

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

Ross Batzer\*

University of Minnesota

January 24, 2022

---

\*Email: [batz0025@umn.edu](mailto:batz0025@umn.edu), Website: <https://sites.google.com/view/ross-batzer>, Most recent slides, Most recent draft

# Introduction

- This paper studies optimal labor income taxation when taxes can depend on past income
- History dependence has large potential for reducing tax distortions
  - History dependent taxes can give incentives based on income history
    - No need to balance incentives across all income levels at once
  - Extract higher income from households with high earning potential
    - Give lower taxes, but only after investing in skills and earning high amounts
- In most countries, income taxes are based mostly on current income
  - But income history is recorded by governments and used in practice
    - e.g. public pensions/social security, income averaging in US (1964-1986)
- Question 1: How should labor income taxation vary with a household's income history?
- Question 2: What are the welfare implications of conditioning taxes on previous income?

# What I Do

- Study history-dependent (HD) taxes in a model of life-cycle labor supply
  - Model ingredients
    - ① Skill investment by households, e.g. decide to go to college or not
    - ② Skill types are imperfectly substitutable in production, e.g. waiter vs. accountant
    - ③ Very general class of tax function
      - Average tax rate = *any* continuously differentiable function of entire income history
      - But taxes can only depend on income, not skills or hours worked
  - Record income history (36 years) + indiv. state variables  $\implies$  up to 41 state variables

# Background

- New feature: skill types are imperfectly substitutable in production
  - Widely held view of data: e.g. Katz and Murphy (1992), Acemoglu (2002)
    - Taxes have nontrivial effects on wages for labor
    - $\uparrow$  output for one worker  $\implies \uparrow$  marginal product (wage) of all workers
  - Creates role for history-dependence
    - Use HD to incentivize households with high earning potential to produce a lot
    - Can get large increases in wages for everyone else
- Existing literature on history-dependent taxation in dynamic models
  - Stantcheva (2017), Farhi and Werning (2013), Weinzierl (2011), and others
  - Common feature: skills are perfect substitutes in production
    - Wages received for labor are invariant to tax system (constant)
  - General finding: very small welfare benefits from HD taxes over just age-dependent

# Methodology

- Imperfectly substitutable skills make standard solution methods intractable
  - Need to compute optimal taxes directly by specifying functional form for income taxes
- Need method to solve model with 41 state variables *and* compute welfare maximizing policy
  - Neural networks (NN) and new numerical libraries (e.g. Google's Tensorflow) make this possible
  - NN's designed to approximate high-dimensional functions quickly
    - Automatically allocate parameters to approximate functions with minimal parameters
- I 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
    - Flexible algorithm: can be applied to study optimal policies in a wide variety of models

# Findings

- ① Optimal history-dependent (HD) tax function is:
  - For most people: tax rates (avg. & marginal) increase over life cycle
    - Wages are estimated to increase deterministically with age
    - Gov. increases taxes with age to smooth consumption
  - For high income history (top 5% of incomes): tax rates decrease over life cycle
    - As gov. learns which HH's have highest earning potential, rewards high earnings with lower taxes
      - Maintains high labor supply during working life
      - Incentivizes skill investment at beginning of life in anticipation of lower future taxes
    - High output  $\implies$  Higher avg. wages  $\implies$  Less redistribution/distortion under optimal tax
- ② Welfare gain from HD tax large compared to optimal age-dependent tax (2% lifetime cons.)
  - Welfare gains are cut in half when wages for each skill are taken as given
  - Virtually no gains from HD taxation when wages *and* skills are taken as given
- ③ 90% of potential welfare gains achieved with parametric tax:  $T(\text{average past income, age})$ 
  - Mimics full HD tax by making optimal taxes flatter in age with higher past income

# Related Literature

## Optimal taxation

- **Parametric (aka Quantitative Ramsey):** Heathcote, Storesletten and Violante (2017, 2020), Kapička (2020), Krueger and Ludwig (2016), Karabarbounis (2016), and others
- **History dependent (aka Dynamic Mirrlees):** Stantcheva (2017), Golosov, Troshkin and Tsyvinski (2016), Farhi and Werning (2013), Weinzierl (2011), and others
- **Static with differentiated labor:** Sachs, Tsyvinski and Werquin (2020), Scheuer and Werning (2016), Rothschild and Scheuer (2013), Stiglitz (1982, 1987), and others
- **Contribution:** General tax function that allows for history-dependence in model with differentiated labor

## Neural networks to solve structural economic models

- Fernández-Villaverde, Nuño, Sorg-Langhans and Vogler (2020), Azinovic, Gaegauf and Scheidegger (2019), Maliar, Maliar and Winant (2019), and others
- **Contribution:** Application to optimal policy (nested neural network solution)

# Plan

## ① Model

- Economic environment
- Optimal taxation problem

## ② Parameter Selection

## ③ Solution Method

## ④ Results

- Optimal taxes
- Welfare analysis
- Simple implementation



# Model

# Agents

- Households (measure one, indexed by  $i \in [0, 1]$ )
  - Work for  $A$  periods, age indexed by  $a = 0, \dots, A - 1$
  - 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,  $T(\cdot)$
  - 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 constant elasticity of substitution (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  $s$
- $\omega$  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

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

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  $(s, \{c_a, h_a, b_{a+1}\}_{a=0}^{A-1})$  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

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

- 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  $T(\cdot)$  to maximize  $W$  subject to its budget

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

and that households solve their problem given the tax function

- Tax function  $T(\cdot)$  is of the form

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

- $\tau_a$  is a continuously differentiable function that depends on age and income history
- Can prove this form has a unique mapping between tax rates 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  $\epsilon_{ia}$
- Assume  $\epsilon_{ia}$  is composed of permanent component  $z$  and transitory component  $\varepsilon$

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

where

$$z_{ia+1} = z_{ia} + \eta_{ia+1}, \quad \eta_{ia} \sim N(0, v_\eta)$$

$$z_{i0} \sim N(0, v_z) \text{ and } \varepsilon_{ia} \sim N(0, v_\varepsilon)$$

- 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
$v_\eta$	Variance of permanent shocks $z$	0.003
$v_\varepsilon$	Variance of transitory shocks $\varepsilon$	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)
$\psi$	Elasticity of skill investment to return	0.65	
$\omega$	Elasticity of substitution across skills	3.124	
$g$	Government spending (% of output)	0.19	
$\beta$	Discount Factor	0.98	Golosov et al. (2016)
$R$	Return on savings	1/0.98	
$m_\phi$	Mean of leisure disutility	0.275	$H = 0.33$
$v_\phi$	Variance of labor disutility 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, splines)
  - ① Additive: additional parameters improve approx. for only a single dimension
  - ② Manual: I have to choose where to place additional parameters
  - With  $A = 36$  and affine/linear functions in each dim, would need  $2^{36} \approx 70$  billion parameters
    - Approximation is infeasible, even for crude approximation
- With neural networks
  - ① Compositional: additional parameters improve approx. for all dimensions simultaneously
  - ② Automatic: self-allocates parameters to find best low-dimensional representation of full data
  - Only have to approximate several thousand of parameters ( $\approx 5,000$ ) 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 ( $\partial W / \partial \tau$ )
    - Calculate how individual choices and GE prices change under perturbed tax function ( $\partial c / \partial \tau$ )
  - ③ Update tax function to maximize welfare (gradient descent:  $\tau_{new} = \tau_{old} + \frac{\partial W}{\partial \tau}$ )
- Optimal tax function is such that no change produces a welfare gain ( $\partial W / \partial \tau = 0$ )

► Technical Version

► Check the Approximation



# Results

# Outline of Results

## ① Optimal taxes

- Baseline: Constant elasticity of substitution production [▶ Here](#)
- Counterfactual 1: Fixed prices [▶ Here](#)
- Counterfactual 2: Fixed prices and skills [▶ 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
  - Developed nested NN method to compute optimal history-dependent taxes in an OLG economy
  - Method can be applied to wide variety of problems to compute optimal policies
- What I found
  - Welfare gain from history-dependence can be large
  - Elasticity of substitution between skill types is critical to optimal policy:
    - Key elasticity to be estimated in future work in optimal taxation
- Why it was useful
  - Demonstrates new possible source of welfare gains from income taxation
  - Full HD optimal tax provides a benchmark to compare with more implementable policies
    - Here, function in just age and avg past income gets close to full HD optimum

Thank You!

# References

- 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.
- 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*.
- Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2016) “Redistribution and Social Insurance,” *American Economic Review*, 106 (2), 359–386.
- Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017) “Optimal Tax Progressivity: An Analytical Framework,” *Quarterly Journal of Economics*, 132 (4), 1693–1754.
- (2020) “Optimal Progressivity with Age-Dependent Taxation,” *Journal of Public Economics*.
- 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.
- Krueger, Dirk and Alexander Ludwig (2016) “On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium,” *Journal of Monetary Economics*, 77, 72–98.
- Maliar, Lilia, Serguei Maliar, and Pablo Winant (2019) “Will Artificial Intelligence Replace Computational Economists Any Time Soon?” *CEPR Working Paper*.
- Rothschild, Casey and Florian Scheuer (2013) “Redistributive Taxation in the Roy Model,” *Quarterly Journal of Economics*, 128 (2), 623–668.
- Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin (2020) “Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium,” *Econometrica*, 88 (2), 469–493.
- Scheuer, Florian and Ivan Werning (2016) “Mirrlees meets Diamond-Mirrlees,” *NBER Working Paper* (22076).
- Stantcheva, Stefanie (2017) “Optimal Taxation and Human Capital Policies over the Life Cycle,” *Journal of Political Economy*, 125 (6).
- Stiglitz, Joseph (1982) “Self-Selection and Pareto Efficient Taxation,” *Journal of Public Economics*, 17, 213–240.
- (1987) “Pareto Efficient and Optimal Taxation and the New New Welfare Economics,” *Handbook of Public Economics*, 2, 991–1042.
- Weinzierl, Matthew (2011) “The Surprising Power of Age-Dependent Taxes,” *Review of Economic Studies*, 78 (4), 1490–1518.

- 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

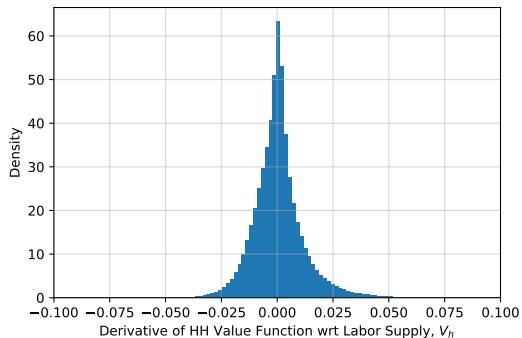
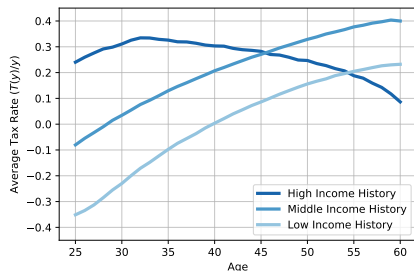
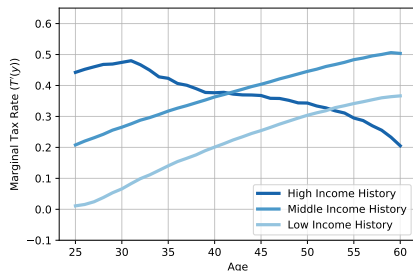
[◀ Back](#)

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



(a) Average Tax Rates ( $T_a/y_a$ )



(b) Marginal Tax Rates ( $\partial T_a / \partial y_a$ )

Figure: Optimal History-Dependent Tax Rates by Age, Baseline Model

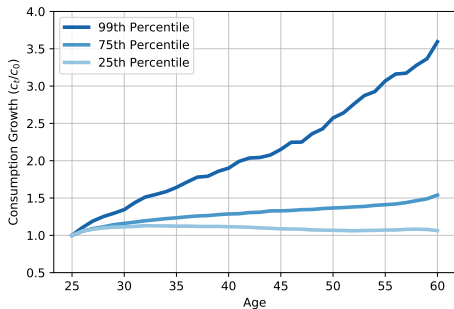
- Follow HH's that receive low (25th percentile), middle (75th) and high (99th) income each period
  - Middle income: taxes increase with deterministic life-cycle profile of wages
  - High income: taxes decrease as government learns their earning potential
    - Maintain high labor supply during life
    - Increase skill investment through expectation of lower future taxes



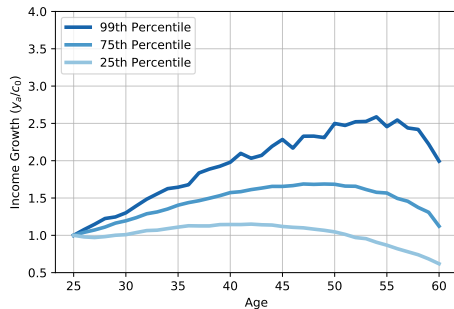
- Why does history-dependent (HD) taxation increase skill investment?
- Optimality condition for skill investment:

$$\underbrace{v'(s_i)}_{\text{Marginal Cost of Skills}} = \underbrace{\sum_{a=0}^{A-1} \beta^a \frac{\partial y_{ia}}{\partial s_i}}_{\text{After-tax Wage}} \underbrace{\left[ 1 - \overbrace{\sum_{k=a}^{A-1} \frac{\partial T_k(\{y_{ij}\}_{j=0}^k)}{\partial y_{ia}}}^{\text{Marginal Effect on Future Taxes}} \right]}_{\text{Marginal Benefit of Skills}} \frac{1}{c_{ia}}$$

- HH's w/ high earning potential anticipate low taxes after many years of high earnings
  - High effective after-tax wage early in life via lower future taxes: low  $\partial T_k / \partial y_{ia}$
- But households can't borrow, so high after-tax wage  $\not\Rightarrow$  high consumption,  $c_{ia}$
- High after-tax wage + low consumption = high skill investment
- Result: high skill investment  $\Rightarrow$  high output  $Y \Rightarrow$  high  $p(s) = \left[ \frac{Y}{N(s)f_s(s)} \right]^{\frac{1}{\omega}}$  for all  $s$



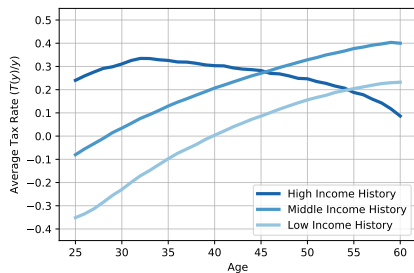
(a) Percentiles of Consumption by Age



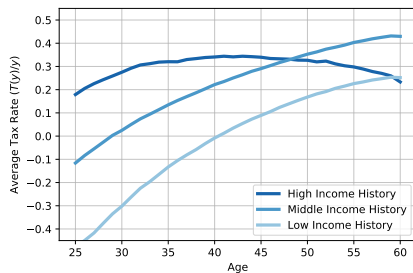
(b) Percentiles of Income by Age

Figure: Life-cycle Percentiles of Consumption and Income

- Consumption and income growth for 25th, 75th and 99th percentile of consumption and income
- Consumption is smooth for middle income HH's, but strongly increases for highest incomes



(a) Baseline Model

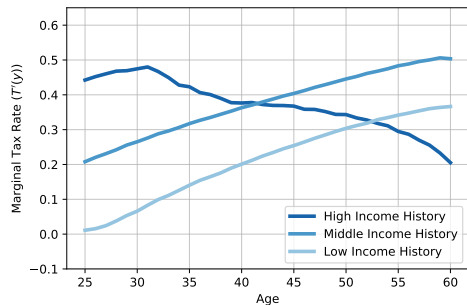


(b) Fixed Wages

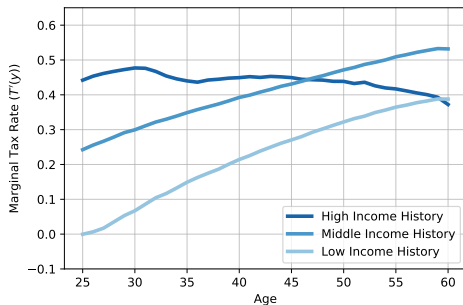
Figure: Optimal History-Dependent Average Tax Rates ( $T_a/y_a$ ) by Age

- Fix prices (under US tax system): endogenous skill investment, but fixed wage for each skill
- Taxes now about flat for high income history
  - History-dependence used less than before: less benefit from maintaining high output

## CF 1: Optimal HD Tax with Fixed Wages [◀ Back](#)



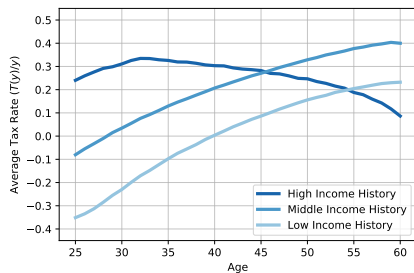
(a) Baseline Model



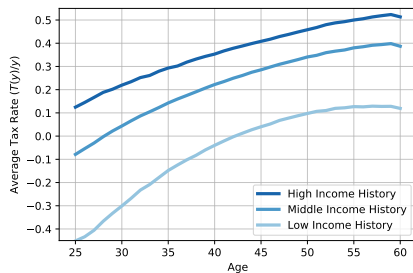
(b) Fixed Wages

Figure: Optimal History-Dependent Marginal Tax Rates ( $\partial T_a / \partial y_a$ ), by Age

- Marginal taxes also flatter than with imperfect substitutability



(a) Baseline Model

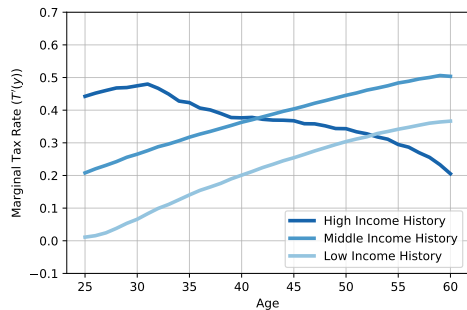


(b) Fixed Wages and Skills

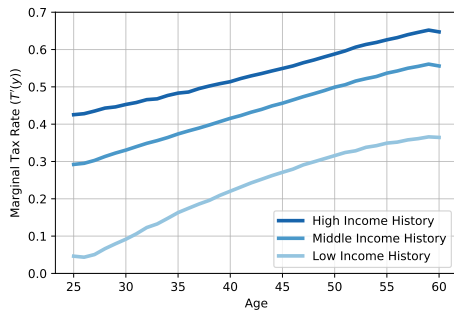
Figure: Optimal History-Dependent Average Tax Rates ( $T_a/y_a$ ) by Age

- Fix prices *and* skills (under US tax system)
- Taxes for high income history increase like others: virtually no benefit from increasing output
  - History-dependent optimum achievable with age-dependent tax

## CF 2: Optimal HD Tax with Fixed Wages and Skills [◀ Back](#)



(a) Baseline Model



(b) Fixed Wages and Skills

Figure: Optimal History-Dependent Marginal Tax Rates ( $\partial T_a / \partial y_a$ ), by Age

- Marginal taxes also increasing in age for all income histories (not much need for HD taxes)

- 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

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

- $g$  : 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  $\tau$  can be *any* differentiable function



- Gains of moving *to* most complex policy *from* more restricted tax systems:

	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

- Gain of HD tax over AD tax equivalent to 1.82 percent increase in lifetime consumption
- About 10-20 times bigger than existing studies of HD taxation ( $< 0.2\%$ )

- Gains of moving *to* optimal HD tax system *from* more restricted tax systems:

Economy	Age-Dependent	Current Income Only
Endogenous Skills and Wages (Baseline)	1.82%	5.43%
Endogenous Skills, Exogenous Wages	0.81%	4.36%
Exogenous Skills and Wages	0.10%	4.20%

**Table:** Welfare Gain of Moving to History-Dependent Tax System from Optimal Restricted Tax

- Removing endogenous effects on wages eliminates most gains from HD over age-dependent
- Removing effects on *both* wages and skill investment virtually eliminates all gains from HD
  - Consistent with existing studies of HD taxation

# What does this mean for policy?

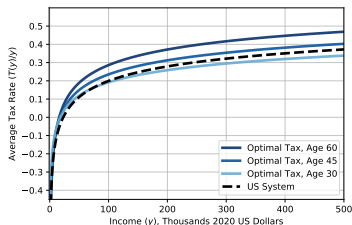
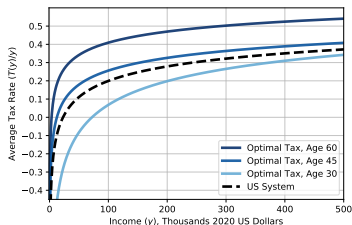
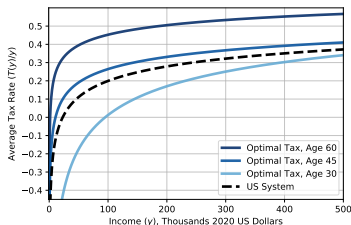
◀ Back

- 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



(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

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

$$y(x; w) \approx \sum_{i=1}^m w_{2,i} f \left( \sum_{j=0}^n w_{1,j,i} x_j \right)$$

- $y$  is the function being approximated
- $x$  is the vector of  $n$  state variables
- $m$  is the degree of approximation (determines accuracy)
- $f$  is a nonlinear function, e.g.  $\tanh$
- Number of parameters  $w$  to estimate  $= m + m \times (n + 1) \implies$  grows linearly in size of state  $x$
- NN finds weighted sums of state variables as low dimension representation of full state
  - In many cases, some average can accurately represent key features of the full state
  - Especially useful with a large number of state variables that behave “similarly”
    - e.g. many locations, countries, sectors, wage shocks, previous income levels
- Universal Approximation Theorem: NN can approximate *any* continuous function

- More complex functions can be approximated by adding additional transformations of data (*layers*)
- With more layers, parameters still grow linearly with number of state variables ( $n$ )
- Two layers

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

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

- $L$  layers (usually at most 5)

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

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

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

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

- 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)

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

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

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

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

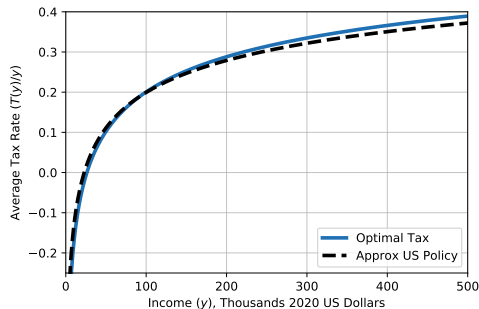
- Compute social welfare under the perturbed tax function:  $W(\tilde{w}_c, \tilde{w}_p, w_T + \Delta)$
- Optimal tax system is the  $w_T$  such that a perturbation produces no welfare gain

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

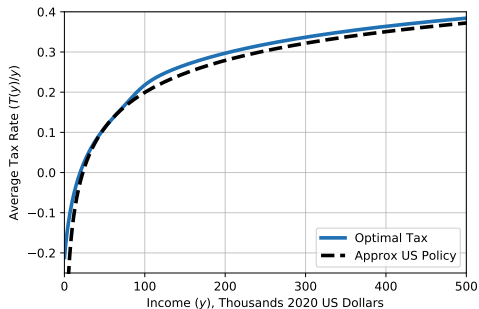
- All optimization done by computer: no taking foc's by hand



# Optimal Tax on Current Income

[◀ Back](#)

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



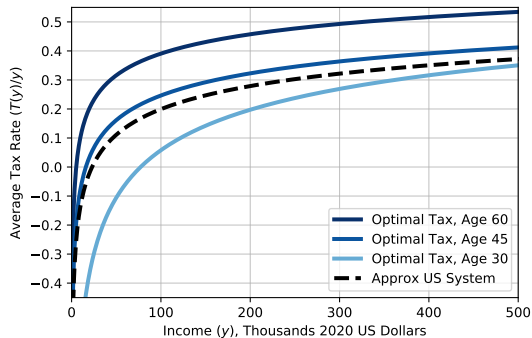
(b) Optimal Nonparametric Tax,  $T(y)$

Figure: Optimal Tax on Current Income

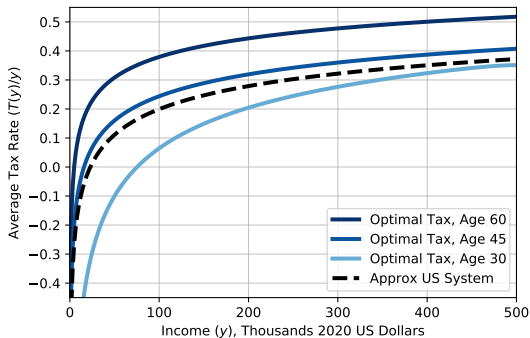
- Optimal taxes have similar progressivity to US system
- Nonparametric tax has lower marginal tax rates for low and middle incomes ( $< \$100K$ )

[▶ 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, 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	Current Income, Thousands of Dollars					
	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

Age	Current Income, Thousands of Dollars					
	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

- 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:

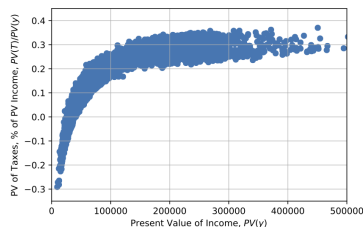
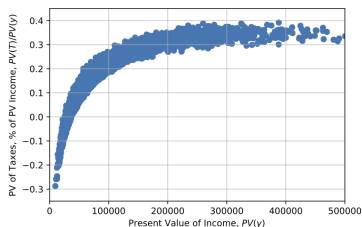
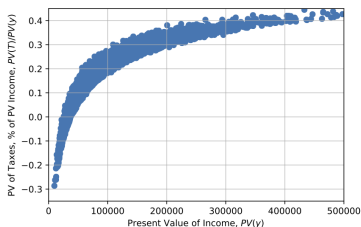
$$PV_i(y) = \left( \frac{1 - R^{-1}}{1 - R^{-A}} \right) \sum_{a=0}^{A-1} R^{-a} y_{ia}$$

	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%
Welfare ( $U$ )	1.97%	1.60%	1.18%	2.58%	1.82%

**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  $s$

# 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

	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

	Current Income, Thousands of Dollars					
Age	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,  $\bar{y} = \$50,000$

	Current Income, Thousands of Dollars					
Age	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,  $\bar{y} = \$500,000$



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

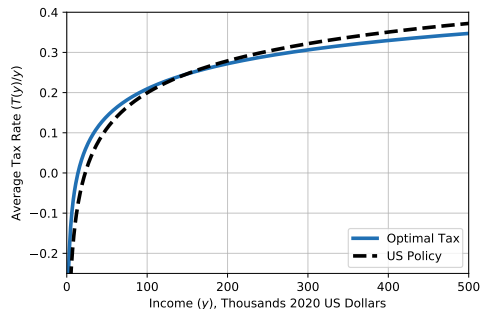
$$u_i(c_{ia}, h_{ia}) = \log c_{ia} - \exp \varphi_i \frac{h_{ia}^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}}, \quad \varphi \sim N(m_\varphi, v_\varphi)$$

- Labor supply elasticity now constant for all households and equals  $\nu$
- Set  $\nu = 0.5$  (standard value), re-calibrate  $(m_\varphi, v_\varphi)$  and compute welfare gains as before

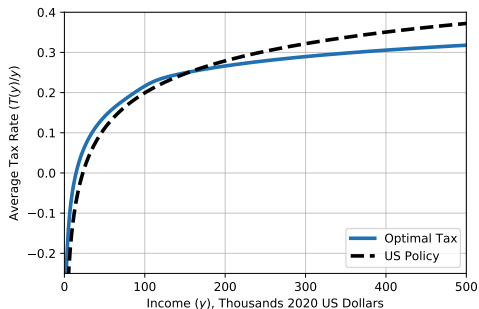
	Income History	Age and Current Income	Only Current Income
Nonparametric	0.0%	1.55%	5.20%
Log-linear		1.65%	5.33%

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

# Separable Utility: Tax on Current Income

[◀ Back](#)

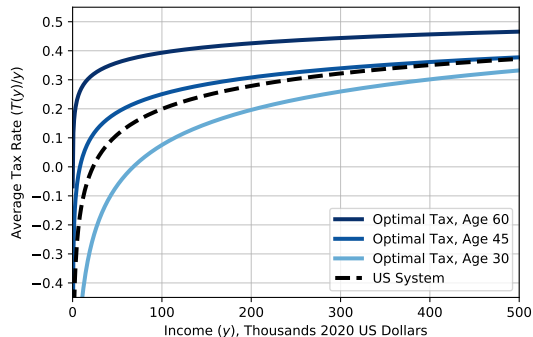
(a) Optimal Parametric Tax,  $T(y) = y - (1 - \tau)y^\rho$



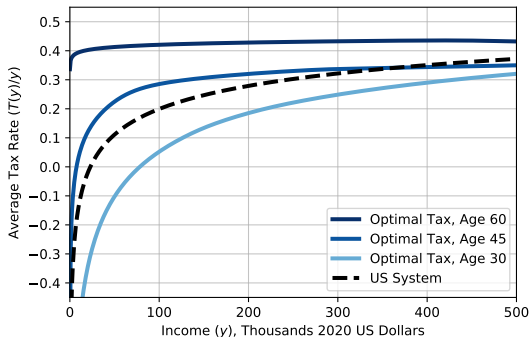
(b) Optimal Nonparametric Tax,  $T(y)$

Figure: Optimal Tax on Current Income, Separable Utility

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

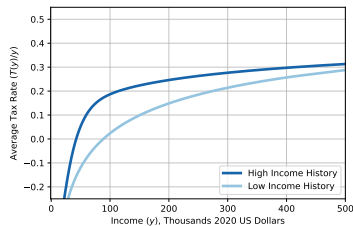


(a) Optimal Parametric Tax,  $T(y, a) = y - (1 - \tau(a))y^{\rho(a)}$

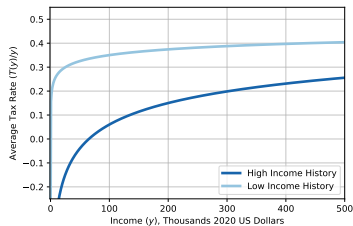


(b) Optimal Nonparametric Tax,  $T(y, a)$

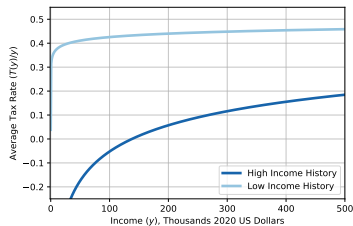
Figure: Optimal Age-Dependent Tax on Current Income, Separable Utility



(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

	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

	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