials (e.g. Chebychev),
  - Number of parameters to estimate grows exponentially in inputs
  - With $A = 36$ and just affine/linear functions, would need $2^{36} \approx 70$ billion parameters
    - Approximation is infeasible, even for crude approximation
- With neural networks
  - Number of parameters grows linearly in inputs
  - Only have to approximate several thousand of parameters ($\approx 4{,}000$) instead of billions

Details

How it Works

- Sketch of optimal tax algorithm:
  1. Approximate tax function as a neural network
  2. Calculate change in social welfare from changes in tax function
     - Calculate how individual choices and GE prices change under perturbed tax function
  3. Update tax function to maximize welfare (gradient descent)
- Similar to variational method used by Saez (2001) to derive optimal nonlinear tax in static economy
  - Optimal tax function is such that no change produces a welfare gain

◂ Back to Intro
▸ Technical Version ▸ Check the Approximation

Results

Outline of Results

1. Optimal taxes
   - Baseline: CES Production with unobservable skill types ▸ Here
   - Counterfactual 1: Skill types are observable (i.e. taxes can depend on s) ▸ Here
   - Counterfactual 2: Skills are perfect substitutes in production ▸ Here
2. Compare steady state welfare under HD optimum to different tax functions ▸ Here
   - Parametric and nonparametric on current income, $T(y)$, then current income and age, $T(y, a)$
3. Simple implementation ▸ Here
   - Parametric tax in avg. lifetime income gets most welfare gains from full HD policy

▸ Results with Separable Utility

Conclusion

- What I did
  - Used NN's to compute optimal history-dependent taxes in an OLG economy
- What I found
  - Welfare gain from history-dependence can be large
  - Elasticity of substitution between skill types is critical to optimal policy
- Why it was useful
  - Computing rich optimal tax function guides which more easily interpretable policies to study
  - Here, simple function in age and avg past income gets close to full HD optimum

Thank You!

References

Albanesi, Stefania and Christopher Sleet (2006) “Dynamic Optimal Taxation with Private Information,” *The Review of Economic Studies*, 73 (1), 1–30.

Azinovic, Marlon, Luca Gaegauf, and Simon Scheidegger (2019) “Deep Equilibrium Nets,” *Working Paper*.

Benabou, Roland (2000) “Unequal Societies: Income Distribution and the Social Contract,” *American Economic Review*, 90 (1), 96–129.

——— (2002) “Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?” *Econometrica*, 70 (2), 481–517.

Chang, Yongsung and Yena Park (2020) “Optimal Taxation with Private Insurance,” *Working Paper*.

Chen, Mingli, Andreas Joseph, Michael Kumhof, and Xinlei Pan (2021) “Deep Reinforcement Learning in a Monetary Model,” *Working paper*.

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger (2009) “Taxing Capital? Not a Bad Idea after All!,” *American Economic Review*, 99 (1), 25–48.

Conesa, Juan Carlos and Dirk Krueger (2006) “On the Optimal Progressivity of the Income Tax Code,” *Journal of Monetary Economics*, 53 (7), 1425–1450.

Duarte, Victor (2018) “Machine Learning for Continuous Time Finance,” *Working Paper*.

Erosa, Andres and Martin Gervais (2002) “Insurance and Taxation over the Life Cycle,” *Review of Economic Studies*, 105 (2), 338–369.

Farhi, Emmanuel and Ivan Werning (2013) “Insurance and Taxation over the Life Cycle,” *Review of Economic Studies*, 810 (2), 596–635.

Fernández-Villaverde, Jesús, Galo Nuño, George Sorg-Langhans, and Maximilian Vogler (2020) “Solving High-Dimensional Dynamic Programming Problems using Deep Learning,” *Working Paper*.

Findeisen, Sebastian and Dominik Sachs (2017) “Redistribution and Insurance with Simple Tax Instruments,” *Journal of Public Economics*, 146, 58–78.

Fukushima, Kenichi (2011) “Quantifying the Welfare Gains from Flexible Dynamic Income Tax Systems,” *Global COE Hi-Stat Discussion Paper Series* (176).

Gervais, Martiin (2012) “On the optimality of age-dependent taxes and the progressive US tax system,” *Journal of Economic Dynamics and Control*, 36 (4), 682–691.

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski (2003) “Optimal Indirect and Capital Taxation,” *Review of Economic Studies*, 70, 569–587.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2016) “Redistribution and Social Insurance,” *American Economic Review*, 106 (2), 359–386.

Golosov, Mikhail and Aleh Tsyvinski (2015) “Policy Implications of Dynamic Public Finance,” *Annual Review of Economics*, 7, 147–171.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017) “Optimal Tax Progressivity: An Analytical Framework,” *The Quarterly Journal of Economics*, 132 (4), 1693–1754.

——— (2020) “Optimal Progressivity with Age-Dependent Taxation,” *Journal of Public Economics*.

Heathcote, Jonathan and Hitoshi Tsujiyama (2020) “Optimal Income Taxation: Mirrlees Meets Ramsey,” *Working Paper*.

Huggett, Mark and Juan Carlos Parra (2010) “How Well Does the U.S. Social Insurance System Provide Social Insurance?” *Journal of Political Economy*, 118 (1).

Kapička, Marik (2020) “Quantifying the Welfare Gains from History Dependent Income Taxation,” *Working paper*.

Karabarbounis, Marios (2016) “A road map for efficiently taxing heterogeneous agents,” *American Economic Journal: Macroeconomics*, 8, 182–214.

Kindermann, Fabian and Dirk Krueger (2021) “High Marginal Tax Rates on the Top 1*American Economic Journal: Macroeconomics*, forthcoming.

Krueger, Dirk and Alexander Ludwig (2016) “On the optimal provision of social insurance: Progressive taxation versus education subsidies in

Checking the Approximation

Back

- I can check if the NN is finding a maximum of the HH problem by manually computing derivatives around the final allocations

	Percentile of Distribution					
	10%	25%	50%	75%	90%	Average
First ($\partial V/\partial h$)	-1.4 $\times 10^{-2}$	-6.5 $\times 10^{-3}$	-1.6 $\times 10^{-4}$	6.4 $\times 10^{-3}$	1.5 $\times 10^{-2}$	7.9 $\times 10^{-4}$
Second ($\partial^2 V/\partial h^2$)	-0.228	-0.193	-0.134	-0.110	-0.086	-0.151

Table: Derivatives of Household Value Function at Approximation

- First derivatives are near zero and second derivatives are negative $\implies$ NN is finding a maximum

Checking the Approximation ◂ Back

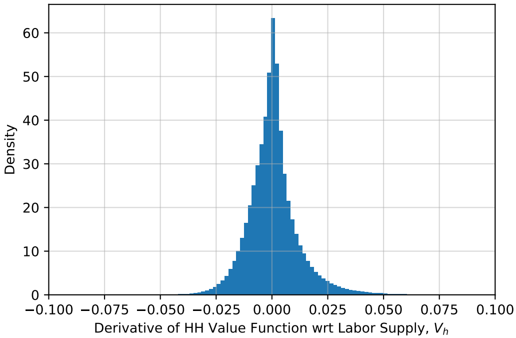
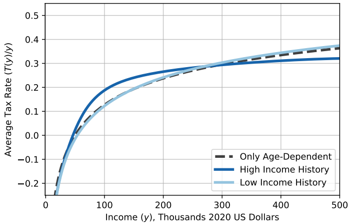


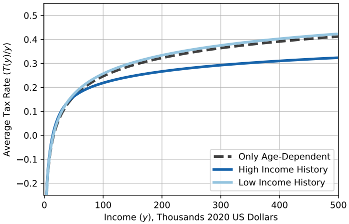
Figure: Histogram of Derivatives of HH Value Function ($\partial V/\partial h$)

Optimal History-Dependent Tax, Differentiated Types

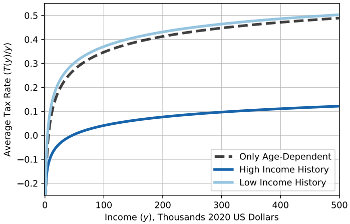
Back



(a) Optimal Tax, Age 30



(b) Optimal Tax, Age 45



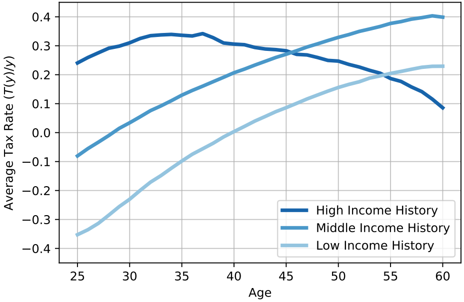
(c) Optimal Tax, Age 60

Figure: Optimal Age and History-Dependent Tax Schedules

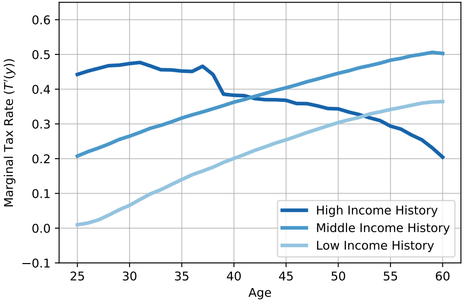
- Tax rates of two households after 5, 20 and 35 years of work
  - Receive low (25th percentile) and high (99th) income each period
  - Low income history: similar to optimal AD tax
  - High income history: tax decreases over lifecycle $\implies$ separate types and reward high output

Tax Rates Over the Life-cycle

Back



(a) Average Tax Rates



(b) Marginal Tax Rates

Figure: Optimal History-Dependent Tax by Age

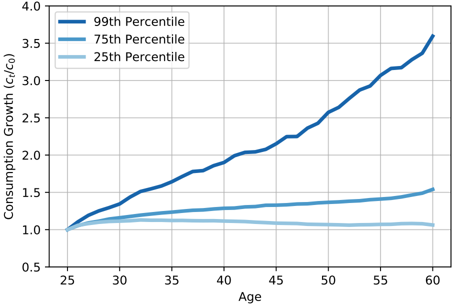
- Follow HH's that receive low (25th percentile), middle (75th) and high (99th) income each period

Discussion ◂ Back

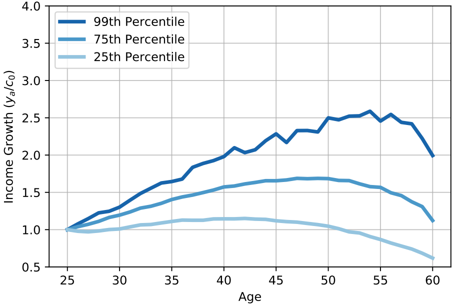
- Why do history-dependent (HD) rewards increase labor supply?
- Tax on **current income** doesn't change labor supply w/ balanced growth pref.
  - Substitution effect ($\uparrow w \implies \uparrow h$) and income effect ($\uparrow c \implies \downarrow h$) cancel
- **HD tax** can change labor supply with pure *substitution effect* early in life
  - HD tax $\implies$ incentivize work today w/ future payoff
  - Increase effective wage w/o changing consumption ($\uparrow w$, no change $c \implies \uparrow h$)
- But **HD tax** gives *income effect* later in life
  - HD tax $\implies$ payoff from previous work ($\uparrow c$)
  - Wages increase w/ age $\implies$ high consumption later in life ($\uparrow c$ when c high => small $\downarrow h$)
- Substitution effect early in life > income effect later in life $\implies$ overall $\uparrow h$

Optimal HD Tax: Consumption and Income Profiles

Back



(a) Percentiles of Consumption by Age



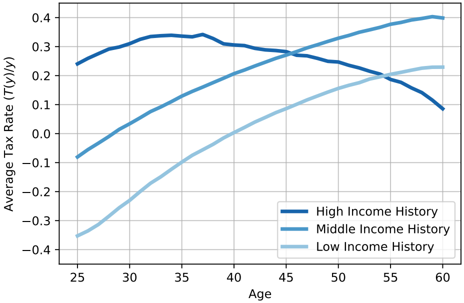
(b) Percentiles of Income by Age

Figure: Life-cycle Percentiles of Consumption and Income

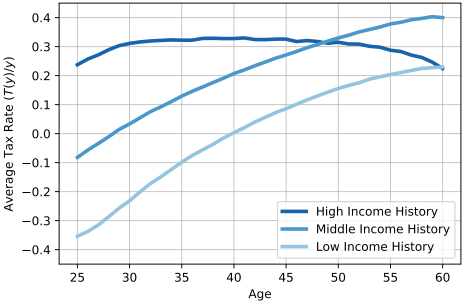
- Consumption and income growth for 25th, 75th and 99th percentile of consumption and income
- Gov. loads consumption later in life when wage is high (smaller income effect)

Optimal HD Tax when Skills Observable

Back



(a) Average Tax Rates, Unobservable Skills



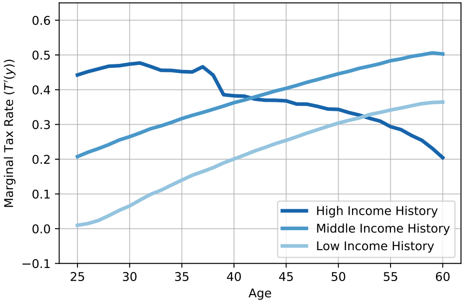
(b) Average Tax Rates, Observable Skills

Figure: Optimal History-Dependent Tax by Age

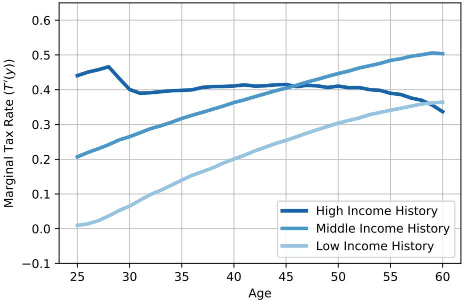
- Follow HH's that receive low (25th percentile), middle (75th) and high (99th) income each period
- Taxes flatter over life-cycle for high income history when skills are observable
  - Don't need to separate types

Optimal HD Tax when Skills Observable

Back



(a) Marginal Tax Rates, Unobservable Skills



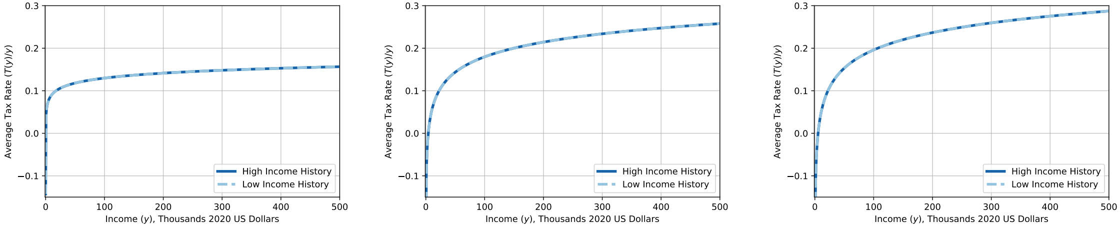
(b) Marginal Tax Rates, Observable Skills

Figure: Optimal History-Dependent Tax by Age

- Marginal taxes also higher and flatter in observable case

Optimal HD Tax when Skills are Perfect Substitutes

Back



(a) Optimal Tax After 10 Years (b) Optimal Tax After 20 Years (c) Optimal Tax After 30 Years

Figure: Optimal Non-parametric History-Dependent Tax

- Tax rates of two households after 10, 20 and 30 years of work
  - Exact same income every year: \$50K (Low) and \$400K (High)
  - Taxes are virtually identical (and so is welfare)
  - Progressivity now increases with age since variance of productivity increases w/ age

Consumption Equivalent Welfare Gains ◂ Back

- To compare welfare impact of each policy, I use consumption equivalent welfare
- Consumption-equivalent welfare gain of moving to a new tax function T^* from T is the g such that

$$W\left(s^*, \{(c_a^*, h_a^*\}_{a=0}^{A-1}; T^*\right) = W\left(s, \{(1+g)c_a, h_a\}_{a=0}^{A-1}; T\right)$$

where W is social welfare

$$W\left(s, \{c_a, h_a\}_{a=0}^{A-1}; T\right) = \int \left\{ -v_i\left(s_i(T)\right) + E_0\left[\left(\frac{1-\beta}{1-\beta^A}\right)\sum_{a=0}^{A-1}\beta^a u_i\left(c_{ia}(T), h_{ia}(T)\right)\right]\right\} di$$

- g : percent gain of lifetime consumption necessary to deliver same welfare as T^*

Comparison with Restricted Tax Systems

- I compare the full history-dependent tax to more restricted systems
- Parametric tax functions that depend on (log-linear class used by Benabou (2000, 2002), Karabarbounis (2016) and Heathcote, Storesletten and Violante (2017, 2020))
  - Current income
  $$T(y) = y - (1-\tau)y^{\rho}$$
  - Current income and age
  $$T_a(y) = y - (1-\tau(a))y^{\rho(a)}$$
  $$\tau(a) = \tau_0 + \tau_1 a + \tau_2 a^2 \text{ and } \rho(a) = \rho_0 + \rho_1 a + \rho_2 a^2$$
- Nonparametric tax functions that can depend on
  - Only current income
  $$T(y) = \tau(y)y$$
  - Age and current income
  $$T_a(y) = \tau(y,a)y$$
  where τ can be *any* differentiable function

Welfare Gains

Back

- Same welfare gains with two different normalizations
- Gains moving *from* simplest policy:

	Income History	Age and Current Income	Only Current Income
Nonparametric	6.31%	4.41%	0.82%
Log-linear		4.25%	0.0%

Table: Welfare Gain of Moving from Parametric Tax on Current Income

- Gains moving *to* most complex policy:

	Income History	Age and Current Income	Only Current Income
Nonparametric	0.0%	1.82%	5.43%
Log-linear		1.98%	6.31%

Table: Welfare Gain of Moving to Nonparametric, History Dependent Taxes

Optimal Restricted Tax Functions | Summary of Tax Functions | Allocations | Present Value of Taxes Paid

What does this mean for policy? Back

- Non-parametric HD policy not easily implemented in reality and hard to interpret
  - What simple parametric policy can achieve similar levels of welfare?

- Full optimum gives you guidance on which simpler policies can achieve highest possible welfare
  - E.g., let log-linear tax vary with age *and* average lifetime income ($\bar{y} = \frac{1}{a-1}\sum_{t=0}^{a-1} y_t$)

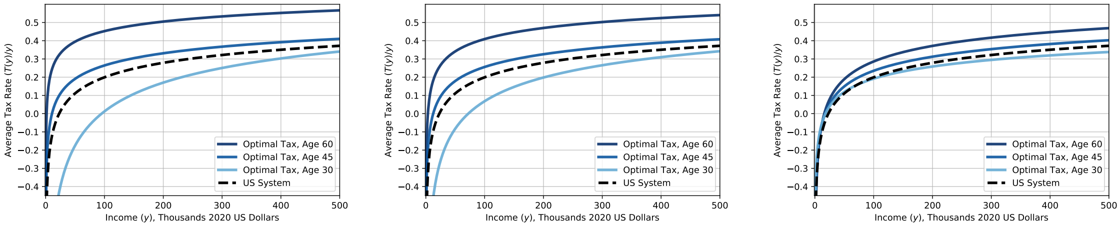
$$T(y; a, \bar{y}) = y - (1 - \tau(a, \bar{y}))y^{\rho(a,\bar{y})}$$

$$\tau(a, \bar{y}) = \tau_0 + \tau_1 a + \tau_2 a^2 + \tau_3 a\bar{y} + \tau_4 a^2 \bar{y} \text{ and } \rho(a, \bar{y}) = \rho_0 + \rho_1 a + \rho_2 a^2 + \rho_3 a\bar{y} + \rho_4 a^2 \bar{y}$$

  - Turns out to capture 90% of welfare gain of HD tax compared to just parametric AD tax
    - Only lose 0.2% of consumption compared to full non-parametric HD tax

History-Dependent Parametric Tax

Back



(a) Optimal Tax by Age, $\bar{y} = \$25,000$ (b) Optimal Tax by Age, $\bar{y} = \$100,000$ (c) Optimal Tax by Age, $\bar{y} = \$500,000$

Figure: Optimal Parametric History-Dependent Tax

- Captures key feature of full HD policy: HH's are rewarded for high past output with smaller increase
  - Low income history $\implies$ taxes increase quickly with age
  - High income history $\implies$ taxes increase more slowly with age

Marginal tax Rates

Comparing Approximation Methods

Back

- **Standard Polynomial Approximation**: *weighted sum of nonlinear transformations*

$$T(x;w) \approx \sum_{i_a=1}^{n} \cdots \sum_{i_0=1}^{n} w_{i_0,\ldots,i_a} f_{i_0,\ldots,i_a}(x), \text{ where } x = (\{y_t\}_{t=0}^{a-1}, a)$$

  - Number of parameters w to estimate $= n^A \implies$ grows exponentially

- **Neural Network**: *nonlinear transformation of weighted sums*

$$T(x;w) \approx \sum_{i=1}^{m} w_{2,i} f\left(\sum_{j=0}^{a} w_{1,j,i} x_j\right), \text{ where } x = (\{y_t\}_{t=0}^{a-1}, a)$$

  - Number of parameters w to estimate $= m + m \times (a+1) \implies$ grows linearly

- NN finds weighted sum of state variables as low dimension representation of state
  - Usually some average can accurately represent the full state vector
  - Especially if state variables are similar (like with income history)

Deep Learning ◂ Back

- With more layers, parameters still grow linearly with number of state variables (m)
- Two layers

$$y(x;w) \approx \sum_{i_2=1}^{n} w_{3,i_2} f\left(\sum_{i_1=1}^{n} w_{2,i_1,i_2} f\left(\sum_{j=1}^{m} w_{1,j,i_{i1}} x_j\right)\right)$$

$$\text{num. parameters} = n \times m + n^2 + n$$

- L layers (usually at most 5)

$$y(x;w) \approx \sum_{i_L=1}^{n} w_{L,i_L} f\left(\sum_{i_{L-1}=1}^{n} w_{L-1,i_{L-1},i_L} f\left(\cdots \sum_{i_2=1}^{n} w_{3,i_2,i_3} f\left(\sum_{i_1=1}^{n} w_{2,i_1,i_2} f\left(\sum_{j=1}^{m} w_{1,j,i_{i1}} x_j\right)\right)\cdots\right)\right)$$

$$\text{num. parameters} = n \times m + (L-1)n^2 + n$$

◂ go back

Solution Method for Optimal Taxes ◂ Back

- Tax function, individual choices and prices are approximated as separate NN's

$$T\left(y_a; a, \{y_t\}_{t=0}^{a-1} \mid w_T\right), \ c(b, z, \varepsilon, s, \phi, a, \{y_t\}_{t=0}^{a-1} \mid w_c; w_p, w_T) \text{ and } p(s \mid w_p; w_c, w_T)$$

- Consider a perturbation to one weight of the tax function: $w_T + \Delta = (w_1, \ldots, w_i + \delta, \ldots, w_n)$
  - Update individual choices (and skill prices) under the perturbed tax function by gradient descent

$$\tilde{w}_c = w_c + \frac{\partial U(w_c; w_T + \Delta, \tilde{w}_p)}{\partial w_c}$$

  - Compute social welfare under the perturbed tax function: $W(\tilde{w}_c, \tilde{w}_p, w_T + \Delta)$

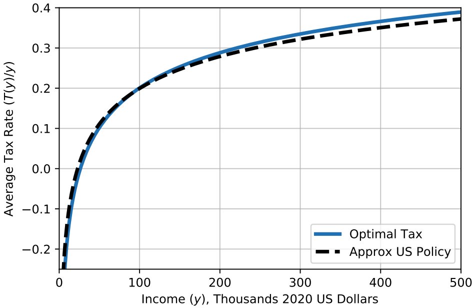
- Optimal tax system is the w_T such that a perturbation produces no welfare gain

$$\frac{\partial W(w_c, w_p, w_T + \Delta)}{\partial \delta} = 0$$

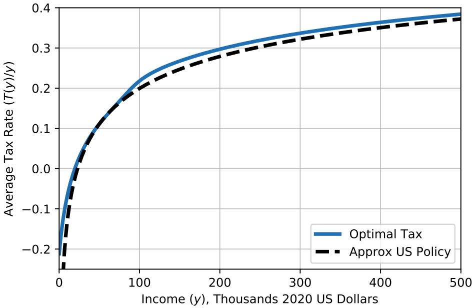
- All optimization done by computer: no taking foc's by hand

Optimal Tax on Current Income

Back



(a) Optimal Parametric Tax, $T(y) = y - (1-\tau)y^{\rho}$



(b) Optimal Nonparametric Tax, $T(y)$

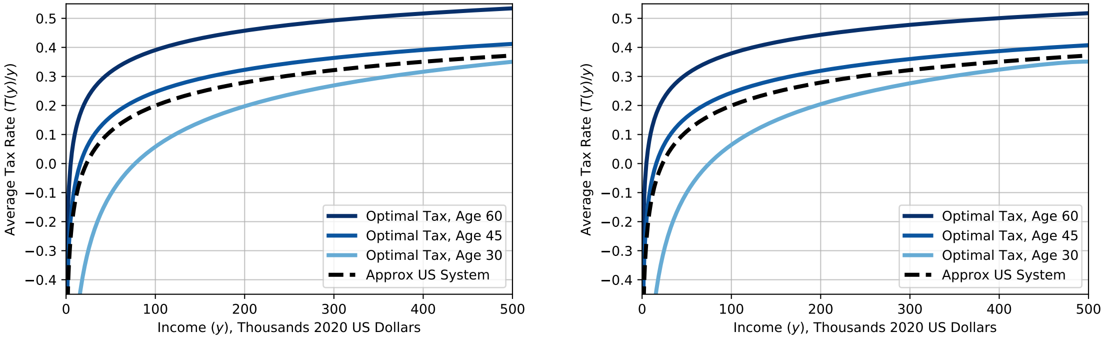
Figure: Optimal Tax on Current Income

- Optimal taxes have similar progressivity to US system
- Nonparametric tax has lower marginal tax rates for low and middle incomes ($< \$100$K)

Marginal Tax Rates

Optimal Age-Dependent Tax

Back



(a) Optimal Parametric Tax, $T(y,a) = y - (1-\tau(a))y^{\rho(a)}$

(b) Optimal Nonparametric Tax, $T(y,a)$

Figure: Optimal Age-Dependent Tax on Current Income

- Optimal taxes increase and become less progressive with age
- Mostly similar except nonparametric has low taxes on young with very high income

Marginal Tax Rates | Present Values

Current Income Marginal Tax Rates

Back

	Current Income, Thousands of Dollars					
	10	25	50	100	200	500
US System	5%	17%	25%	32%	39%	47%
Parametric	4%	22%	25%	34%	41%	49%
Nonparametric	3%	15%	24%	40%	40%	47%

Table: Marginal Tax Rates, Taxes on Current Income

Age-Dependent Marginal Tax Rates

Back

	Current Income, Thousands of Dollars					
Age	10	25	50	100	200	500
30	-21%	1%	15%	28%	38%	50%
45	10%	22%	29%	36%	43%	50%
60	26%	36%	43%	49%	55%	61%

Table: Marginal Tax Rates, Parametric Age-Dependent Tax

	Current Income, Thousands of Dollars					
Age	10	25	50	100	200	500
30	-20%	0%	16%	28%	39%	19%
45	10%	21%	29%	36%	42%	50%
60	26%	35%	42%	48%	53%	59%

Table: Marginal Tax Rates, Nonparametric Age-Dependent Tax

Allocations with HD Policy

Back

- Compare steady state allocations under HD tax to allocations under AD tax
- Average gain of switching from AD tax to HD tax by present value of income:

$$PV_i(y) = \left(\frac{1-R^{-1}}{1-R^{-A}}\right)\sum_{a=0}^{A-1} R^{-a} y_{ia}$$

	Quartile, Present Value of Income				
	0-25%	25-50%	50-75%	75-100%	Total
Income ($PV(y)$)	-0.82%	-0.56%	-0.17%	5.20%	2.12%
Consumption ($PV(c)$)	0.35%	-0.10%	-0.23%	7.08%	2.85%
Leisure ($PV(1-h)$)	0.38%	0.40%	0.38%	-1.15%	0.01%
Skills (s)	-2.26%	-0.66%	0.65%	1.64%	0.51%
Skill Price ($p(s)$)	-0.39%	0.19%	0.75%	1.72%	0.70%

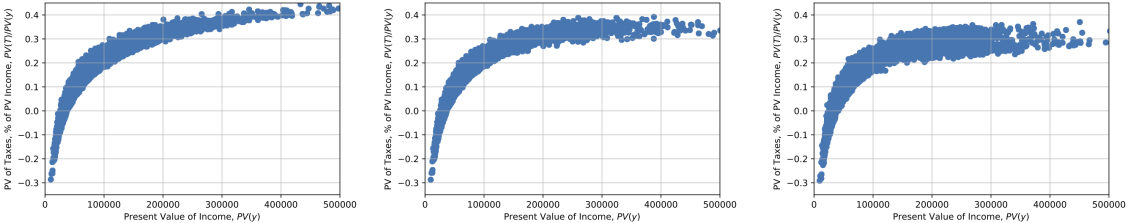
Table: Percent Gain in Average Allocations by Quartile of Present Value of Income, AD to HD Tax

- Note: $\phi = 0.275$, so leisure is valued about $3.6\times$ consumption
- Higher $Y \implies$ higher $p(s) = (Y/[N(s)f_s(s)])^{\frac{1}{\omega}} \implies$ similar c with lower h and s

Levels

Present Value of Taxes Paid

Back



(a) Parametric Age-Dependent Taxes (b) Nonparametric Age-Dependent Taxes (c) History-Dependent Taxes

Figure: Present Value of Taxes Paid by Present Value of Income

- Compare present value of taxes paid in initial period: $PV_i(T) = \left(\frac{1-R^{-1}}{1-R^{-A}}\right)\sum_{a=0}^{A-1} R^{-a} T_a(y_{ia})$
  - PV of taxes paid are similar in all three for incomes below \$100K
  - Nonparametric taxes flatter for incomes over \$200K
  - HD taxes more dispersed for incomes over \$100K

Allocations with HD Policy

Back

	Present Value of Income Quartile				
	0-25%	25-50%	50-75%	75-100%	Total
Income $(PV(y))$	37,991	58,902	82,549	149,060	82,124
Consumption $(PV(c))$	36,466	51,568	67,279	106,729	65,510
Taxes $(PV(T))$	1,525	7,333	15,267	42,315	16,610
Avg. Tax Rates $(PV(T/y))$	-4.4%	3.9%	9.4%	17.6%	6.6%
Leisure $(PV(1-h))$	0.648	0.643	0.639	0.634	0.641
Skills (s)	0.221	0.333	0.460	0.823	0.459
Skill Price $(p(s))$	0.179	0.197	0.218	0.279	0.218

Table: Average Allocations by Quartile of PV Income Distribution, AD Tax

	Quartile, Present Value of Income				
	0-25%	25-50%	50-75%	75-100%	Total
Income $(PV(y))$	37,679	58,573	82,407	156,819	83,868
Consumption $(PV(c))$	36,588	51,518	67,124	114,281	67,377
Taxes $(PV(T))$	1,164	7,117	15,288	42,588	16,539
Avg. Tax Rates $(PV(T/y))$	-5.6%	3.1%	9.0%	17.4%	6.0%
Leisure $(PV(l))$	0.650	0.646	0.642	0.626	0.641
Skills (s)	0.216	0.331	0.463	0.837	0.462
Skill Price $(p(s))$	0.178	0.198	0.219	0.284	0.220

Table: Average Allocations by Quartile of PV Income Distribution, HD Tax

Parametric History-Dependent Marginal Tax Rates

Back

Age	Current Income, Thousands of Dollars: 10	25	50	100	200	500
30	-25%	-1%	14%	27%	38%	50%
45	13%	23%	30%	36%	42%	49%
60	33%	41%	47%	52%	57%	62%

Table: Marginal Tax Rates, Parametric History-Dependent Tax, $\bar{y} = \$50,000$

Age	Current Income, Thousands of Dollars: 10	25	50	100	200	500
30	7%	16%	23%	29%	35%	42%
45	9%	20%	28%	35%	42%	49%
60	12%	25%	34%	42%	49%	57%

Table: Marginal Tax Rates, Parametric History-Dependent Tax, $\bar{y} = \$500,000$

Separable Utility

Back

- Now consider the same exercise, but with a separable utility function

$$u_i(c_{ia}, h_{ia}) = \log c_{ia} - \exp \varphi_i \frac{h_{ia}^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}}, \ \varphi \sim N(m_\varphi, v_\varphi)$$

  - Labor supply elasticity now constant for all households and equals ν

- Set $\nu = 0.5$ (standard value), re-calibrate (m_φ, v_φ) and compute welfare gains as before

	Income History	Age and Current Income	Only Current Income
Nonparametric	0.0%	1.55%	5.20%
Log-linear		1.65%	5.33%

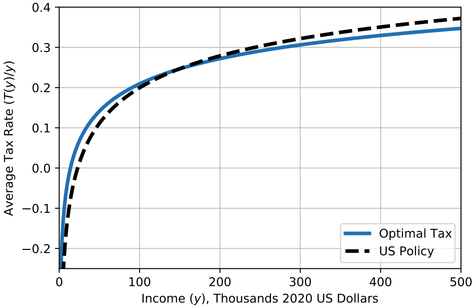
Table: Welfare Gain of Moving to Nonparametric, History Dependent Taxes

Optimal Tax Functions

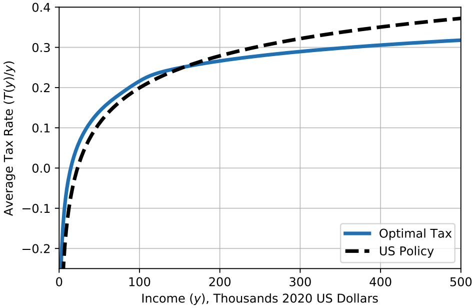
Marginal Tax Rates

Separable Utility: Tax on Current Income

Back



(a) Optimal Parametric Tax, $T(y) = y - (1-\tau)y^{\rho}$



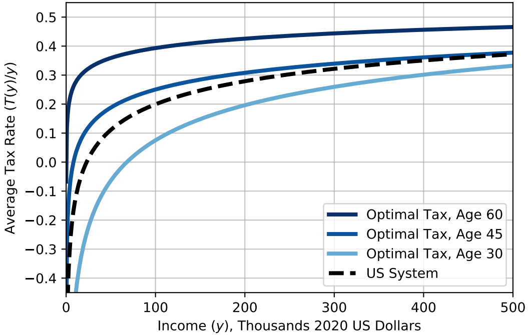
(b) Optimal Nonparametric Tax, $T(y)$

Figure: Optimal Tax on Current Income, Separable Utility

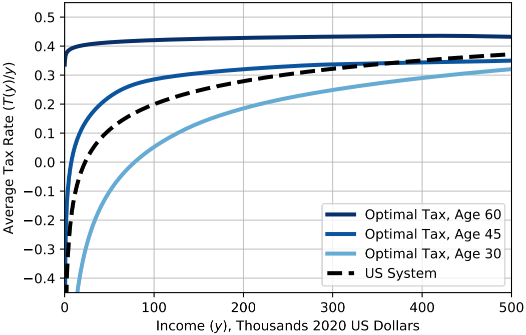
- Dotted line is parametric approximation of US income tax system as $T_{US}(y) = y - (1-\tau_{US})y^{\rho_{US}}$

Separable Utility: Age-Dependent Tax

Back



(a) Optimal Parametric Tax, $T(y,a) = y - (1-\tau(a))y^{\rho(a)}$

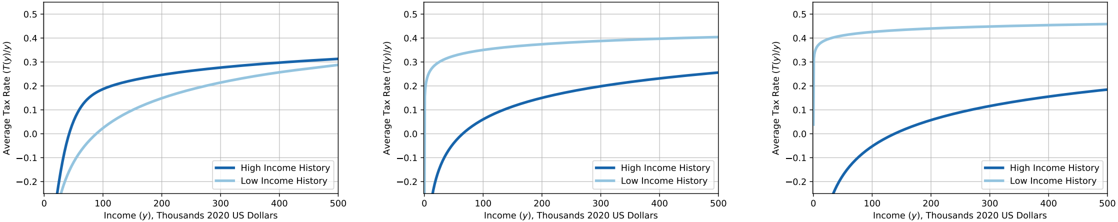


(b) Optimal Nonparametric Tax, $T(y,a)$

Figure: Optimal Age-Dependent Tax on Current Income, Separable Utility

Separable Utility: History-Dependent Tax

Back



(a) Optimal Tax After 10 Years (b) Optimal Tax After 20 Years (c) Optimal Tax After 30 Years

Figure: Optimal Non-parametric History-Dependent Tax, Nonseparable Utility

- Tax rates of two households who each earn \$50K and \$400K for first 10, 20 and 30 years of working life

Separable Utility: History-Dependent Marginal Tax Rates

Back

	Age			
	30	40	50	60
Low Income History ($50K)	-25%	30%	35%	49%
Middle Income History ($100K)	-5%	21%	38%	49%
High Income History ($400K)	29%	33%	20%	9%

Table: Average Tax Rates, History-Dependent Tax, Separable Utility

	Age			
	30	40	50	60
Low Income History ($50K)	6%	36%	39%	49%
Middle Income History ($100K)	20%	32%	41%	49%
High Income History ($400K)	37%	40%	32%	26%

Table: Marginal Tax Rates, History-Dependent Tax, Separable Utility

Summary of Tax Functions

Back

	History of Income	Age and Current Income	Only Current Income
Nonparametric	$T(y;\{y_t\}_{t=0}^{a-1},a)$	$T(y;a)$	$T(y)$
Log-linear		$y-(1-\tau(a))y^{\rho(a)}$	$y-(1-\tau)y^{\rho}$

Table: Summary of Tax Functions