Appendix

"Optimal Fiscal Policy in a Model with Uninsurable Idiosyncratic Income Risk"

Sebastian Dyrda
University of Toronto

Marcelo Pedroni
University of Amsterdam

Contents

List of Figures

A Data

In what follows we describe our procedure to obtain macroeconomic data and cross-sectional moments at the household level. We use the Current Population Survey (CPS) to construct the cross-sectional moments for hours, the Survey of the Consumer Finances (SCF) to construct the cross-sectional moments for wealth, earnings and income, and the Consumption Expenditure Survey (CEX) to construct the cross-sectional moments for consumption. Finally, we discuss the computation of our targets for the statistical properties of the labor income process based on the data provided by Pruitt and Turner (2020) from the Internal Revenue Services.

A.1 National Income and Product Accounts (NIPA)

Following Aiyagari and McGrattan (1998) we define physical capital as the sum of nonresidential and residential private fixed assets and purchases of consumer durables. Therefore, our definition excludes government's fixed assets. We compute the average capital-output ratio, following the outlined definition, for period the 1995-2007 using two tables provided by the U.S. Bureau of Economic Analysis: Table 1.1. Current-Cost Net Stock of Fixed Assets and Consumer Durable Goods for the capital series and Table 1.1.5. Gross Domestic Product for the GDP series. We obtain the ratio of 2.49, which is very close to 2.5 used by Aiyagari and McGrattan (1998).

We define the investment-output ratio in a way consistent with the capital-output ratio, that is, we compute investment as the sum of nonresidential and residential private fixed assets and purchases of consumer durables and exclude the government's investment. We compute the average investment-output ratio, following the outlined definition, for 1995-2007 period using two tables provided by the U.S. Bureau of Economic Analysis: Table 1.5. Investment in Fixed Assets and Consumer Durable Goods for the capital series and Table 1.1.5. Gross Domestic Product for the GDP series. We obtain the ratio of 0.26.

The third statistic we discipline using data from the NIPA tables is the transfers-to-output ratio. We define transfers in the data as personal current transfer receipts, which include social security transfers, medicare, medicaid, unemployment benefits, and veteran benefits. We choose this for two reasons: First, we include retired and unemployed households in our inequality moments. Second, lump-sum transfers in the model can be interpreted as a basic income in the case of not working. We compute the transfers to output ratio, following the outlined definition, for 1995-2007 period using two tables provided by the U.S. Bureau of Economic Analysis: Table 2.1. Personal Income and Its Disposition for the construction of the transfers and Table 1.1.5. Gross Domestic Product for the GDP series. We obtain the average ratio of 0.114.

Finally, we exploit NIPA tables to discipline the capital income share in GDP, i.e. parameter α in our model. We follow closely the approach proposed by Ríos-Rull and Santaeulàlia-Llopis (2010) and described in detail in their Appendix A. In short, we define labor share of income as one minus capital income divided by output. Several sources of income, mainly proprietor's income, cannot be unambiguously allocated to labor or capital income. Thus, following Ríos-Rull and Santaeulàlia-Llopis (2010) we assume that the proportion of ambiguous capital income to ambiguous income is the same as the proportion of unambiguous capital income to unambiguous income, and we compute these series following definitions provided in Appendix A.2 of that paper. For consistency with our definition of the capital stock and investment from the paragraphs above, we use a definition of labor share with durables that extends measured output and capital income with flow services from consumer durables—see equation (39) in Appendix A of Ríos-Rull and Santaeulàlia-Llopis (2010). We compute the average labor share, following the outlined definition, for 1995-2007 and obtain the capital income share 0.378, the number that we set α to in our calibration.

A.2 Equivalence Scale

We construct cross-sectional statistics at the household level. To account for the distribution of household sizes we use an equivalence scale. This way we take into the consideration the number of people living in the household and how these people share resources and take advantage of economies of scale. We use an equivalence adjustment based on a three-parameter scale that reflects and follows the procedure of the U.S. Census Bureau:[1]

1. On average, children consume less than adults.
2. As family size increases, expenses do not increase at the same rate.
3. The increase in expenses is larger for a first child of a single-parent family than the first child of a two-adult family.

The three-parameter scale is calculated in the following way:

- One and two adults: scale = $(\text{number of adults})^{0.5}$
- Single parents: scale = $(\text{number of adults} + (0.8 \times \text{first child}) + 0.5 \times \text{other children})^{0.7}$
- All other families: scale = $(\text{number of adults} + 0.5 \times \text{number of children})^{0.7}$

We apply the same equivalence measures to the SCF and the CEX.

A.3 Current Population Survey (CPS)

Weights. We use the March supplement weights to produce our estimates. We use individual weights for individual-level variables, and household weights for household-level variables.

[1]See the link here: https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/equivalence.html

Sample selection. We use waves of the CPS from 1995 to 2007. Our CPS sample selection builds upon Heathcote, Perri, and Violante (2010). We start with the general population sample in the CPS and proceed as follows:

1. We drop households in which there are household members with negative or zero weights, no sex or no age information.
2. We drop households in which there are members with positive earnings but zero weeks worked.
3. We drop households in which there is an individual whose hourly wage is less than half the legal minimum in that year.

With that sample at hand we apply an income top-coding procedure, following Heathcote et al. (2010) and using the codes they provide. We refer the reader to their description of the top-coding procedure. Next, we define employment status and hours worked, again at *the household level*, as follows.

Hours. We use a definition of hours worked based on the CPS variable "hours worked last week" and obtain the annual number by multiplying hours by the number of weeks worked last year. Next, we aggregate hours to the household level by adding all the hours worked within the household. We define the average hours worked at the household level by dividing the total hours worked of the household over the number of the working age individuals within the household. Working age individuals are those between 25 and 60 years old, again following the definition in Heathcote et al. (2010). This procedure yields an average (over 1995-2007 period) of hours worked per person within the household of 1673 hours annually. Assuming 52 working weeks per year and 100 hours per week of available time we end up with an average hours worked as a fraction of time available of 0.32, which we use as a target in the benchmark economy. Table 1 presents the comparison of cross-sectional statistics of hours worked computed using individual hours, total household hours worked and average hours worked.

Employment status. We define the employment status of the household applying the definition used by Heathcote et al. (2010) to the household. We call the household *employed* if its total hours worked exceed 260 hours annually. As a result, households with total hours less than that are labeled as *non-employed*. Averaging over 1995-2007 period we calculate that 79 percent of households are employed.

A.4 Survey of the Consumer Finances (SCF)

We use 2007 wave of the Survey of Consumer Finances to compute the cross-sectional statistics on wealth, income and earnings, as well as the composition of total income. We also compute the self-employed statistics using SCF. The unit of observation in the SCF is a household and we use the equivalence scale described in Appendix A.2 to take into account the households' composition in the cross-section. We

Table 1: Cross-sectional statistics on hours worked.

Statistic	Hours		
	Individual	Total hhs	Average per hhs
Mean	1999.81	2630.60	1673.19
Std Dev	711.37	1981.45	1049.65
Skewness	0.10	0.62	0.79
Coeff. of Var.	0.36	0.75	0.63
% in bottom 5%	1.13	0.00	0.00
% in 1st quintile	9.35	0.01	3.02
% in 2nd quintile	18.46	11.63	13.69
% in 3th quintile	20.80	18.53	20.68
% in 4th quintile	21.84	28.48	25.44
% in 5th quintile	29.55	41.34	37.17
% in top 5%	9.05	13.71	12.91
Gini index	0.19	0.42	0.34

Note: Data come from the 1995-2007 Current Population Survey.

follow closely Kuhn and Ríos-Rull (2016) and define wealth, income and earnings and self-employed as follows:

Wealth. Our measure of a household's *wealth* is its net worth. We use the definition net worth provided by the code book of the SCF. The detailed categories incorporated in the net worth variable are provided here.

Income. The notion of *income* we use adds labor income, business income, capital income (see above for the variables included in these three categories), withdrawals from the pension accounts, SSA benefits and other pension benefits, and government transfers. The government transfers include UI/workers compensation, child support/alimony payments, income from TANF/SSI/Foodstamps and other income on income tax return.

Earnings. Our definition of *earnings* includes wages and salaries (labor income) and a fraction of business income (income from a sole proprietorship, a farm, other businesses or investments, net rent, trusts, or royalties). We set this fraction equal the sample average ratio of unambiguous labor income (wages plus salaries) to the sum of unambiguous labor income and unambiguous capital income. For the *capital income* we use the sum of income from non-taxable investments such as municipal bonds, income from other interest, income from dividends and income from gains or losses from mutual funds

Table 2: The impact of the equivalence scale on the cross-sectional statistics

Statistic	**Households**			**Households using Eq. Scale**		
	Wealth	**Income**	**Earnings**	**Wealth**	**Income**	**Earnings**
Mean	1.4E+07	1070199.4	600075.7	8987178.3	661119.4	366418.8
Std Dev	6.5E+07	5783490.8	3998315.8	4.2E+07	4099064.9	2983323.7
Skewness	10.49	14.81	22.01	11.23	24.57	36.55
Coeff. of Var.	4.56	5.40	6.66	4.69	6.20	8.14
% in bottom 5%	-0.24	0.25	-0.14	-0.23	0.25	-0.16
% in 1st quintile	-0.19	2.79	-0.14	-0.19	3.04	-0.16
% in 2nd quintile	1.07	6.73	4.21	0.99	6.82	4.12
% in 3th quintile	4.45	11.28	11.65	4.21	11.54	11.58
% in 4th quintile	11.25	18.32	20.81	11.21	18.00	20.87
% in 5th quintile	83.42	60.88	63.47	83.78	60.59	63.58
% in top 5%	33.59	20.98	18.69	33.49	20.84	18.66
Gini index	0.82	0.58	0.64	0.82	0.57	0.64

Note: Data come from the 2007 Survey of the Consumer Finances.

or from the sale of stocks, bonds, or real estate (capital gains).

Self-employed. We define *self-employed households* as those who own a business and have active management role in it. We compute their fraction in population, fraction of wealth and income held by them.

In Table 2 we present the impact of the equivalence scale on the cross-sectional moments we use to discipline our model.

A.5 Consumer Expenditure Survey (CEX)

To compute the Lorenz curve for consumption we rely on CEX data and follow closely, in terms of definition of consumption and sample selection, the work of Heathcote et al. (2010). Our sample selection procedure, top coding, and weighting follows exactly the procedure described in Appendix C of Heathcote et al. (2010). The only change we make relative to their procedure is the equivalence scale. Instead of the OECD equivalence scale we use the one defined in the Appendix A.2 for the comparability with the CPS and SCF data that we use in the paper. Table 3 presents a comparison between the cross-sectional statistics computed by Heathcote et al. (2010) and the same statistics computed using our equivalence scale.

Table 3: The cross-sectional statistics from CEX

Statistic	Heathcote, Perri, Violante (2010)	Equivalence Scale from Appendix A.2
Mean	1392.44	1586.78
Std Dev	1005.00	1126.39
Skewness	5.59	5.58
Coeff. of Var.	0.72	0.71
% in bottom 5%	1.30	1.34
% in 1st quintile	7.94	8.11
% in 2nd quintile	12.77	12.94
% in 3th quintile	16.93	16.99
% in 4th quintile	22.39	22.31
% in 5th quintile	39.97	39.66
% in top 5%	15.73	15.59
Gini index	0.32	0.31

Note: Data come from the 2006 Consumer Expenditure Survey.

A.6 Internal Revenue Services data (IRS)

Pruitt and Turner (2020) document statistical properties of the labor income process for households using administrative data from the IRS. In particular, they report quantiles 10, 25, 50, 75, and 90 of the growth rate of labor income for each percentile of income in the base year and for every year from 2000 to 2013. First, for each percentile and each year we calculate Kelly skewness and Moors kurtosis from the provided quantiles:

$$\text{Kelly Skewness}_{p,t} = \frac{(Q_{90}^{p,t} - Q_{50}^{p,t}) - (Q_{50}^{p,t} - Q_{10}^{p,t})}{Q_{90}^{p,t} - Q_{10}^{p,t}},$$

$$\text{Moors Kurtosis}_{p,t} = \frac{(Q_{87.5}^{p,t} - Q_{62.5}^{p,t}) - (Q_{37.5}^{p,t} - Q_{12.5}^{p,t})}{Q_{75}^{p,t} - Q_{25}^{p,t}}.$$

Since some of the quantiles necessary to calculate Moors kurtosis are not reported, we interpolate them from the reported ones. The targets we use in the paper are simple averages of these measures across percentiles and years.

It would be more directly useful for our purposes if the quantiles were reported in aggregate terms, not by percentile of income. Unfortunately, these are not available. To calculate comparable measures in the model we generate Monte-Carlo simulated data and then use exactly the same procedure used to obtain the targets in the data. That is, in the model, we also first split the households by income percentile, calculate the quantiles of income growth for each percentile, then calculate the measures above before averaging across percentiles.

B Two Period Economies

Let the utility function be given by

$$u(c,h) = \frac{(c^{\gamma}(1-h)^{1-\gamma})^{1-\sigma}}{1-\sigma}, \tag{B.1}$$

let $F(K,N)$ denote the production function including undepreciated capital.

B.1 Risk Economy

We can define equilibrium as follows:

Definition 1 *A tax distorted competitive equilibrium is* $(K, h_L, h_H, w, R, \tau^h, \tau_R^k, T)$ *such that*

1. (K, h_L, h_H) *solves*

$$\max_{a, h_L, h_H} u(\omega - a, \bar{h}) + \beta E[u(c_i, h_i)], \quad s.t.\ c_i = (1-\tau^h) w e_i h_i + (1-\tau_R^k) R a + T;$$

2. $R = F_K(K,N)$, $w = F_N(K,N)$, *where* $N = \pi_L e_L h_L + \pi_H e_H h_H$;
3. *and,* $\tau^h w N + \tau_R^k R K = G + T$.

For this economy we can establish the following proposition:

Proposition 1 *The optimal tax system is such that*

$$\tau^h = \frac{\Omega}{1 - N + \gamma\Omega}, \quad and \quad \tau_R^k = \frac{(1-\gamma)\tau^h}{1-\gamma\tau^h}, \quad where \quad \Omega \equiv \frac{\pi_L(1-e_L)u_{c,L} + \pi_H(1-e_H)u_{c,H}}{\pi_L u_{c,L} + \pi_H u_{c,H}} \geq 0.$$

Proof. To simplify notation, we define

$$u_{c,0} \equiv u_c(\omega - K, \bar{h}), \quad u_{c,L} \equiv u_c(c_L, h_L), \quad \text{and} \quad u_{c,H} \equiv u_c(c_H, h_H),$$

with analogous definitions for derivatives with respect h and those of higher order. Also, define after-tax prices $\tilde{w} \equiv (1-\tau^h)w$ and $\tilde{R} \equiv (1-\tau_R^k)R$, then the equations that characterize the equilibrium can be written as

$$\begin{aligned}
&\tilde{R} = \frac{u_{c,0}}{\beta(\pi_L u_{c,L} + \pi_H u_{c,H})}, \qquad -\tilde{w} = \frac{u_{h,L}}{e_L u_{c,L}} = \frac{u_{h,H}}{e_H u_{c,H}},\\
&c_L = \tilde{w} e_L h_L + \tilde{R} K + T, \qquad c_H = \tilde{w} e_H h_H + \tilde{R} K + T,\\
&T = \pi_L c_L + \pi_H c_H - (\tilde{w} N + \tilde{R} K), \qquad N = \pi_L e_L h_L + \pi_H e_H h_H.
\end{aligned}$$

Replacing the government budget constraint by the resource constraint, and using the utility function to simplify the intratemporal conditions and budget constraints, we can write the dual version of the Ramsey planner's problem as that of solving

$$\max_{\tilde{w},\tilde{R},T,K} u(\omega - K, \bar{h}) + \beta(\pi u(c_L, h_L) + (1-\pi)u(c_H, h_H)),$$

subject to

$$\begin{aligned} u_c(\omega - K, \bar{h}) &= \tilde{R}\beta(\pi_L u_c(c_L, h_L) + \pi_H u_c(c_H, h_H)), \\ G + T &= F(K, N) - (\tilde{w}N + \tilde{R}K), \end{aligned}$$

where

$$\begin{aligned} c_L &= \gamma(\tilde{w}e_L + \tilde{R}K + T), \qquad & c_H &= \gamma(\tilde{w}e_H + \tilde{R}K + T), \\ h_L &= \gamma - (1-\gamma)\frac{\tilde{R}K + T}{\tilde{w}e_L}, \qquad & h_L &= \gamma - (1-\gamma)\frac{\tilde{R}K + T}{\tilde{w}e_H}, \\ N &= \pi_L e_L h_L + \pi_H e_H h_H. \end{aligned}$$

Letting μ be the Lagrange multiplier on the intertemporal condition, and λ the one on the resource constraint, we obtain the first-order conditions of the planner's problem

$$[\tilde{w}] : \beta \left\{ \begin{array}{r} \pi_L \left[(u_{c,L} - \mu\tilde{R}u_{cc,L})\gamma e_L + (u_{h,L} - \mu\tilde{R}u_{ch,L})\frac{(1-\gamma)}{\tilde{w}e_L}\frac{\tilde{R}K+T}{\tilde{w}}\right] \\ +\pi_H \left[(u_{c,H} - \mu\tilde{R}u_{cc,H})\gamma e_H + (u_{h,H} - \mu\tilde{R}u_{ch,H})\frac{(1-\gamma)}{\tilde{w}e_H}\frac{\tilde{R}K+T}{\tilde{w}}\right] \end{array} \right\} - \lambda\left((F_N - \tilde{w})\frac{(1-\gamma)}{\tilde{w}}\frac{\tilde{R}K+T}{\tilde{w}} - N\right) = 0,$$

$$[\tilde{R}] : \beta \left\{ \begin{array}{r} \pi_L \left[(u_{c,L} - \mu\tilde{R}u_{cc,L})\gamma - (u_{h,L} - \mu\tilde{R}u_{ch,L})\frac{(1-\gamma)}{\tilde{w}e_L}\right] \\ +\pi_H \left[(u_{c,H} - \mu\tilde{R}u_{cc,H})\gamma - (u_{h,H} - \mu\tilde{R}u_{ch,H})\frac{(1-\gamma)}{\tilde{w}e_H}\right] \end{array} \right\} - \mu\beta\left(\frac{\pi_L u_{c,L} + \pi_H u_{c,H}}{K}\right) + \lambda\left((F_N - \tilde{w})\frac{(1-\gamma)}{\tilde{w}} + 1\right) = 0,$$

$$[T] : \beta \left\{ \begin{array}{r} \pi_L \left[(u_{c,L} - \mu\tilde{R}u_{cc,L})\gamma - (u_{h,L} - \mu\tilde{R}u_{ch,L})\frac{(1-\gamma)}{\tilde{w}e_L}\right] \\ +\pi_H \left[(u_{c,H} - \mu\tilde{R}u_{cc,H})\gamma - (u_{h,H} - \mu\tilde{R}u_{ch,H})\frac{(1-\gamma)}{\tilde{w}e_H}\right] \end{array} \right\} + \lambda\left((F_N - \tilde{w})\frac{(1-\gamma)}{\tilde{w}} + 1\right) = 0,$$

$$[K] : \beta \left\{ \begin{array}{r} \pi_L \left[(u_{c,L} - \mu\tilde{R}u_{cc,L})\gamma - (u_{h,L} - \mu\tilde{R}u_{ch,L})\frac{(1-\gamma)}{\tilde{w}e_L}\right] \\ +\pi_H \left[(u_{c,H} - \mu\tilde{R}u_{cc,H})\gamma - (u_{h,H} - \mu\tilde{R}u_{ch,H})\frac{(1-\gamma)}{\tilde{w}e_H}\right] \end{array} \right\} - \frac{u_{c,0} + \mu u_{cc,0}}{\tilde{R}} + \lambda\left((F_N - \tilde{w})\frac{(1-\gamma)}{\tilde{w}} + 1 - \frac{F_K}{\tilde{R}}\right) = 0.$$

We need to manipulate these equations, for clarity we keep the initial identifiers $[\tilde{w}]$, $[\tilde{R}]$, $[T]$, or $[K]$ throughout and refer to the equations by them. First notice that, subtracting $[T]$ from $[\tilde{R}]$, it follows that

$$\mu = 0,$$

so that the intertemporal Euler equation is not binding for the planner at the optimum. Using this fact and subtracting $[T]$ from $[K]$, it follows that

$$\lambda = -\frac{u_{c,0}}{F_K}.$$

Hence, it follows that we can rewrite $[\tilde{w}]$ and $[\tilde{R}]$ as

$$[\tilde{w}] : \beta \left\{ \begin{array}{c} \pi_L \left[u_{c,L}\gamma e_L + u_{h,L}\frac{(1-\gamma)}{\tilde{w}e_L}\frac{\tilde{R}K+T}{\tilde{w}} \right] \\ +\pi_H \left[u_{c,H}\gamma e_H + u_{h,H}\frac{(1-\gamma)}{\tilde{w}e_H}\frac{\tilde{R}K+T}{\tilde{w}} \right] \end{array} \right\} + \frac{u_{c,0}}{F_K}\left((F_N - \tilde{w})\frac{(1-\gamma)}{\tilde{w}}\frac{\tilde{R}K+T}{\tilde{w}} - N \right) = 0,$$

$$[\tilde{R}] : \beta \left\{ \begin{array}{c} \pi_L \left[u_{c,L}\gamma - u_{h,L}\frac{(1-\gamma)}{\tilde{w}e_L} \right] \\ +\pi_H \left[u_{c,H}\gamma - u_{h,H}\frac{(1-\gamma)}{\tilde{w}e_H} \right] \end{array} \right\} - \frac{u_{c,0}}{F_K}\left((F_N - \tilde{w})\frac{(1-\gamma)}{\tilde{w}} + 1 \right) = 0.$$

Using the agents' intratemporal and intertemporal optimality conditions, and defining

$$\Omega \equiv \frac{\pi_L(1-e_L)u_{c,L} + \pi_H(1-e_H)u_{c,H}}{\pi_L u_{c,L} + \pi_H u_{c,H}},$$

we obtain

$$[\tilde{w}] : (1-\gamma)(\tilde{R}K+T) = \gamma(1-\Omega)\tilde{w} + \frac{\tilde{R}}{F_K}((\frac{F_N}{\tilde{w}} - 1)(1-\gamma), (\tilde{R}K+T) - \tilde{w}N),$$

$$[\tilde{R}] : \frac{F_K}{\tilde{R}} = (1-\gamma)\frac{F_N}{\tilde{w}} + \gamma.$$

The agents' budget constraints aggregate to

$$C = \tilde{w}N + \tilde{R}K + T,$$

and with the normalization $\pi_L e_L + \pi_H e_H = 1$, the intratemporal conditions can be aggregated to

$$\frac{1-N}{C} = \frac{1-\gamma}{\gamma}\frac{1}{\tilde{w}}.$$

This allows us to rewrite the equations as

$$[\tilde{w}] : \frac{F_N - \tilde{w}}{\tilde{w}} = \frac{\gamma}{1-\gamma}\frac{\Omega\tilde{w}}{C - \gamma\Omega\tilde{w}},$$

$$[\tilde{R}] : \frac{F_K - \tilde{R}}{\tilde{R}} = (1-\gamma)\frac{F_N - \tilde{w}}{\tilde{w}}.$$

Finally, it follows that

$$\tau^h = \frac{\Omega}{1 - N + \gamma\Omega}, \quad \text{and} \quad \tau^k_R = \frac{(1-\gamma)\tau^h}{1-\gamma\tau^h}.$$

■

Monotonicity results Next we show that consumption and labor supply are increasing in productivity levels and that marginal utility of consumption (in the second period) is decreasing. To see this first notice that

$$c(e) = \gamma(\tilde{w}e + \tilde{R}K + T) \;\Rightarrow\; \frac{\partial c(e)}{\partial e} = \gamma\tilde{w} > 0,$$
$$h(e) = \gamma - (1-\gamma)\frac{\tilde{R}K + T}{\tilde{w}e} \;\Rightarrow\; \frac{\partial h(e)}{\partial e} = (1-\gamma)\frac{\tilde{R}K + T}{\tilde{w}e^2} > 0.$$

Next, since

$$u_c(e) = \gamma(c(e))^{\gamma(1-\sigma)-1}(1-h(e))^{(1-\gamma)(1-\sigma)}$$

we have that

$$\begin{aligned}
\frac{\partial u_c(e)}{\partial e} &= \gamma(c(e))^{\gamma(1-\sigma)-1}(1-h(e))^{(1-\gamma)(1-\sigma)}\left((\gamma(1-\sigma)-1)\frac{\partial c(e)/\partial e}{c(e)} - \gamma(1-\gamma)(1-\sigma)\frac{\partial h(e)/\partial e}{(1-h(e))}\right) \\
&= \gamma(c(e))^{\gamma(1-\sigma)-1}(1-h(e))^{(1-\gamma)(1-\sigma)}\left(\frac{(\gamma(1-\sigma)-1)\tilde{w}}{(\tilde{w}e + \tilde{R}K + T)} - \frac{\gamma(1-\gamma)(1-\sigma)(\tilde{R}K+T)}{(\tilde{w}e + \tilde{R}K + T)e}\right) \\
&= \frac{\gamma u_c(e)}{c(e)}\tilde{w}(\gamma(1-\sigma)(1-\gamma+h(e))-1) < 0,
\end{aligned}$$

where the inequality follows from the fact that

$$(\gamma(1-\sigma)(1-\gamma+h(e))-1) < (\gamma(1-\gamma+h(e))-1) = -(1-\gamma)\left(1+\gamma\frac{\tilde{R}K+T}{\tilde{w}e}\right) < 0.$$

Finally, notice that if there is no risk, then $\Omega = 0$. Otherwise, start from any pair (e_L, e_H) and increase the amount of risk by adding a mean-preserving spread of ε such that the pair of productivities becomes $(e_L - \varepsilon/\pi_L, e_H + \varepsilon/\pi_H)$, then it follows that

$$\frac{\partial\Omega(\varepsilon)}{\partial\varepsilon} = \frac{-\left[(1+\frac{\pi_H}{\pi_L})u_{c,H}\frac{\partial u_{c,L}}{\partial\varepsilon} + (1+\frac{\pi_L}{\pi_H})u_{c,L}\frac{\partial u_{c,H}}{\partial\varepsilon}\right]\varepsilon + (u_{c,L}-u_{c,H})(\pi_L u_{c,L} + \pi_H u_{c,H})}{(\pi_L u_{c,L} + \pi_H u_{c,H})^2} > 0.$$

B.2 Inequality Economy

We can define equilibrium as follows:

Definition 2 *A tax distorted competitive equilibrium is* $(a_L, a_H, K, h_L, h_H, N, w, R, \tau^h, \tau^k_R, T)$ *such that*

1. *for* $i \in \{L, H\}$, (a_i, h_i) *solves*

$$\max_{a_i, h_i} u(\omega_i - a_i, \bar{h}) + \beta u((1-\tau^h)wn_i + (1-\tau^k_R)Ra_i + T, h_i);$$

2. $R = F_K(K, N)$, *and* $w = F_N(K, N)$;

3. $K = p_L a_L + p_H a_H$, *and* $N = p_L h_L + p_H h_H$;

4. *and,* $\tau^h wN + \tau_R^k RK = G + T$.

Assuming the planner is utilitarian, we can establish the following proposition:

Proposition 2 *The optimal tax system is such that*

$$\tau_R^k = \frac{\gamma+\beta}{\beta}\frac{\Lambda}{1-K+\Lambda}, \quad \textit{and} \quad \tau^h = 0, \quad \textit{where} \quad \Lambda \equiv \frac{p_L(K-a_L)u_{c,L} + p_H(K-a_H)u_{c,H}}{p_L u_{c,L} + p_H u_{c,H}} \geq 0.$$

Proof. Using the same notation introduced for the risk economy, the equations that characterize the equilibrium can be written as

$$\begin{aligned}
\tilde{R} &= \frac{u_{c,0,L}}{\beta u_{c,L}} = \frac{u_{c,0,H}}{\beta u_{c,H}}, \qquad -\tilde{w} = \frac{u_{h,L}}{u_{c,L}} = \frac{u_{h,H}}{u_{c,H}},\\
c_L &= \tilde{w}h_L + \tilde{R}a_L + T, \qquad c_H = \tilde{w}h_H + \tilde{R}a_H + T,\\
K &= p_L a_L + p_H a_H, \qquad N = p_L h_L + p_H h_H,\\
T &= p_L c_L + p_H c_H - (\tilde{w}N + \tilde{R}K).
\end{aligned}$$

The next step, and most of what follows, is analogous to the derivations for the risk economy. Replacing the government budget constraint by the resource constraint, and using the utility function to simplify the intratemporal conditions and budget constraints, we can write the dual version of the Ramsey planner's problem as that solving

$$\max_{\tilde{w},\tilde{R},T,a_L,a_H} \; p_L(u(\omega_L - a_L, \bar{h}) + \beta u(c_L, h_L)) + p_H(u(\omega_H - a_H, \bar{h}) + \beta u(c_H, h_H))$$

subject to

$$\begin{aligned}
u_c(\omega_L - a_L, \bar{h}) &= \beta \tilde{R} u_c(c_L, h_L), \qquad u_c(\omega_H - a_H, \bar{h}) = \beta \tilde{R} u_c(c_H, h_H),\\
G + T &= F(K,N) - (\tilde{w}N + \tilde{R}K),
\end{aligned}$$

where

$$\begin{aligned}
c_L &= \gamma(\tilde{w} + \tilde{R}a_L + T), \qquad c_H = \gamma(\tilde{w} + \tilde{R}a_H + T),\\
h_L &= \gamma - (1-\gamma)\frac{\tilde{R}a_L + T}{\tilde{w}}, \qquad h_H = \gamma - (1-\gamma)\frac{\tilde{R}a_H + T}{\tilde{w}}.
\end{aligned}$$

Letting μ_L and μ_H be the Lagrange multiplier on the intertemporal conditions of each type of agent, and λ the one on the resource constraint, we obtain the first-order conditions of the planner's problem

$$[\tilde{w}] : \beta \left\{ \begin{array}{c} p_L((u_{c,L} - \mu_L \tilde{R} u_{cc,L})\gamma + (u_{h,L} - \mu_L \tilde{R} u_{ch,L})\frac{(1-\gamma)}{\tilde{w}}\frac{\tilde{R}a_L + T}{\tilde{w}}), \\ +p_H((u_{c,H} - \mu_H \tilde{R} u_{cc,H})\gamma + (u_{h,H} - \mu_H \tilde{R} u_{ch,H})\frac{(1-\gamma)}{\tilde{w}}\frac{\tilde{R}a_H + T}{\tilde{w}}) \end{array} \right\} - \lambda \left((1-\gamma)\frac{F_N - \tilde{w}}{\tilde{w}}\frac{\tilde{R}K + T}{\tilde{w}} - N \right) = 0,$$

$$[\tilde{R}]:\beta\left\{\begin{array}{c} p_L((u_{c,L}-\mu_L\tilde{R}u_{cc,L})\gamma a_L-(u_{h,L}-\mu_L\tilde{R}u_{ch,L})\frac{(1-\gamma)}{\tilde{w}}a_L) \\ +p_H((u_{c,H}-\mu_H\tilde{R}u_{cc,H})\gamma a_H-(u_{h,H}-\mu_H\tilde{R}u_{ch,H})\frac{(1-\gamma)}{\tilde{w}}a_H) \end{array}\right\}+\left\{\begin{array}{c} -(p_L\mu_L u_{c,L}+p_H\mu_H u_{c,H}) \\ +\lambda\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}+1\right)K \end{array}\right\}=0,$$

$$[T]:\beta\left\{\begin{array}{c} p_L((u_{c,L}-\mu_L\tilde{R}u_{cc,L})\gamma-(u_{h,L}-\mu_L\tilde{R}u_{ch,L})\frac{(1-\gamma)}{\tilde{w}}) \\ +p_H((u_{c,H}-\mu_H\tilde{R}u_{cc,H})\gamma-(u_{h,H}-\mu_H\tilde{R}u_{ch,H})\frac{(1-\gamma)}{\tilde{w}}) \end{array}\right\}+\lambda\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}+1\right)=0,$$

$$[a_L]:\left\{\begin{array}{c} \tilde{R}\beta((u_{c,L}-\tilde{R}\mu_L u_{cc,L})\gamma-(u_{h,L}-\tilde{R}\mu_L u_{ch,L})\frac{(1-\gamma)}{\tilde{w}}) \\ -u_{c,0,L}-\mu_L u_{cc,0,L} \end{array}\right\}+\lambda\tilde{R}\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}-\frac{F_K-\tilde{R}}{\tilde{R}}\right)=0,$$

$$[a_H]:\left\{\begin{array}{c} \tilde{R}\beta((u_{c,H}-\tilde{R}\mu_H u_{cc,H})\gamma-(u_{h,H}-\tilde{R}\mu_H u_{ch,H})\frac{(1-\gamma)}{\tilde{w}}) \\ -u_{c,0,H}-\mu_H u_{cc,0,H} \end{array}\right\}+\lambda\tilde{R}\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}-\frac{F_K-\tilde{R}}{\tilde{R}}\right)=0.$$

Using the following properties of the utility function

$$\begin{aligned} \frac{u_{cc}}{u_c} &= -(\gamma(\sigma-1)+1)\frac{1}{c} \;\Rightarrow\; (u_c-\mu\tilde{R}u_{cc})=\left(1+(\gamma(\sigma-1)+1)\tilde{R}\frac{\mu}{c}\right)u_c, \\ \frac{u_{ch}}{u_h} &= -\gamma(\sigma-1)\frac{1}{c} \;\Rightarrow\; (u_h-\mu\tilde{R}u_{ch})=-\left(1+\gamma(\sigma-1)\tilde{R}\frac{\mu}{c}\right)\tilde{w}u_c, \end{aligned}$$

the intertemporal and intratemporal conditions of the consumer's problem to substitute away $u_{c,0,L}$, $u_{c,0,H}$, $u_{h,L}$, and $u_{h,H}$, and the following definitions

$$\begin{aligned} \Phi_0 &\equiv p_L\frac{1}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_L}{\omega_L-a_L}}+p_H\frac{1}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_H}{\omega_H-a_H}}, \\ \Phi_a &\equiv p_L\frac{a_L}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_L}{\omega_L-a_L}}+p_H\frac{a_H}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_H}{\omega_H-a_H}}, \\ \Phi_c &\equiv p_L\frac{c_L}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_L}{\omega_L-a_L}}+p_H\frac{c_H}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_H}{\omega_H-a_H}}, \end{aligned}$$

we can rearrange $[T]$, $[a_L]$, and $[a_H]$ into

$$\begin{aligned} \lambda &= \frac{\beta(p_L u_{c,L}+p_H u_{c,H})}{\tilde{R}\gamma\sigma\Phi_0\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}-\frac{F_K-\tilde{R}}{\tilde{R}}\right)-\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}+1\right)}, \\ \mu_L &= \frac{\lambda c_L}{\beta u_{c,L}}\frac{\frac{F_K-\tilde{R}}{\tilde{R}}-(1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_L}{\omega_L-a_L}}, \\ \mu_H &= \frac{\lambda c_H}{\beta u_{c,H}}\frac{\frac{F_K-\tilde{R}}{\tilde{R}}-(1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_H}{\omega_H-a_H}}, \end{aligned}$$

and, then $[\tilde{w}]$ and $[\tilde{R}]$ become

$$\begin{aligned} [\tilde{w}]: &\left(\gamma-(1-\gamma)\frac{T}{\tilde{w}}\right)(p_L u_{c,L}+p_H u_{c,H})-(1-\gamma)\frac{\tilde{R}}{\tilde{w}}(p_L u_{c,L}a_L+p_H u_{c,H}a_H) \\ &+\frac{(p_L u_{c,L}+p_H u_{c,H})\left\{\begin{array}{c} -\tilde{R}\gamma\left((\gamma(\sigma-1)+1)-(\sigma-1)(1-\gamma)\frac{T}{\tilde{w}}\right)\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}-\frac{F_K-\tilde{R}}{\tilde{R}}\right)\Phi_0 \\ +\tilde{R}\gamma(\sigma-1)(1-\gamma)\frac{\tilde{R}}{\tilde{w}}\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}-\frac{F_K-\tilde{R}}{\tilde{R}}\right)\Phi_a-\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}\frac{\tilde{R}K+T}{\tilde{w}}-N\right) \end{array}\right\}}{\tilde{R}\gamma\sigma\Phi_0\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}-\frac{F_K-\tilde{R}}{\tilde{R}}\right)-\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}+1\right)}=0, \end{aligned}$$

$$[\tilde{R}]:\frac{p_L u_{c,L} a_L + p_H u_{c,H} a_H}{p_L u_{c,L} + p_H u_{c,H}} = \frac{\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}} - \frac{F_K-\tilde{R}}{\tilde{R}}\right)(\gamma\sigma\tilde{R}\Phi_a - \Phi_c) - \left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}+1\right)K}{\tilde{R}\gamma\sigma\Phi_0\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}} - \frac{F_K-\tilde{R}}{\tilde{R}}\right) - \left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}+1\right)}.$$

Next, notice that the second-period budget constraints and intratemporal conditions aggregate to

$$C = \tilde{w}N + \tilde{R}K + T, \qquad \text{, and} \qquad \frac{1-N}{C} = \frac{1-\gamma}{\gamma}\frac{1}{\tilde{w}},$$

and, therefore

$$T = \frac{\gamma}{1-\gamma}\tilde{w} - \frac{1}{1-\gamma}\tilde{w}N - \tilde{R}K,$$

which we can use to substitute away T from $[\tilde{w}]$ to get

$$[\tilde{w}]:\gamma(F_N-\tilde{w})(1-N) = -\left(\gamma\left(\frac{1}{1-\gamma}\tilde{w}(1-N) - \tilde{R}K\right)\Phi_0 - \Phi_c + \gamma\tilde{R}\Phi_a\right)\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}} - \frac{F_K-\tilde{R}}{\tilde{R}}\right)(1-\gamma)\tilde{R}.$$

Define

$$\Lambda = \frac{p_L u_{c,L}(K-a_L) + p_H u_{c,H}(K-a_H)}{p_L u_{c,L} + p_H u_{c,H}},$$

and $[\tilde{R}]$ can be rewritten as

$$[\tilde{R}]:\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}}+1\right)\Lambda = -(\gamma\sigma\tilde{R}(K-\Lambda)\Phi_0 - \gamma\sigma\tilde{R}\Phi_a + \Phi_c)\left((1-\gamma)\frac{F_N-\tilde{w}}{\tilde{w}} - \frac{F_K-\tilde{R}}{\tilde{R}}\right).$$

To proceed, notice that the second-period budget constraints combined with the equation we obtained for T implies that

$$c_L = \gamma\tilde{R}a_L + \gamma\left(\frac{1}{1-\gamma}\tilde{w}(1-N) - \tilde{R}K\right),$$
$$c_H = \gamma\tilde{R}a_H + \gamma\left(\frac{1}{1-\gamma}\tilde{w}(1-N) - \tilde{R}K\right),$$

and, therefore,

$$\Phi_c = \gamma\tilde{R}\Phi_a + \gamma\left(\frac{1}{1-\gamma}\tilde{w}(1-N) - \tilde{R}K\right)\Phi_0.$$

This equation can be used to to simplify the system of equations to

$$[\tilde{w}]:\tilde{w} = F_N,$$
$$[\tilde{R}]:\frac{F_K-\tilde{R}}{\tilde{R}} = \frac{\Lambda}{p_L\frac{\sigma(C-\gamma\tilde{R}\Lambda)+(1-\sigma)c_L}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_L}{\omega_L-a_L}} + p_H\frac{\sigma(C-\gamma\tilde{R}\Lambda)+(1-\sigma)c_H}{\gamma\sigma\tilde{R}+(\gamma(\sigma-1)+1)\frac{c_H}{\omega_H-a_H}}}.$$

First notice that, $[\tilde{w}]$ implies that there should be no taxation of labor income

$$\tau^h = 0.$$

When $\sigma = 1$, the intertemporal optimality conditions become

$$\beta \tilde{R} = \frac{c_L}{\omega_L - a_L} = \frac{c_H}{\omega_H - a_H},$$

which allows us to simplify $[\tilde{R}]$ into

$$[\tilde{R}] : \frac{F_K - \tilde{R}}{\tilde{R}} = \frac{\Lambda(\gamma + \beta)\tilde{R}}{C - \gamma \tilde{R}\Lambda},$$

and it follows that

$$\tau_R^k = \frac{(\gamma + \beta)\tilde{R}\Lambda}{C + \beta \tilde{R}\Lambda}.$$

Further, it follows from the intertemporal optimality condition that

$$\beta \tilde{R}(\bar{\omega} - K) = C,$$

where $\bar{\omega} = p_L \omega_L + p_H \omega_H$, we get

$$\tau_R^k = \frac{\gamma + \beta}{\beta} \frac{\Lambda}{\bar{\omega} - K + \Lambda}.$$

■

Monotonicity results We first show savings are increasing in the initial endowment. This follows from the intertemporal condition which we can write as

$$\left(\gamma(\tilde{w} + \tilde{R}a + T)\right)^{(\sigma-1)(2\gamma-1)+1} - \frac{1}{\beta \tilde{R}} \left(\frac{1-\gamma}{\gamma \tilde{w}} \frac{1}{1-\bar{h}} \right)^{(\sigma-1)(1-\gamma)} (\omega - a)^{(\sigma-1)\gamma+1} = 0,$$

so that implicit differentiation yields

$$\frac{\partial a(\omega)}{\partial \omega} = \frac{(\gamma\sigma + (1-\gamma))(\tilde{w} + \tilde{R}a + T)}{((\sigma-1)(2\gamma-1)+1)\tilde{R}(\omega - a) + (\gamma\sigma + (1-\gamma))(\tilde{w} + \tilde{R}a + T)} > 0.$$

Next we establish that, in the second period, consumption is increasing in assets, and labor supply and marginal utility of consumption are decreasing in assets, since

$$c(a) = \gamma(\tilde{w} + \tilde{R}a + T) \;\Rightarrow\; \frac{\partial c(a)}{\partial a} = \gamma \tilde{R} > 0,$$
$$h(a) = \gamma - (1-\gamma)\frac{\tilde{R}a + T}{\tilde{w}} \;\Rightarrow\; \frac{\partial h(a)}{\partial a} = -(1-\gamma)\frac{\tilde{R}}{\tilde{w}} < 0.$$

Therefore, since

$$u_c(a) = \gamma(c(a))^{\gamma(1-\sigma)-1}(1-h(a))^{(1-\gamma)(1-\sigma)},$$

we have that

$$u_c'(a) = \gamma(c(a))^{\gamma(1-\sigma)-1}(1-h(a))^{(1-\gamma)(1-\sigma)} \left((\gamma(1-\sigma)-1)\frac{c'(a)}{c(a)} - \gamma(1-\gamma)(1-\sigma)\frac{h'(a)}{(1-h(a))} \right)$$

$$= \gamma(c(a))^{\gamma(1-\sigma)-1}(1-h(a))^{(1-\gamma)(1-\sigma)}\left((\gamma(1-\sigma)-1)\frac{\tilde{R}}{\tilde{w}+\tilde{R}a+T}+\gamma(1-\gamma)(1-\sigma)\frac{\tilde{R}}{\tilde{w}+\tilde{R}a+T}\right)$$
$$= \frac{\gamma(c(a))^{\gamma(1-\sigma)-1}(1-h(a))^{(1-\gamma)(1-\sigma)}\tilde{R}}{\tilde{w}+\tilde{R}a+T}((\gamma(1-\sigma)-1)+\gamma(1-\gamma)(1-\sigma))$$
$$= \frac{\gamma u_c(a)\tilde{R}}{c(a)}((\gamma(1-\sigma)-1)+\gamma(1-\gamma)(1-\sigma)) < 0,$$

where the inequality follows from the fact that

$$((\gamma(1-\sigma)-1)+\gamma(1-\gamma)(1-\sigma)) = \gamma(2-\gamma)(1-\sigma)-1 < 0.$$

Finally, notice that if there is no inequality, then $\Lambda = 0$. Otherwise, start from any pair (ω_L, ω_H) and increase the amount of inequality by adding a mean-preserving spread of ε such that the pair of endowments becomes $(\omega_L - \varepsilon/p_L, \omega_H + \varepsilon/p_H)$, then it follows that

$$\frac{\partial\Lambda(\varepsilon)}{\partial\varepsilon} = \frac{\left\{\begin{array}{c} -\left[p_H u_{c,H}\frac{\partial u_{c,L}}{\partial a_L}\frac{\partial a_L}{\partial\omega_L}+p_L u_{c,L}\frac{\partial u_{c,H}}{\partial a_H}\frac{\partial a_H}{\partial\omega_H}\right](a_H-a_L) \\ +\left((u_{c,L}-(p_H u_{c,H}+p_L u_{c,L}))\frac{\partial a_L}{\partial\omega_L}-(u_{c,H}-(p_L u_{c,L}+p_H u_{c,H}))\frac{\partial a_H}{\partial\omega_H}\right)(p_L u_{c,L}+p_H u_{c,H}) \end{array}\right\}}{(p_L u_{c,L}+p_H u_{c,H})^2} > 0.$$

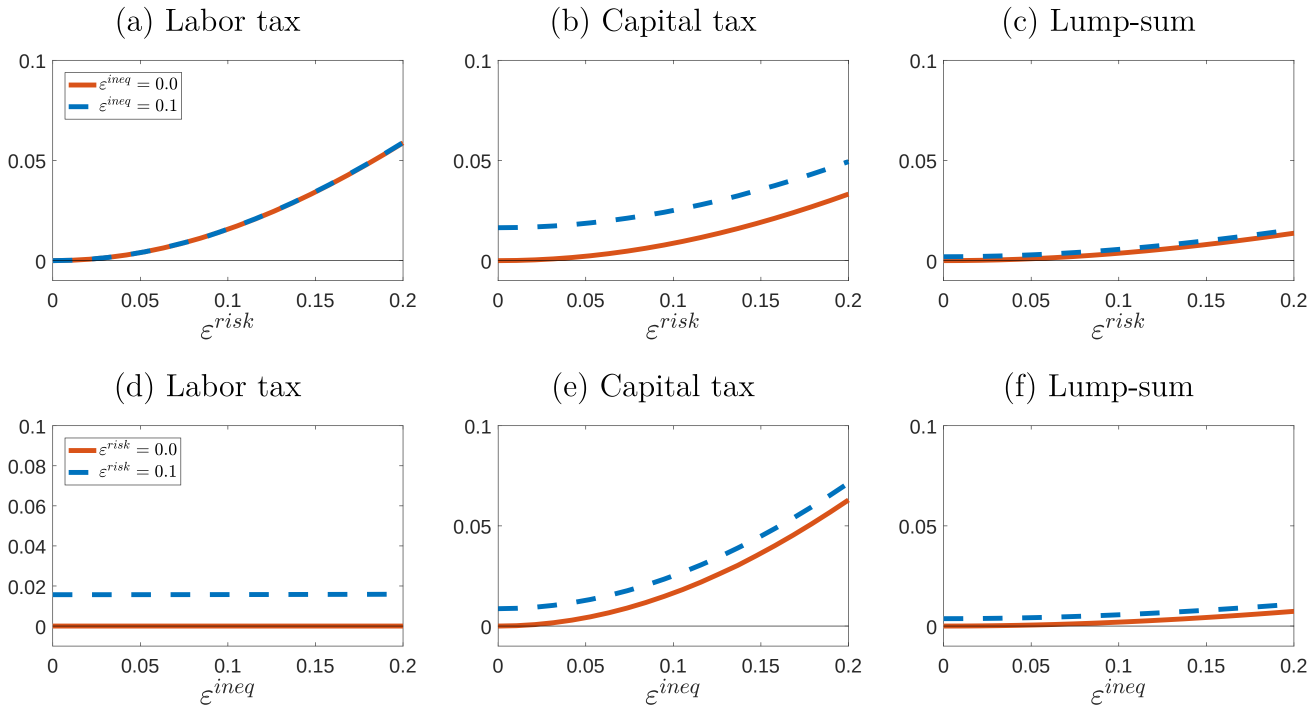
B.3 Risk and Inequality

If both risk and inequality are present, the optimal tax system has to balance three objectives: minimize distortions, and provide insurance and redistribution. A reasonable conjecture is that the optimal tax system is a convex combination of the results derived above for the risk and inequality economies, that is, positive labor and capital taxes with magnitudes associated with the levels of risk and inequality in the economy. A more subtle conjecture, associated with the result that labor should not be taxed in the inequality economy, is that, given some level of risk, the optimal labor taxes should not vary with the level of inequality. We corroborate these conjectures with a numerical example.[2] The results are plotted in Figure 1.

The first row of Figure 1 shows how the optimal tax system varies with the level of risk (controlled by the parameter ϵ^{risk}) for two levels of inequality: $\epsilon^{ineq} = 0$ (solid line) and $\epsilon^{ineq} = 0.1$ (dashed line). The solid lines corroborate the results for the risk economy. The comparison between the dashed and the solid lines corroborates the conjectures made above. The labor tax is increasing with the level of risk and independent on the level of inequality whereas capital taxes increase with the level of inequality and are independent on level of risk. The second row of Figure 1 shows the results for the analogous experiment with ϵ^{ineq} on the x-axis and $\epsilon^{risk} = 0$ (solid) and $\epsilon^{risk} = 0.1$ (dashed).

[2] The most relevant interpretation of this two-period economy is that each period corresponds to half of the working life of a person. Accordingly, we set $\beta = 0.95^{20}$ and $\delta = 1 - 0.9^{20}$. Other parameters are set to satisfy the usual targets: $\sigma = 2$, $\gamma = 0.45$, $\bar{h} = 0.33$, $\pi = p = 0.5$, and $F(K,N) = K^\alpha N^{1-\alpha} + (1-\delta)K$ with $\alpha = 0.36$. G is set to 0, but any other feasible level would just shift the lump-sum transfers correspondingly.

Figure 1: Optimal taxes in the presence of both risk and inequality.



B.4 Relationship with Dávila, Hong, Krusell, and Ríos-Rull (2012)

The results established in Dávila, Hong, Krusell, and Ríos-Rull (2012) have an interesting relationship to the ones we obtain in this paper. We use the last result to explain this relationship. Among other things, Dávila et al. (2012) show that the competitive equilibrium allocation in the SIM model is constrained inefficient. That is, the incomplete market structure itself induces outcomes that could be improved upon if consumers merely acted differently, using the same set of markets but departing from purely self-interested optimization. The constrained inefficiency results from a pecuniary externality. The savings and labor supply decisions of the agents affect the wage and interest rates and, therefore, the risk and inequality in the economy. These effects are not internalized by the agents and inefficiency follows. Note that the planner's problem in their environment is significantly different from the Ramsey problem described here. There the planner affects allocations directly and prices indirectly, as a result redistribution and insurance can only occur via the manipulation of equilibrium prices. Whereas here the Ramsey planner affects (after tax) prices directly and allocations indirectly.

In a setting similar to the inequality economy just described above, for instance, Dávila et al. (2012) show that there is under accumulation of capital. A higher level of capital would decrease interest rates and increase wages, reducing inequality. A naive extrapolation of this logic would suggest that capital income taxes should be negative so as to encourage savings. This logic, however, does not take into account the more relevant direct effect of the tax system on after tax prices. Proposition 2 shows that the opposite is true: capital income taxes should be positive.

C Budget Balancing and Debt Determination

This appendix provides the derivation of the final level of labor income tax that balances the government's budget constraint,

$$G + r_t B_t = B_{t+1} - B_t + \tau_t^c C_t + \tau_t^h w_t N_t + \tau_t^k r_t \left(K_t + B_t\right) - T_t.$$

We assume that the budget is balanced if government debt is bounded. Manipulating the equation above we obtain

$$B_{t+1} = \left(1 + \left(1 - \tau_t^k\right) r_t\right) B_t + G + T_t - \tau_t^k r_t K_t - \tau_t^h w_t N_t - \tau_t^c C_t.$$

Next, define

$$R_t \equiv 1 + \left(1 - \tau_t^k\right) r_t, \quad \text{and} \quad D_t \equiv G + T_t - \tau_t^k r_t K_t - \tau_t^h w_t N_t - \tau_t^c C_t,$$

and it follows that

$$B_{t+1} = R_t B_t + D_t \tag{C.1}$$

Iterating this equation forward we obtain

$$B_t = \left(\prod_{i=1}^{t-1} R_i\right) B_1 + \sum_{j=1}^{t-2} \left(\prod_{i=j+1}^{t-1} R_i\right) D_j + D_{t-1}$$

That is, given debt at $t = 1$, debt at t is given by the current value of B_1, plus the current value of the accumulated deficits.

Now, we compute the path $\{B_t\}_{t=1}^{t^*}$. Let $t^* + 1$ be the period in which taxes become constant (i.e. taxes are set to their final levels at $t^* + 1$). Suppose the paths $\left\{\tau_t^k, \tau_t^h, \tau_t^c, T_t\right\}_{t=1}^{t^*}$ are given. First, we calculate the debt at period $t^* + 1$ associated with these paths for taxes,

$$B_{t^*+1} = \left(\prod_{i=1}^{t^*} R_i\right) B_1 + \sum_{j=1}^{t^*-1} \left(\prod_{i=j+1}^{t^*} R_i\right) D_j + D_{t^*}.$$

To compute the debt levels for $t \in \{1, ..., t^*\}$ we can use equation ($C.1$) to solve for it backwards,

$$B_t = \frac{B_{t+1} - D_t}{R_t}.$$

Finally, we can compute the final lump-sum transfer and $\{B_t\}_{t=t^*+2}^{\bar{t}}$. Suppose the paths $\left\{\tau_t^k, \tau_t^h, \tau_t^c\right\}_{t=t^*+1}^{\bar{t}}$ are given and constant over time. We solve for T that implies $B_{\bar{t}} = B_{\bar{t}-1}$ where $\bar{t}$ is very large. We start by computing $B_{\bar{t}-1}$ taking as given B_{t^*+1} and the constant final level of taxes τ^k, τ^h, τ^c and T,

$$B_{\bar{t}-1} = \left(\prod_{i=t^*+1}^{\bar{t}-2} R_i\right) B_{t^*+1} + \sum_{j=t^*+1}^{\bar{t}-3} \left(\prod_{i=j+1}^{\bar{t}-2} R_i\right) D_j + D_{\bar{t}-2},$$

Using the definition for D_t we obtain

$$B_{\bar{t}-1} = \Psi + \Omega T. \tag{C.2}$$

where

$$\Psi \equiv \left(\prod_{i=t^*+1}^{\bar{t}-2} R_i\right) B_{t^*+1} + \sum_{j=t^*+1}^{\bar{t}-3} \left(\prod_{i=j+1}^{\bar{t}-2} R_i\right) \left(G - \tau^k r_j K_j - \tau^h w_j N_j - \tau^c C_j\right)$$
$$+ G - \tau^k r_{\bar{t}-2} K_{\bar{t}-2} - \tau^h w_{\bar{t}-2} N_{\bar{t}-2} - \tau^c C_{\bar{t}-2},$$

$$\Omega \equiv \sum_{j=t^*+1}^{\bar{t}-3} \left(\prod_{i=j+1}^{\bar{t}-2} R_i\right) + 1.$$

Substituting this and $B_{\bar{t}} = B_{\bar{t}-1}$ on (C.1) evaluated at $\bar{t}-1$ we obtain:

$$(\Psi + \Omega T) = R_{\bar{t}-1} (\Psi + \Omega T) + D_{\bar{t}-1},$$

using the definition of D_t,

$$(\Psi + \Omega T) = R_{\bar{t}-1} (\Psi + \Omega T) + G + T - \tau^k r_{\bar{t}-1} K_{\bar{t}-1} - \tau^h w_{\bar{t}-1} N_{\bar{t}-1} - \tau^c C_{\bar{t}-1},$$

and therefore

$$T = \frac{(R_{\bar{t}-1} - 1)\Psi + G - \tau^k r_{\bar{t}-1} K_{\bar{t}-1} - \tau^h w_{\bar{t}-1} N_{\bar{t}-1} - \tau^c C_{\bar{t}-1}}{(1 - R_{\bar{t}})\Omega - 1}.$$

Then, $B_{\bar{t}-1}$ is given by C.2 and we can solve for $\{B_t\}_{t=t^*+2}^{\bar{t}-2}$ backwards using (C.1).

D Algorithms

Here we describe the algorithms used to obtain our results.

D.1 Endogenous Grid Method for BGP Preferences

We develop a version of the endogenous grid method algorithm suited for the balanced growth path preferences. The dynamic programming problem is

$$v(a,e) = \max_{c,n,a'} \; u(c,h) + \beta \sum_{e'} P(e'|e)\, v(a',e'), \tag{D.1}$$

subject to

$$(1+\tau_c)\, c + a' = \tilde{w} e h + (1+\tilde{r})\, a + T, \qquad a' \geq \underline{a}, \qquad n \geq 0,$$

where $\tilde{w} = (1-\tau^n) w$ and $\tilde{r} = (1-\tau^k) r$ and we impose the balanced-growth-path (BGP) preferences,

$$u(c,h) = \frac{\left(c^{\gamma}(1-h)^{1-\gamma}\right)^{1-\sigma}}{1-\sigma}. \tag{D.2}$$

Note that the intratemporal first-order condition implies

$$h(a,e) = \max\left\{1 - \frac{(1-\gamma)}{\gamma}\frac{(1+\tau_c)\, c(a,e)}{\tilde{w} e}, 0\right\}. \tag{D.3}$$

Algorithm 1 *Then, the endogenous grid method is applied as follows:*

1. *Create a grid $\mathcal{A}$ for next period asset positions of the household and a grid $\mathcal{E}$ for household productivities.*

2. *Guess $c'(a',e')$ for each $a' \in \mathcal{A}$ and $e' \in \mathcal{E}$ and use intertemporal condition with equality to obtain $c(a',e)$ as follows.*

   (a) *Note that the intratemporal condition (D.3) implies that*

   $$h > 0 \quad \Leftrightarrow \quad \frac{(1-\gamma)}{\gamma}\frac{(1+\tau_c)}{\tilde{w} e} c^{\frac{-\sigma}{(1-\gamma)(1-\sigma)}} < c^{\frac{(1-\sigma)\gamma - 1}{(1-\gamma)(1-\sigma)}}. \tag{D.4}$$

   (b) *Rewrite the intertemporal condition using (D.4) accounting for the potentially binding lower bound on the labor supply:*

   $$\min\left\{\frac{(1-\gamma)}{\gamma}\frac{(1+\tau_c)}{\tilde{w} e} c^{-\frac{\sigma}{(1-\gamma)(1-\sigma)}}, c^{\frac{(1-\sigma)\gamma-1}{(1-\gamma)(1-\sigma)}}\right\}^{(1-\gamma)(1-\sigma)}$$

$$= \beta (1+\tilde{r}) \sum_{e'} P(e'|e) \min \left\{ \frac{(1-\gamma)}{\gamma} \frac{(1+\tau_c')}{\tilde{w}'e'} (c')^{-\frac{\sigma}{(1-\gamma)(1-\sigma)}}, (c')^{\frac{(1-\sigma)\gamma-1}{(1-\gamma)(1-\sigma)}} \right\}^{(1-\gamma)(1-\sigma)}.$$

(c) *Obtain the* $c(a', e)$ *depending whether the constraint on the labor supply binds, that is*

$$c(a', e) = \left[\left(\frac{\gamma}{(1-\gamma)} \frac{\tilde{w}e}{(1+\tau_c)} \right)^{-\frac{(1-\gamma)(1-\sigma)}{\sigma}} \right] \left[\beta (1+\tilde{r}) \sum_{e'} P(e'|e) \min \left\{ \frac{(1-\gamma)}{\gamma} \frac{(1+\tau_c')}{\tilde{w}'e'} (c')^{-\frac{\sigma}{(1-\gamma)(1-\sigma)}}, (c')^{\frac{(1-\sigma)\gamma-1}{(1-\gamma)(1-\sigma)}} \right\}^{(1-\gamma)(1-\sigma)} \right]^{-\frac{1}{\sigma}},$$

for $n > 0$*, and*

$$c(a', e) = \left[\beta (1+\tilde{r}) \sum_{e'} P(e'|e) \min \left\{ \frac{(1-\gamma)}{\gamma} \frac{(1+\tau_c')}{\tilde{w}'e'} (c')^{-\frac{\sigma}{(1-\gamma)(1-\sigma)}}, (c')^{\frac{(1-\sigma)\gamma-1}{(1-\gamma)(1-\sigma)}} \right\}^{(1-\gamma)(1-\sigma)} \right]^{\frac{1}{(1-\sigma)\gamma-1}},$$

for $n = 0$.

3. *From the budget constraint and the intratemporal condition we then obtain*

$$a(a', e) = \frac{\min \left\{ \gamma^{-1} (1+\tau_c) c + a' - \tilde{w}e - T, (1+\tau_c) c + a' - T \right\}}{(1+\tilde{r})},$$

which can be inverted to get $a'(a, e)$.

4. *Impose the borrowing constraint by setting* $a'(a, e) = \max(a'(a, e), \underline{a})$.

5. *Use the budget constraint and the intratemporal condition the update the consumption policy functio,n*

$$a'(a, e) > (1+\tilde{r}) a + T - \frac{\gamma}{(1-\gamma)} \tilde{w}e \Rightarrow c(a, e) = \frac{\gamma}{(1+\tau_c)} (\tilde{w}e + (1+\tilde{r}) a + T - a'(a, e)),$$
$$a'(a, e) < (1+\tilde{r}) a + T - \frac{\gamma}{(1-\gamma)} \tilde{w}e \Rightarrow c(a, e) = \frac{1}{(1+\tau_c)} ((1+\tilde{r}) a + T - a'(a, e)).$$

6. *Given* $c(a, e)$ *obtain* $n(a, e)$ *from the intratemporal condition (D.3).*

D.2 Transition

We compute the transition between steady states as follows[3]:

Algorithm 2 *Pick* $\bar{t} > 0$ *i.e. the transition length.*

[3] This is an extension of the procedure proposed by Domeij and Heathcote (2004).

1. *Solve for the initial stationary equilibrium using endogenous grid method to solve for the decision rules and density function iteration to solve for stationary distribution.*

2. *Assume the economy converges to a new stationary equilibrium in $\bar{t}$ periods and guess sequences of capital $\{K_i\}_{i=2}^{\bar{t}-1}$ and labor inputs $\{N_i\}_{i=2}^{\bar{t}-1}$. Back out factor prices from the firm's first order conditions.*

3. *Adjust the level of entire path of lump-sum such that given $\{K_i\}_{i=2}^{\bar{t}-1}$, $\{N_i\}_{i=2}^{\bar{t}-1}$ and the paths for the other taxes, government debt is unchanged between $\bar{t}-1$ and $\bar{t}$. Compute the associated path for the government debt, $\{B_i\}_{i=1}^{\bar{t}-1}$ (for details see Appendix C).*

4. *Solve for the final stationary equilibrium given final tax rates τ^k, τ^h, τ^c and T, and $B_{\bar{t}}$. Compute $K_{\bar{t}}$.*

5. *Solve for households savings decisions in transition.*

6. *Update the path of capital and labor input, i.e. take the initial stationary distribution over wealth and productivity and use the decision rules computed above to simulate the economy forward. Then, check for market clearing at each date and adjust $\{K_i\}_{i=2}^{\bar{t}-1}$ and $\{N_i\}_{i=2}^{\bar{t}-1}$ appropriately.*

7. *If the new sequences for capital and labor inputs are the close enough to the old, we have found the equilibrium path. Otherwise go back to step 5.*

8. *Increase $\bar{t}$ until the solution stops changing.*

D.3 Global Optimization

We parameterize the time paths of fiscal instruments as follows:

$$x_t = \left(\sum_{i=0}^{m_{x0}} \alpha_i^x P_i(t)\right) \exp\left(-\lambda^x t\right) + \left(1 - \exp\left(-\lambda^x t\right)\right) \left(\sum_{j=0}^{m_{xF}} \beta_j^x P_j(t)\right), \tag{D.5}$$

where x_t is any of the fiscal instruments, τ_t^k, τ_t^h, or T_t; $\{P_i(t)\}_{i=0}^{m_{x0}}$ and $\{P_j(t)\}_{j=0}^{m_{xF}}$ are families of Chebyshev polynomials, $\{\alpha_i^x\}_{i=0}^{m_{x0}}$ and $\left\{\beta_j^x\right\}_{j=0}^{m_{xF}}$ are weights on the consecutive elements of the family, and λ^x controls the convergence rate of the fiscal instrument. The orders of the polynomial approximations are given by m_{x0} and m_{xF} for the short-run and long-run dynamics. In our baseline experiment, we set at $m_{\tau_k 0} = m_{\tau_h 0} = 2$, $m_{\tau_k F} = m_{\tau_h F} = 0$, $m_{T0} = 4$ and $m_{TF} = 2$. Given these choices, we end up with the following 17 parameters:

$$\pi_A = \{\alpha_0^k, \alpha_1^k, \alpha_2^k, \beta_0^k, \lambda^k, \alpha_0^h, \alpha_1^h, \alpha_2^h, \beta_0^n, \lambda^n, \alpha_1^T, \alpha_2^T, \alpha_3^T, \alpha_4^T, \beta_0^T, \beta_1^T, \lambda^T\}, \tag{D.6}$$

which determine the time paths of fiscal instruments.

The global optimization algorithm we use to solve the optimal policy problem is an application of the procedure (with some adjustments) described in Kan and Timmer (1987a), and Kan and Timmer (1987b), Kucherenko and Sytsko (2005) and Guvenen (2011). We refer the interested reader to Arnoud, Guvenen, and Kleineberg (2019) for an examination of the performance of this class of algorithms in the standard global optimization tests functions. Our objective function shares many characteristics of these functions, such as: (1) a large number of local optima, (2) flatness near the global optimum, and (3) non-smoothness with regard to the changes in the arguments. These examples in particular highlight the importance of a thorough search through the functions' domains in the global stage in order to identify the global optimum. Also, as documented by Arnoud, Guvenen, and Kleineberg (2019), this class of algorithms is superior to other methods sometimes used for global optimization in economic applications. These facts, coupled with the massive amount of computational power we used and a parallel implementation of the algorithm—the main experiment is conducted on 1200 cores—guarantees that we search the policy space thoroughly enough to find the global optimum.

Algorithm 3 *Let N denote the dimensionality of the parameter space, i.e. for our baseline experiment we have $N = 17$. Then, the algorithm is as follows:*

1. *Initialization*

   (a) *Set the bounds $[b^0_{\min}(n), b^0_{\max}(n)]$ for each parameter $n \in N$.*

   (b) *Generate a matrix of policies of dimension $N \times I \times G_{\max}$ with the use of a quasi-random low-discrepancy sequence (we use Sobol sequence), where I is the number of function evaluations at each global stage and $G_{\max}$ is the maximum number of global iterations.*

   (c) *Let P_g be the $N \times I$ matrix of parameters associated with global iteration g. Set the global iteration $g = 1$. Set the number of local maxima $NLM = 1$.*

2. *Global stage (pre-testing): for each $i \leq I$ do the following steps:*

   (a) *If the library (see part (e)) is non-empty, find within it the already computed transition that has the parameter vector, $P^*_g(\cdot, i)$, closest to $P_g(\cdot, i)$.*

   (b) *Use the previously saved paths for capital and labor associated with $P^*_g(\cdot, i)$ as an initial guess to compute the transitional dynamics associated with parameter vector $P_g(\cdot, i)$. If the library is empty, use as an initial condition an interpolation between the initial and final stationary levels of capital.*

   (c) *Evaluate welfare over transition.*

   (d) *Save the welfare gain/loss at $W(i)$ position of the welfare vector W.*

   (e) *Save $P_g(\cdot, i)$ and its associated equilibrium path for capital and labor into a library of initial conditions.*

3. *Local stage:*

   (a) *Organize vector W in the ascending order and the vector of parameters P_g accordingly.*

   (b) *Define a reduced sample set P_R of size $N \times I_R$, where $I_R \leq I$, i.e. $P_R \equiv \{P_g(\cdot, i) : i \leq I_R\}$.*

   (c) *For each parameter vector in $i \leq I_R$ run the local solver BOBYQA*[4] *i.e. search for the welfare maximizing parameters starting from $P_R(\cdot, i)$.*

   (d) *Denote by W_R to be the vector of dimension I_R of welfare gains/losses for the reduced sample. Save the parameters and welfare associated with it at $P_R(\cdot, i)$ and $W_R(i)$.*

   (e) *Organize vector W_R in the ascending order and the vector of parameters P_R accordingly.*

4. *Update the set of local maxima:*

   (a) *If $g = 1$ then for each k, l where $k \neq l$ and $k, l \leq I_R$ check*

$$\|P_R(\cdot, k) - P_R(\cdot, l)\| > \varepsilon_{LM}. \tag{D.7}$$

   If condition (D.7) holds then we call the two local maxima distinct and set: $NLM = NLM+1$, $P_{LM}(\cdot, NLM) = P_R(\cdot, k)$ and $W_{LM}(NLM) = W_R(k)$, where P_{LM} and W_{LM} are accordingly a matrix of parameters and associated vector of welfare gains/losses.

   (b) *If $g > 1$ then for each k, l where $k \neq l$ and $k, l \leq I_R$ check*

$$\|P_R(\cdot, k) - P_R(\cdot, l)\| > \varepsilon_{LM}, \tag{D.8}$$

   and for each $k \leq I_R$ and $j \leq NLM$ check

$$\|P_R(\cdot, k) - P_{LM}(\cdot, j)\| > \varepsilon_{LM}. \tag{D.9}$$

   If conditions (D.8) and (D.9) are satisfied, set $NLM = NLM+1$, $P_{LM}(\cdot, NLM) = P_R(\cdot, k)$, and $W_{LM}(NLM) = W_R(k)$.

5. *Adjust the bounds:*

   (a) *For each $n \in N$ compute the following auxiliary variables*

$$\begin{aligned} X_{\max}(n) &= \max_{i \in \{1,\ldots,\min(NLM, NLM_b)\}} \left\{b^0_{\min}(n) + P_{LM}(n, i) \times \left(b^0_{\max}(n) - b^0_{\min}(n)\right)\right\} \\ X_{\min}(n) &= \min_{i \in \{1,\ldots,\min(NLM, NLM_b)\}} \left\{b^0_{\min}(n) + P_{LM}(n, i) \times \left(b^0_{\max}(n) - b^0_{\min}(n)\right)\right\} \end{aligned}$$

   where NLM_b is the upper bound on the number of local minima used in the adjustment of the bounds.

[4] See Powell (2009). The parameters for BOBYQA were: RHOBEG$= 10^{-1}$, and RHOEND$= 10^{-5}$

(b) *For each $n \in N$, adjust the bounds as follows*

(i) *If* $(X_{\max}(n) - X_{\min}(n)) > \xi_1 (b^0_{\max}(n) - b^0_{\min}(n))$ *then*

$$\begin{aligned} b^1_{\max}(n) &= X_{\max}(n) + \theta_1 (X_{\max}(n) - X_{\min}(n)) \\ b^1_{\min}(n) &= X_{\min}(n) - \theta_1 (X_{\max}(n) - X_{\min}(n)) \\ stop(n) &= 0 \end{aligned}$$

(ii) *If* $X_{\min}(n) > (b^0_{\max}(n) - \xi_2 (b^0_{\max}(n) - b^0_{\min}(n)))$ *then*

$$\begin{aligned} b^1_{\max}(n) &= b^0_{\max}(n) + \theta_2 \left(b^0_{\max}(n) - b^0_{\min}(n)\right) \\ b^1_{\min}(n) &= b^0_{\max}(n) - \theta_3 \left(b^0_{\max}(n) - b^0_{\min}(n)\right) \\ stop(n) &= 0 \end{aligned}$$

(iii) *If* $X_{\max}(n) < (b^0_{\min}(n) + \xi_2 (b^0_{\max}(n) - b^0_{\min}(n)))$ *then*

$$\begin{aligned} b^1_{\max}(n) &= b^0_{\min}(n) + \theta_3 \left(b^0_{\max}(n) - b^0_{\min}(n)\right) \\ b^1_{\min}(n) &= b^0_{\min}(n) - \theta_2 \left(b^0_{\max}(n) - b^0_{\min}(n)\right) \\ stop(n) &= 0 \end{aligned}$$

(iv) *Else*

$$\begin{aligned} b^1_{\max}(n) &= X_{\max}(n) + \theta_4 \left(b^0_{\max}(n) - b^0_{\min}(n)\right) \\ b^1_{\min}(n) &= X_{\min}(n) - \theta_4 \left(b^0_{\max}(n) - b^0_{\min}(n)\right) \\ stop(n) &= 1 \end{aligned}$$

6. *Stopping rule: if $stop(n) = 1$ for all $n \in N$, then, let NLM_g be the number of local minima found in the current global iteration and compute*

$$NLM_{exp} = \frac{NLM_g(I_R - 1)}{I_R - NLM_g - 2}$$

provided that $I_R > NLM_g + 2$. If $NLM_{exp} < NLM_g + 0.5$ then STOP and go to Step 7.[5] *Otherwise set the global iteration to $g = g + 1$, use bound updates i.e. set $b^0_{\max} = b^1_{\max}$, $b^0_{\min} = b^1_{\min}$ and go to Step 2.*

7. *Pick the global optimum, i.e.*

$$j_{\max} = \arg \max_{j \in NLM} W_R(j)$$

[5] This is a Baysian rule, see Guvenen (2011) for an heuristic explanation for it.

$$P_{GM}(\cdot) \quad = \quad P_{LM}(\cdot, j_{\max})$$

In the computational implementation of the algorithm presented above we have to impose the values of the following parameters: (1) number of parameters in the approximation of the time paths N (ii) initial bounds on the parameters $\left[b^0_{n,\min}, b^0_{n,\max}\right]$ (iii) number of function evaluations in the global stage I (iv) maximum number of global iterations $G_{\max}$ (5) the size of the reduced sample I_R (6) the distance separating two local maxima ε_{LM} (7) bounds adjustment parameters $\{NLM_b, \xi_1, \xi_2, \theta_1, \theta_2, \theta_3, \theta_4\}$ (8) Stopping tolerance for the bounds ε_B. In the main experiment of the paper we set $N = 17$ - see the detailed description of the main experiment in Section 3.2 of the paper.

The number of function evaluations in the global stage is set to $I = 240,000$, which is the multiple of 1200, the number of cores we use in the computational implementation of the algorithm. Further, we set the maximum number of global iterations, $G_{\max}$, to 8, and in numerous robustness checks we have not hit this bound. The size of the reduced sample, I_R, is set to 1200, so that each core conducts one local search. The distance separating two local maxima is set to

$$\varepsilon_{LM} \equiv \sqrt{\left(\sum_{n=1}^{N}(b^0_{max}(n) - b^0_{min}(n))^2\right)},$$

where $b^0_{max}(n)$ and $b^0_{min}(n)$ are the initial bounds set in step 1. The bounds adjustment parameters are set to the following values: $NLM_b = 8$, $\xi_1 = 0.5$, $\xi_2 = 0.1$, $\theta_1 = 0.2$, $\theta_2 = 0.7$, $\theta_3 = 0.1$, $\theta_4 = 0.15$. The idea behind the adjustment of the bounds in step 5 part (b) is the following. Fix a particular $n \in N$, then

(i) If the local maxima are distributed somewhat evenly over the bounds, i.e. $(X_{\max}(n) - X_{\min}(n)) < 0.5 = \alpha_1$, then we increase the bounds on both sides by 20 percent (θ_1) of $(X_{\max}(n) - X_{\min}(n))$.

(ii) If the local maxima are bunched up in the top 10 percent (ξ_2) of the bounds, then we increase the upper and lower bounds in proportion to $(b^0_{\max}(n) - b^0_{\min}(n))$ using the parameters θ_2, and θ_3.

(iii) If the local maxima are bunched up in the bottom 10 percent (ξ_2) of the the bounds, then we decrease the upper and lower bounds in proportion to $(b^0_{\max}(n) - b^0_{\min}(n))$ using the parameters θ_2, and θ_3.

(iv) If none of the other conditions are satisfied, that is, it the local maxima are bunched up in the middle of the bounds we reduce the upper bound and increase the lower bound by by 15 percent (θ_4).

We also conducted robustness checks with regard to the bound adjustment procedure and concluded that the values of these parameters affect just the pace of convergence rather than the results of the global optimization.

Scalability and Computational Resources. Our algorithm is implemented in modern Fortran language using MPI library and our experiments were conducted at the Niagara cluster at the SciNet HPC Consortium located at the University of Toronto. The algorithm is highly scalable and we run our benchmark experiment multiple times on 1200 cores for about 96 hours. Other experiments in the paper were run on 400 to 1200 cores for various time depending on their scale. The basic hardware specifications of the Niagara cluster are: (1) 2024 nodes, each with 40 Intel Skylake or Cascadelake cores at 2.4GHz, for a total of 80,640 cores (2) 202 GB (188 GiB) of RAM per node. (3) EDR Infiniband network in a so-called 'Dragonfly+' topology (4) Peak performance of the cluster is about 3.6 PFlops (6.25 PFlops theoretical). See Ponce et al. (2019) and Loken et al. (2010) for more details on the cluster and here for more information on hardware specifications. Niagara was the 53rd fastest supercomputer on the TOP500 list of June 2018, and is at number 82 as of November 2020, see the list.

E Welfare Decomposition

This appendix provides the proofs for the propositions associated with the welfare decomposition. We repeat definitions and the propositions themselves here for convenience.

E.1 Definitions

Average welfare gain Consider a policy reform and denote by $\{c_t^j, h_t^j\}$ the equilibrium consumption and labor paths of a household with and without the reform, with $j = R$ or $j = NR$ respectively. The average welfare gain, Δ, that results from implementing the reform is defined as the constant (over time and across agents) percentage increase to c_t^{NR} that equalizes the utilitarian welfare to the value associated with the reform, that is,

$$\int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u\left((1+\Delta)c_t^{NR}, h_t^{NR}\right) \right] d\lambda_0 = \int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t^R, h_t^R\right) \right] d\lambda_0, \tag{E.1}$$

where λ_0 is the initial distribution over states (a_0, e_0). These welfare gains associated with the utilitarian welfare function can be decomposed into three parts:

1. Level effect Let the aggregate level of c_t and h_t at each t be

$$C_t^j \equiv \int c_t^j d\lambda_t^j, \quad \text{and} \quad H_t^j \equiv \int h_t^j d\lambda_t^j,$$

where λ_t^j is the distribution over (a_0, e^t) conditional on whether or not the reform is implemented. Then, the level effect, Δ_L, is given by

$$\sum_{t=0}^{\infty} \beta^t u\left((1+\Delta_L)C_t^{NR}, H_t^{NR}\right) = \sum_{t=0}^{\infty} \beta^t u\left(C_t^R, H_t^R\right). \tag{E.2}$$

2. Insurance effect Let $\{\bar{c}_t^j(a_0, e_0), \bar{h}_t^j(a_0, e_0)\}$ denote a certainty-equivalent sequence of consumption and labor conditional on a household's initial state that satisfies

$$\sum_{t=0}^{\infty} \beta^t u\left(\bar{c}_t^j(a_0, e_0), \bar{h}_t^j(a_0, e_0)\right) = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t^j, h_t^j\right) \right]. \tag{E.3}$$

Next, let $\bar{C}_t^j$ and $\bar{H}_t^j$ denote aggregate certainty equivalents, that is

$$\bar{C}_t^j = \int \bar{c}_t^j(a_0, e_0) d\lambda_0, \quad \text{and} \quad \bar{H}_t^j = \int \bar{h}_t^j(a_0, e_0) d\lambda_0, \quad \text{for } j = R, NR. \tag{E.4}$$

The insurance effect, Δ_I, is defined by

$$1+\Delta_I \equiv \frac{1-p^R_{risk}}{1-p^{NR}_{risk}}, \quad \text{where} \sum_{t=0}^{\infty} \beta^t u\left((1-p^j_{risk})C^j_t, H^j_t\right) = \sum_{t=0}^{\infty} \beta^t u\left(\bar{C}^j_t, \bar{H}^j_t\right). \tag{E.5}$$

Here, p^j_{risk} is the welfare cost of risk in the economies with and without reform.

3. Redistribution effect The redistribution effect, Δ_R, satisfies

$$1+\Delta_R \equiv \frac{1-p^R_{ineq}}{1-p^{NR}_{ineq}}, \quad \text{where} \sum_{t=0}^{\infty} \beta^t u\left((1-p^j_{ineq})\bar{C}^j_t, \bar{H}^j_t\right) = \int \sum_{t=0}^{\infty} \beta^t u\left(\bar{c}^j_t(a_0,e_0), \bar{h}^j_t(a_0,e_0)\right) d\lambda_0. \tag{E.6}$$

Analogously to p^j_{risk}, p^j_{ineq} denotes the cost of inequality.

Choice of certainty equivalents. Notice that there can be many certainty-equivalent paths that satisfy equation (5.3). These paths could differ over time and over levels of consumption and labor. In general, these choices can affect the components of the decomposition. If the certainty equivalents for consumption and leisure follow parallel paths over time, however, these choices are immaterial.

Assumption 4 *The certainty equivalents display parallel patterns if* $\bar{c}^j_t(a_0,e_0) = \eta^j(a_0,e_0)\tilde{C}^j_t$*, and* $1-\bar{h}^j_t(a_0,e_0) = \eta^j(a_0,e_0)(1-\tilde{H}^j_t)$*, for some function* $\eta^j(a_0,e_0)$ *and paths* $\{\tilde{C}^j_t\}$*, and* $\{\tilde{H}^j_t\}$*.*

There are two ways in which this assumption is restrictive. First, it assumes that the certainty equivalents of households with different initial conditions are a proportion of the same paths, with only the degree of proportionality, $\eta^j(a_0,e_0)$, changing; this is the property we are referring to as "parallel patterns." Second, it assumes that the degree of proportionality applies in the same way to the path of consumption and leisure. Reasonable deviations from the first restriction lead to small changes in the results benchmark results in Table 4 below. The second restriction is more consequential, because the way one decomposes the differences between consumption and leisure affects the amount of curvature that is absorbed by the insurance and redistribution effects.[6] The choice of certainty equivalents, however, never matters for the magnitude of the level effect. Under this assumption, we can establish Proposition 4. In particular, one could impose that the certainty-equivalent paths should follow their corresponding aggregates, that is $\tilde{C}^j_t = C^j_t$ and $\tilde{H}^j_t = H^j_t$. In any case, as long as Assumption 4 is satisfied this choice does not matter.

[6]More precisely, the degree of proportionality is taken to a different power if it multiplies only consumption, for instance, $(\eta c)^\gamma(1-h)^{1-\gamma} = \eta^\gamma(c)^\gamma(1-h)^{1-\gamma}$ versus if it multiplies consumption and leisure as in Assumption 4, $(\eta c)^\gamma(\eta(1-h))^{1-\gamma} = \eta(c)^\gamma(1-h)^{1-\gamma}$.

E.2 Proofs

Proposition 3 *If preferences are such that, for any scalar x, $u(xc,h) = g(x)u(c,h)$ for some totally multiplicative function $g(\cdot)$, then*

$$1 + \Delta = (1 + \Delta_L)(1 + \Delta_I)(1 + \Delta_R).$$

Proof. For any sequence $\{c_t, h_t\}$, let

$$U(\{c_t, h_t\}) \equiv \sum_{t=0}^{\infty} \beta^t u(c_t, h_t),$$

and notice that U inherits the property that, for any scalar x,

$$U(\{xc_t, h_t\}) = g(x) U(\{c_t, h_t\}). \tag{E.7}$$

Suppressing the dependence on (a_0, e_0), it follows that

$$\begin{aligned}
\int \mathbb{E}_0 \left[U\left(\left\{c_t^R, h_t^R\right\}\right)\right] d\lambda_0 &\overset{(E.3)}{=} \int U\left(\left\{\bar{c}_t^R, h_t^R\right\}\right) d\lambda_0 \overset{(E.6)}{=} U\left(\left\{\left(1 - p_{ineq}^R\right) \bar{C}_t^R, \bar{H}_t^R\right\}\right) \\
&\overset{(E.7)}{=} g\left(1 - p_{ineq}^R\right) U\left(\left\{\bar{C}_t^R, \bar{H}_t^R\right\}\right) \\
&\overset{(E.5)}{=} g\left(1 - p_{ineq}^R\right) U\left(\left\{\left(1 - p_{risk}^R\right) C_t^R, H_t^R\right\}\right) \\
&\overset{(E.7)}{=} g\left(\left(1 - p_{ineq}^R\right)\left(1 - p_{risk}^R\right)\right) U\left(\left\{C_t^R, H_t^R\right\}\right) \\
&\overset{(E.2)}{=} g\left(\left(1 - p_{ineq}^R\right)\left(1 - p_{risk}^R\right)\right) U\left(\left\{\left(1 + \Delta_L\right) C_t^{NR}, H_t^{NR}\right\}\right) \\
&\overset{(E.7)}{=} g\left(\left(1 + \Delta_L\right)\left(1 - p_{ineq}^R\right)\left(1 - p_{risk}^R\right)\right) U\left(\left\{C_t^{NR}, H_t^{NR}\right\}\right) \\
&\overset{(E.7)}{=} g\left(\left(1 + \Delta_L\right)\left(1 - p_{ineq}^R\right)\frac{\left(1 - p_{risk}^R\right)}{\left(1 - p_{risk}^{NR}\right)}\right) U\left(\left\{\left(1 - p_{risk}^{NR}\right) C_t^{NR}, H_t^{NR}\right\}\right) \\
&\overset{(E.5)}{=} g\left(\left(1 + \Delta_L\right)\left(1 + \Delta_I\right)\left(1 - p_{ineq}^R\right)\right) U\left(\left\{\bar{C}_t^{NR}, \bar{H}_t^{NR}\right\}\right) \\
&\overset{(E.7)}{=} g\left(\left(1 + \Delta_L\right)\left(1 + \Delta_I\right)\frac{\left(1 - p_{ineq}^R\right)}{\left(1 - p_{ineq}^{NR}\right)}\right) U\left(\left\{\left(1 - p_{ineq}^{NR}\right) \bar{C}_t^{NR}, \bar{H}_t^{NR}\right\}\right) \\
&\overset{(E.6)}{=} g\left(\left(1 + \Delta_L\right)\left(1 + \Delta_I\right)\left(1 + \Delta_R\right)\right) \int U\left(\left\{\bar{c}_t^{NR}, \bar{h}_t^{NR}\right\}\right) d\lambda_0 \\
&\overset{(E.5)}{=} g\left(\left(1 + \Delta_L\right)\left(1 + \Delta_I\right)\left(1 + \Delta_R\right)\right) \int \mathbb{E}_0 \left[U\left(\left\{c_t^{NR}, h_t^{NR}\right\}\right)\right] d\lambda_0 \\
&\overset{(E.7)}{=} \int \mathbb{E}_0 \left[U\left(\left\{\left(1 + \Delta_L\right)\left(1 + \Delta_I\right)\left(1 + \Delta_R\right) c_t^{NR}, h_t^{NR}\right\}\right)\right] d\lambda_0.
\end{aligned}$$

The result, then, follows from the definition of Δ in equation ($E.1$). ■

Proposition 4 *For balanced-growth-path preferences, as specified in equation (B.1), if the certainty equivalents satisfy Assumption 4, then the components Δ_L, Δ_I, and Δ_R are independent of the paths $\{\tilde{C}_t^j\}$, and $\{\tilde{H}_t^j\}$.*

Proof. The level effect, Δ_L, is independent of the choice of certainty equivalents by definition. The result in Proposition 5, then, implies that the insurance effect, Δ_I, is also invariant. Finally, it follows from the result in Proposition 3 that the redistribution effect, Δ_R, is invariant. ■

Proposition 5 *If the certainty equivalents satisfy Assumption 4, then, maximizing*

$$W^0 = \left(\int E_0 \left[U\left(\{c_t, h_t\}\right)\right]^{\frac{1}{1-\sigma}} d\lambda_0 \right)^{1-\sigma}$$

is equivalent to maximizing $(1+\Delta_L)(1+\Delta_I)$.

Proof. First notice that, for $j = R, NR$, it follows from Assumption 4 that

$$E_0 \left[U\left(\left\{ c_t^j, h_t^j \right\} \right) \right] \overset{(E.3)}{=} U\left(\left\{ \eta_0^j \tilde{C}_t^j, 1 - \eta_0^j (1 - \tilde{H}_t^j) \right\} \right) = \left(\eta_0^j \right)^{1-\sigma} U\left(\left\{ \tilde{C}_t^j, \tilde{H}_t^j \right\} \right), \tag{E.8}$$

and, from equation (E.4), it follows that

$$\bar{C}_t^j = \int \eta^j(a_0, e_0) d\lambda_0 \, \tilde{C}_t^j, \quad \text{and} \quad \bar{H}_t^j = 1 - \int \eta^j(a_0, e_0) d\lambda_0 \, (1 - \tilde{H}_t^j). \tag{E.9}$$

Therefore,

$$\begin{aligned}
\left(\int E_0 \left[U\left(\{c_t^R, h_t^R\}\right)\right]^{\frac{1}{1-\sigma}} d\lambda_0 \right)^{1-\sigma} &\overset{(E.8)}{=} \left(\int \eta_0^R d\lambda_0 \right)^{1-\sigma} U\left(\left\{ \tilde{C}_t^R, \tilde{H}_t^R \right\} \right) \\
&\overset{(E.8)}{=} U\left(\left\{ \int \eta_0^R d\lambda_0 \tilde{C}_t^R, 1 - \int \eta_0^R d\lambda_0 (1 - \tilde{H}_t^R) \right\} \right) \\
&\overset{(E.9)}{=} U\left(\{ \bar{C}_t^R, \bar{H}_t^R \} \right) \\
&\overset{(E.5)}{=} U\left(\{ \left(1 - p_{risk}^R\right) C_t^R, H_t^R \} \right) \\
&= \left(1 - p_{risk}^R\right)^{1-\sigma} U\left(\{ C_t^R, H_t^R \} \right) \\
&\overset{(E.2)}{=} \left(1 - p_{risk}^R\right)^{1-\sigma} U\left(\{ (1+\Delta_L) C_t^{NR}, H_t^{NR} \} \right) \\
&= \left(\frac{\left(1 - p_{risk}^R\right)}{\left(1 - p_{risk}^{NR}\right)} (1+\Delta_L) \right)^{1-\sigma} U\left(\{ \left(1 - p_{risk}^{NR}\right) C_t^{NR}, H_t^{NR} \} \right) \\
&\overset{(E.5)}{=} \left((1+\Delta_I)(1+\Delta_L) \right)^{1-\sigma} U\left(\{ \bar{C}_t^{NR}, \bar{H}_t^{NR} \} \right) \\
&\overset{(E.9)}{=} \left((1+\Delta_I)(1+\Delta_L) \right)^{1-\sigma} U\left(\left\{ \int \eta_0^{NR} d\lambda_0 \tilde{C}_t^{NR}, 1 - \int \eta_0^{NR} d\lambda_0 (1 - \tilde{H}_t^{NR}) \right\} \right) \\
&\overset{(E.8)}{=} \left((1+\Delta_I)(1+\Delta_L) \right)^{1-\sigma} \left(\int \eta_0^{NR} d\lambda_0 \right)^{1-\sigma} U\left(\left\{ \tilde{C}_t^{NR}, \tilde{H}_t^{NR} \right\} \right) \\
&\overset{(E.8)}{=} \left((1+\Delta_I)(1+\Delta_L) \right)^{1-\sigma} \left(\int E_0 \left[U\left(\{c_t^{NR}, h_t^{NR}\}\right)\right]^{\frac{1}{1-\sigma}} d\lambda_0 \right)^{1-\sigma},
\end{aligned}$$

which completes the proof. ■

E.3 Alternative Decomposition: Consumption versus Labor

In this appendix we consider an alternative decomposition, which helps understand how much of the welfare gains from a particular policy are associated with changes in consumption behavior and how much comes from changes in labor-supply decisions.

Consider a policy reform and and denote by $\{c_t^j, h_t^j\}$ the equilibrium consumption and labor paths of a household with and without the reform, with $j = R$ or $j = NR$ respectively. The average welfare gain, Δ, that results from implementing the reform is defined as the constant (over time and across households) percentage increase to c_t^{NR} that equalizes the utilitarian welfare to the value associated with the reform; that is,

$$\int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left((1+\Delta) c_t^{NR}, h_t^{NR} \right) \right] d\lambda_0 = \int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left(c_t^R, h_t^R \right) \right] d\lambda_0,$$

where λ_0 is the initial distribution over states (a_0, e_0). Then, the consumption component of the welfare gains, Δ_C, comes from first switching only consumption decisions to the ones that follow from the reform, that is Δ_C solves[7]

$$\int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left((1+\Delta_C) c_t^{NR}, h_t^{NR} \right) \right] d\lambda_0 = \int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left(c_t^R, h_t^{NR} \right) \right] d\lambda_0.$$

Then, the labor component, Δ_H, is such that

$$\int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left((1+\Delta_H) c_t^{R}, h_t^{NR} \right) \right] d\lambda_0 = \int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left(c_t^R, h_t^R \right) \right] d\lambda_0.$$

It follows immediately from these definitions that $(1+\Delta) = (1+\Delta_C)(1+\Delta_H)$.

Computing this decomposition for our benchmark results yielded $\Delta_C = 2.63\%$ and $\Delta_H = 0.87\%$, so that most of the average welfare gains of $\Delta = 3.52\%$ are accounted for by the consumption component. Combining this finding with the fact that most of the welfare gains come from redistribution, this is indicative that the majority of the utilitarian welfare gains come from the reduction of consumption inequality.

[7] To compute

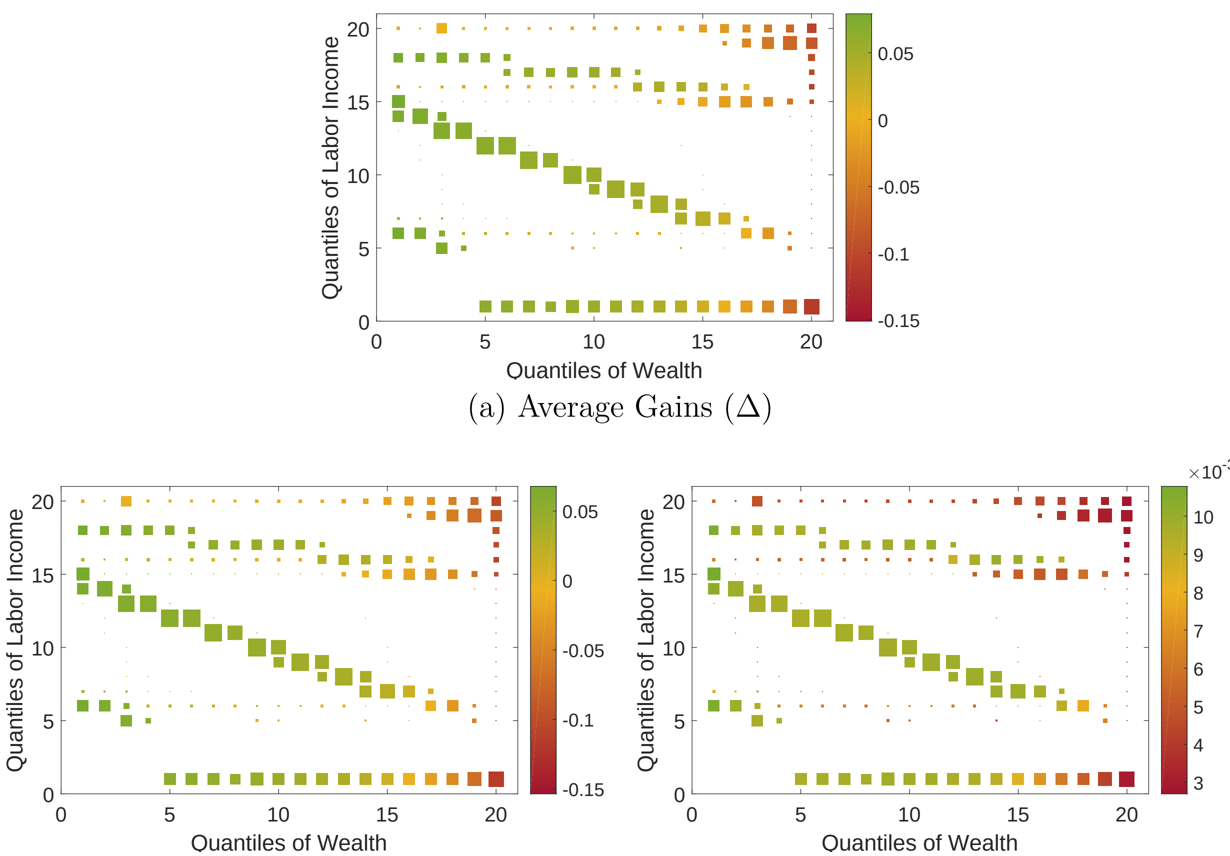
$$\int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left(c_t^R, h_t^{NR} \right) \right] d\lambda_0,$$

we define an asset policy function residually using the households' budget constraints to guarantee that the corresponding sequences $\{c_t^R, h_t^{NR}\}$ are budget feasible.

One complicating factor in the analysis of these results is that the policy functions used to compute

$$\int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u \left(c_t^R, h_t^{NR} \right) \right] d\lambda_0$$

are not consistent with one another: c^R and h^{NR} would not have been chosen together (generically) with intratemporal as well as intertemporal (since utility is non-separable) optimality conditions being violated. As a result, Δ_C captures the benefit of switching from c^{NR} to c^R but also the losses from moving from consumption choices that are compatible with the labor-supply decisions to ones that are not. The opposite is true for Δ_H which, to some extent, captures the result of realigning consumption and labor-supply decisions. Moreover, Δ_C and Δ_H both include gains steaming from redistribution, insurance and the reductions in overall distortions, so it may be interesting to consider an even more elaborate decomposition combining the one we present in the main text of the paper with this one.



(a) Average Gains (Δ)

(b) Consumption Component (Δ_C)

(c) Labor Component (Δ_H)

Figure 2: Conditional Welfare Gains: Consumption-Labor Decomposition

Note: In all three figures the axis display 20-quantiles of wealth and labor income. The size of the tiles is proportional to the density of households in the initial stationary distribution. The color of the tile represents the welfare gain or loss associated with the optimal policy conditional on the household's level of wealth and labor income with the corresponding scale to the right of each figure. The first panel presents Δ, the second Δ_C, and the third Δ_H.

Figure 2 displays the results for the decomposition conditional on wealth and income. Panel 2a shows how the average welfare gains, Δ, are distributed between households with different levels of wealth and labor income. In line with the redistribution achieved, to a large extent via high initial capital income taxes, wealthy households lose and asset-poor households win. Conditional on wealth quantile, however, the welfare gains remain similar across quantiles of labor income. This is because the provision of insurance benefits all households in a similar way—risk is more consequential to low-productivity households, but since transitory shocks are roughly multiplicative these households actually face less income risk.

Panel 2b shows what the consumption component, Δ_C, is distributed very similarly to the average component, Δ, plotted in Panel 2a. In particular, notice that the scales to the right of these two panels are exactly the same, which is indicative again of the fact that the consumption component captures most of the average gains. In contrast, that scale in Panel 2c which shows the labor component, Δ_H, is of significantly lower magnitude and the distribution of welfare gains for this component is more concentrated. This latter point can be appreciated by the fact that there is less yellow (intermedite levels of welfare gains) in Panel 2c than in the other two. This is consistent with the finding that labor productivity increases as a result of a relatively increase in labor supply of the higher productivity households (which also tend to be wealthier). Finally, we want clarify that while high-productivity households do work*relative* more as a result of the optimal reform most of them do not actually work more in absolute terms, since the policy implies, in particular, a substantial increase in labor-income taxes; this explains why the scale in Panel 2c only contains only positive numbers.

F Complete-Markets Model

F.1 Environment

Consider an economy populated by a continuum of infinitely-lived agents divided into types $i \in I$ of size π_i. Each agent of type $i \in I$ ranks streams of consumption and hours worked $\{c_{i,t}, h_{i,t}\}$ according to the preferences

$$\sum_{t=0}^{\infty} \beta^t u\left(c_{i,t}, h_{i,t}\right), \tag{F.1}$$

with period utility function given by

$$u\left(c, h\right) = \frac{\left(c^{\gamma}\left(1-h\right)^{1-\gamma}\right)^{1-\sigma}}{1-\sigma}.$$

An agent of type $i \in I$ with productivity e_i works $h_{i,t}$ each period. Aggregates are denoted without the subscript i: $C_t = \sum_i \pi_i c_{i,t}$, $N_t = \sum_i \pi_i e_i h_{i,t}$ and $K_t = \sum_i \pi_i k_{i,t}$.

Consumption-capital good is produced with a concave, constant returns to scale technology, $F\left(K, N\right)$, that uses aggregate capital, K, and aggregate labor, N. Thus, the resource constraint of the economy is given by

$$C_t + G_t + K_{t+1} = F\left(K_t, N_t\right) + \left(1-\delta\right) K_t, \quad \text{for} \quad t \geq 0 \tag{F.2}$$

where $\{G_t\}_{t=0}^{\infty}$ is an exogenous sequence of government spending and δ is the rate of depreciation of the capital stock.

F.1.1 Agent's problem

Let p_t denote the price of the consumption good in period t in terms of consumption in period 0 (so that $p_0 = 1$), w_t and r_t denote the real wage and the rental rate of capital in period t. Let $b_{i,t}$ and $k_{i,t}$ denote the number of units of government debt and capital held between periods $t-1$ and t, and R_t denote its gross return (between $t-1$ and t). Given $k_{i,0}$, $b_{i,0}$, prices $\{p_t, w_t, r_t\}_{t=0}^{\infty}$ and policies $\left\{\tau_t^h, \tau_t^k, T_t\right\}_{t=0}^{\infty}$, the agent chooses $\{c_{i,t}, h_{i,t}, k_{i,t+1}, b_{i,t+1}\}$ to maximize (1) subject to the intertemporal budget constraint

$$\sum_{t=0}^{\infty} p_t\left(\left(1+\tau^c\right) c_{i,t} + k_{i,t+1} + b_{i,t+1}\right) \leq \sum_{t=0}^{\infty} p_t\left(\left(1-\tau_t^h\right) w_t e_i h_{i,t} + R_t\left(k_{i,t} + b_{i,t}\right) + T_t\right),$$

where $R_t \equiv 1 + \left(1-\tau_t^k\right)\left(r_t - \delta\right)$, for $t \geq 0$. Since $p_t = R_{t+1} p_{t+1}$, and defining $T \equiv \sum_{t=0}^{\infty} p_t T_t$, this is equivalent to

$$\sum_{t=0}^{\infty} p_t\left(\left(1+\tau^c\right) c_{i,t} - \left(1-\tau_t^h\right) w_t e_i h_{i,t}\right) \leq R_0 a_{i,0} + T, \tag{F.3}$$

where $a_{i,0} = k_{i,0} + b_{i,0}$. The first order conditions of agent i's problem are:

$$[c_{i,t}] : \frac{\beta^t u_c(c_{i,t}, h_{i,t})}{(1+\tau^c) p_t} = \phi, \quad \forall\, t \geq 0,$$
$$[h_{i,t}] : \beta^t u_h(c_{i,t}, h_{i,t}) = -\phi p_t \left(1 - \tau_t^h\right) w_t e_i, \quad \forall\, t \geq 0,$$

thus, in particular,

$$p_t = \beta^t \frac{u_c(c_{i,t}, h_{i,t})}{u_c(c_{i,0}, h_{i,0})}, \quad \forall\, t \geq 0, \tag{F.4}$$
$$\frac{u_h(c_{i,t}, h_{i,t})}{u_c(c_{i,t}, h_{i,t})} = -e_i w_t \frac{\left(1 - \tau_t^h\right)}{(1+\tau^c)}, \quad \forall\, t \geq 0, \tag{F.5}$$

which holds across all agents.

F.1.2 Firm's problem

The first order conditions for the firm problem are:

$$r_t = F_k(K_t, N_t), \quad \forall\, t \geq 0, \tag{F.6}$$
$$w_t = F_h(K_t, N_t), \quad \forall\, t \geq 0. \tag{F.7}$$

F.1.3 Government's budget constraint

Each period the government finances the expenses G_t and lump sum transfers T_t with proportional income taxes on capital τ_t^k and labor τ_t^h. The government's intertemporal budget constraint is

$$\sum_t p_t (G_t + R_t B_t + T_t) = \sum_t p_t \left(\tau^c C_t + \tau_t^h w_t N_t + \tau_t^k (r_t - \delta) K_t + B_{t+1}\right),$$

which is equivalent to

$$R_0 B_0 + T + \sum_t p_t G_t = \sum_t p_t \left(\tau^c C_t + \tau_t^h w_t N_t + \tau_t^k (r_t - \delta) K_t\right). \tag{F.8}$$

F.1.4 Competitive equilibrium

Definition Given $\{a_{i,0}\}$, K_0, and B_0, a competitive equilibrium is a policy $\left\{\tau_t^h, \tau_t^k, T_t\right\}_{t=0}^{\infty}$, a price system $\{p_t, w_t, r_t\}_{t=0}^{\infty}$ and an allocation $\{c_{i,t}, h_{i,t}, K_{t+1}\}_{t=0}^{\infty}$ such that: (i) agents choose $\{c_{i,t}, h_{i,t}\}_{t=0}^{\infty}$ to maximize utility subject to budget constraint ($F.3$) taking policies and prices (that satisfy $p_t = R_{t+1} p_{t+1}$) as given; (ii) firms maximize profits; (iii) the government's budget constraint ($F.8$) holds; (iv) markets clear: the resource constraint ($F.2$) holds.

F.2 A Simple Characterization of Equilibrium

Let $\varphi \equiv \{\varphi_i\}$ be the market weights normalized so that $\sum_i \varphi_i = 1$ with $\varphi_i \geq 0$. Then, given aggregate levels C_t and N_t, the individual levels can be found by solving the following static subproblem for each period t:

$$U(C_t, N_t; \varphi) \equiv \max_{c_{i,t}, h_{i,t}} \sum_i \pi_i \varphi_i u(c_{i,t}, h_{i,t}) \quad \text{s.t.} \sum_i \pi_i c_{i,t} = C_t, \quad \text{and} \quad \sum_i \pi_i e_i h_{i,t} = N_t. \tag{F.9}$$

In what follows, we obtain a simple formula for the aggregate indirect utility $U(C_t, N_t; \varphi)$. The Lagrangian for this problem is

$$L = \sum_i \pi_i \varphi_i \left[\frac{\left(c_{i,t}^{\gamma} (1 - h_{i,t})^{1-\gamma}\right)^{1-\sigma}}{1-\sigma} \right] + \theta_t^c \left(C_t - \sum_i \pi_i c_{i,t} \right) - \theta_t^h \left(N_t - \sum_i \pi_i e_i h_{i,t} \right),$$

where θ_t^c and θ_t^h are Lagrange multipliers. The first order conditions are

$$[c_{i,t}] : \varphi_i \left(c_{i,t}^{\gamma} (1 - h_{i,t})^{1-\gamma}\right)^{1-\sigma} \gamma c_{i,t}^{-1} = \theta_t^c, \quad \forall\, t \geq 0, \tag{F.10}$$

$$[h_{i,t}] : \varphi_i \left(c_{i,t}^{\gamma} (1 - h_{i,t})^{1-\gamma}\right)^{1-\sigma} (1-\gamma)(1 - h_{i,t})^{-1} = e_i \theta_t^h, \quad \forall\, t \geq 0, \tag{F.11}$$

rearranging yields

$$c_{i,t} = \frac{\gamma}{(1-\gamma)} \frac{\theta_t^h}{\theta_t^c} e_i (1 - h_{i,t}),$$

so that

$$c_{i,t} = \left(\frac{(\theta_t^c)^{\sigma+\gamma(1-\sigma)} \left(\theta_t^h\right)^{(1-\gamma)(1-\sigma)}}{\gamma^{\sigma+\gamma(1-\sigma)} (1-\gamma)^{(1-\gamma)(1-\sigma)}} \frac{(e_i)^{(1-\gamma)(1-\sigma)}}{\varphi_i} \right)^{-\frac{1}{\sigma}},$$

$$h_{i,t} = 1 - \left(\frac{(\theta_t^c)^{\gamma(1-\sigma)} \left(\theta_t^h\right)^{1-\gamma(1-\sigma)}}{\gamma^{\gamma(1-\sigma)} (1-\gamma)^{1-\gamma(1-\sigma)}} \frac{(e_i)^{1-\gamma(1-\sigma)}}{\varphi_i} \right)^{-\frac{1}{\sigma}},$$

and summing across types (given that $C_t = \sum_j \pi_j c_{j,t}$, and $N_t = \sum_j \pi_j e_j h_{j,t}$)

$$C_t = \left(\frac{(\theta_t^c)^{\sigma+\gamma(1-\sigma)} \left(\theta_t^h\right)^{(1-\gamma)(1-\sigma)}}{\gamma^{\sigma+\gamma(1-\sigma)} (1-\gamma)^{(1-\gamma)(1-\sigma)}} \right)^{-\frac{1}{\sigma}} \sum_i \pi_i \left(\frac{(e_i)^{(1-\gamma)(1-\sigma)}}{\varphi_i} \right)^{-\frac{1}{\sigma}}, \tag{F.12}$$

$$N_t = 1 - \left(\frac{(\theta_t^c)^{\gamma(1-\sigma)} \left(\theta_t^h\right)^{1-\gamma(1-\sigma)}}{\gamma^{\gamma(1-\sigma)} (1-\gamma)^{1-\gamma(1-\sigma)}} \right)^{-\frac{1}{\sigma}} \sum_i \pi_i \left(\frac{(e_i)^{1-\gamma(1-\sigma)}}{\varphi_i} \right)^{-\frac{1}{\sigma}}. \tag{F.13}$$

It follows that

$$c_{i,t}^{m}\left(C_t, N_t; \varphi\right) = \omega_i^c C_t, \tag{F.14}$$

$$h_{i,t}^{m}\left(C_t, N_t; \varphi\right) = 1 - \omega_i^h \left(1 - N_t\right), \tag{F.15}$$

where

$$\omega_i^c \equiv \frac{\left(\frac{(e_i)^{(1-\gamma)(1-\sigma)}}{\varphi_i}\right)^{-\frac{1}{\sigma}}}{\sum_j \pi_j \left(\frac{(e_j)^{(1-\gamma)(1-\sigma)}}{\varphi_j}\right)^{-\frac{1}{\sigma}}}, \quad \text{and} \quad \omega_i^h \equiv \frac{\left(\frac{(e_i)^{1-\gamma(1-\sigma)}}{\varphi_i}\right)^{-\frac{1}{\sigma}}}{\sum_j \pi_j \left(\frac{(e_j)^{1-\gamma(1-\sigma)}}{\varphi_j}\right)^{-\frac{1}{\sigma}}}. \tag{F.16}$$

Hence, we can write aggregate indirect utility $U\left(C_t, N_t; \varphi\right)$ in terms of the aggregates C_t, and N_t

$$U\left(C_t, N_t; \varphi\right) = \Omega \frac{\left(C_t^\gamma \left(1 - N_t\right)^{1-\gamma}\right)^{1-\sigma}}{1-\sigma}, \tag{F.17}$$

where

$$\Omega \equiv \sum_i \pi_i \varphi_i \left(\left(\omega_i^c\right)^\gamma \left(\omega_i^h\right)^{1-\gamma}\right)^{1-\sigma}. \tag{F.18}$$

F.3 Implementability Condition

Using the simple characterization from the previous section we can now derive the implementability condition. Applying the envelope theorem to problem ($F.9$) we get

$$U_C\left(C_t, N_t; \varphi\right) = \theta_t^c, \quad \text{and} \quad U_N\left(C_t, N_t; \varphi\right) = -\theta_t^h.$$

On the other hand, from the first order conditions of the problem ($F.9$) we have

$$\varphi_i u_c\left(c_{i,t}, h_{i,t}\right) = \theta_t^c, \quad \text{and} \quad \varphi_i u_h\left(c_{i,t}, h_{i,t}\right) = -e_i \theta_t^h.$$

It follows that

$$U_C\left(C_t, N_t; \varphi\right) = \varphi_i u_c\left(c_{i,t}, h_{i,t}\right), \tag{F.19}$$

$$U_N\left(C_t, N_t; \varphi\right) = \frac{\varphi_i u_h\left(c_{i,t}, h_{i,t}\right)}{e_i}. \tag{F.20}$$

In any competitive equilibrium these optimality conditions must hold for every agent i. Hence, using ($F.19$), ($F.20$), ($F.4$), and ($F.5$), we obtain

$$\frac{U_N\left(C_t, N_t; \varphi\right)}{U_C\left(C_t, N_t; \varphi\right)} = \frac{u_h\left(c_{i,t}, h_{i,t}\right)}{u_c\left(c_{i,t}, h_{i,t}\right) e_i} = -w_t \frac{\left(1 - \tau_t^h\right)}{\left(1 + \tau^c\right)}, \tag{F.21}$$

and
$$\frac{U_C(C_t,N_t;\varphi)}{U_C(C_0,N_0;\varphi)}=\frac{u_c(c_{i,t},h_{i,t})}{u_c(c_{i,0},h_{i,0})}=\frac{p_t}{\beta^t}. \tag{F.22}$$

Given the relationships above we can derive the implementation condition which relies only on the aggregates C_t, N_t and market weights φ. Let $c^m_{i,t}(C_t,N_t;\varphi)$ and $h^m_{i,t}(C_t,N_t;\varphi)$ be the arg max of problem ($F.9$) given by the ($F.14$) and ($F.15$) respectively. The budget constraint of agent i implies

$$\sum_{t=0}^{\infty}p_t\left(c^m_{i,t}(C_t,N_t;\varphi)-\frac{(1-\tau^h_t)}{(1+\tau^c)}w_te_ih^m_{i,t}(C_t,N_t;\varphi)\right)\leq\frac{R_0a_{i,0}+T}{(1+\tau^c)},$$

which using ($F.21$) and ($F.21$) can be restated as

$$\sum_{t=0}^{\infty}\beta^t\left(U_C(C_t,N_t;\varphi)c^m_{i,t}(C_t,N_t;\varphi)+U_N(C_t,N_t;\varphi)e_ih^m_{i,t}(C_t,N_t;\varphi)\right)\leq U_C(C_0,N_0;\varphi)\left(\frac{R_0a_{i,0}+T}{1+\tau^c}\right),\quad\forall\, i. \tag{F.23}$$

The following Proposition follows immediately from the arguments above.

Proposition 6 *An aggregate allocation $\{C_t,N_t,K_{t+1}\}_{t=0}^{\infty}$ can be supported by a competitive equilibrium if and only if the resource constraints ($F.2$) hold and there exist market weights φ and a lump-sum tax T such that the implementability conditions ($F.23$) hold for all $i\in I$. Individual allocations can then be computed using functions $c^m_{i,t}$ and $h^m_{i,t}$, prices and taxes can be computed using the usual equilibrium conditions.*

In the Ramsey problem considered in this paper we restrict the capital income tax to be less than or equal to one. The following Lemma is useful to define our Ramsey problem taking this restriction into consideration.

Lemma 5 *In any competitive equilibrium $\tau^k_{t+1}\leq 1$ if and only if $U_C(C_t,N_t;\varphi)\geq\beta U_C(C_{t+1},N_{t+1};\varphi)$.*

Proof. $(\Rightarrow)$: Take any competitive equilibrium with $\tau^k_{t+1}\leq 1$. Then, the first order conditions for agent i imply

$$[c_{i,t}]:\ \frac{\beta^t}{p_t}u_c(c_{i,t},h_{i,t})=\phi,$$
$$[c_{i,t+1}]:\ \frac{\beta^{t+1}}{p_{t+1}}u_c(c_{i,t+1},h_{i,t+1})=\phi,$$

and, thus,
$$u_c(c_{i,t},h_{i,t})=\beta\frac{p_t}{p_{t+1}}u_c(c_{i,t+1},h_{i,t+1}).$$

Since $R_{t+1}\equiv 1+\left(1-\tau^k_{t+1}\right)(r_{t+1}-\delta)$, for $t\geq 0$ and $p_t=R_{t+1}p_{t+1}$ we get

$$u^i_c(c_{i,t},h_{i,t})=\beta\left(1+\left(1-\tau^k_{t+1}\right)(r_{t+1}-\delta)\right)u^i_c(c_{i,t+1},h_{i,t+1}).$$

Further, using $U_C(C_t, N_t; \varphi) = \varphi_i u_c(c_{i,t}, h_{i,t})$ from $(F.19)$, we get

$$U_C(C_t, N_t; \varphi) = \beta \left(1 + \left(1 - \tau_{t+1}^k\right)(r_{t+1} - \delta)\right) U_C(C_{t+1}, N_{t+1}; \varphi),$$

and since $\tau_{t+1}^k \leq 1$ and hence $\left(1 - \tau_{t+1}^k\right) > 0$

$$U_C(C_t, N_t; \varphi) = \beta \left(1 + \left(1 - \tau_{t+1}^k\right)(r_{t+1} - \delta)\right) U_C(C_{t+1}, N_{t+1}; \varphi) \geqslant \beta U_C(C_{t+1}, N_{t+1}; \varphi).$$

which completes the first part of the proof.
$(\Leftarrow)$: Suppose that $U_C(C_t, N_t; \varphi) \geq \beta U_C(C_{t+1}, N_{t+1}; \varphi)$. In any competitive equilibrium we have

$$U_C(C_t, N_t; \varphi) = \beta \left(1 + \left(1 - \tau_{t+1}^k\right)(r_{t+1} - \delta)\right) U_C(C_{t+1}, N_{t+1}; \varphi),$$

thus, by $U_C(C_t, N_t; \varphi) \geq \beta U_C(C_{t+1}, N_{t+1}; \varphi)$,

$$\beta \left(1 + \left(1 - \tau_{t+1}^k\right)(r_{t+1} - \delta)\right) U_C(C_{t+1}, N_{t+1}; \varphi) \geq \beta U_C(C_{t+1}, N_{t+1}; \varphi),$$

implying that

$$1 + \left(1 - \tau_{t+1}^k\right)(r_{t+1} - \delta) \geq 1,$$

and hence

$$\tau_{t+1}^k \leq 1.$$

■

F.4 Ramsey Problem

Let $\lambda \equiv \{\lambda_i\}$ be the planner's welfare weight on type i, with $\sum_i \pi_i \lambda_i = 1$, the Ramsey planner problem is

$$\max_{\{C_t, N_t, K_{t+1}\}_{t=0}^{\infty}, \tau_0^k, T, \varphi} \sum_{t,i} \beta^t \lambda_i \pi_i u\left(c_{i,t}^m(C_t, N_t; \varphi), h_{i,t}^m(C_t, N_t; \varphi)\right),$$

subject to

$$\sum_{t=0}^{\infty} \beta^t \left(U_C(C_t, N_t; \varphi) c_{i,t}^m(C_t, N_t; \varphi) + U_N(C_t, N_t; \varphi) e_i h_{i,t}^m(C_t, N_t; \varphi)\right) \leq U_C(C_0, N_0; \varphi) \left(\frac{R_0 a_{i,0} + T}{1 + \tau^c}\right), \quad \forall\ i,$$

$$C_t + G_t + K_{t+1} = F(K_t, N_t) + (1 - \delta) K_t, \quad \forall\ t \geq 0,$$

$$U_C(C_t, N_t; \varphi) \geq \beta U_C(C_{t+1}, N_{t+1}; \varphi), \quad \forall\ t \geq 0.$$

Define

$$W\left(C_t,N_t;\varphi,\mu,\lambda\right)\equiv\sum_i\lambda_i\pi_i u\left(c^m_{i,t}\left(C_t,N_t;\varphi\right),h^m_{i,t}\left(C_t,N_t;\varphi\right)\right)$$
$$+\sum_i\pi_i\mu_i\left[U_C\left(C_t,N_t;\varphi\right)c^m_{i,t}\left(C_t,N_t;\varphi\right)+U_N\left(C_t,N_t;\varphi\right)e_i h^m_{i,t}\left(C_t,N_t;\varphi\right)\right],$$

where μ_i is the Lagrange multiplier on the implementability constraint of agent i, and $\mu\equiv\{\mu_i\}$. Rewrite the Ramsey problem as

$$\max_{\{C_t,N_t,K_{t+1}\}_{t=0}^{\infty},T,\varphi,\tau_0^k\leq 1}\sum_{t,i}\beta^t W\left(C_t,N_t;\varphi,\mu,\lambda\right)-U_C\left(C_0,N_0;\varphi\right)\sum_i\pi_i\mu_i\left(\frac{R_0a_{i,0}+T}{1+\tau^c}\right),$$

subject to

$$C_t+G_t+K_{t+1}=F\left(K_t,N_t\right)+\left(1-\delta\right)K_t,\quad\forall\ t\geq 0,$$
$$U_C\left(C_t,N_t;\varphi\right)\geq\beta U_C\left(C_{t+1},N_{t+1};\varphi\right),\quad\forall\ t\geq 0,$$

where $\beta^t\nu_t$ and $\beta^t\eta_t$ are the Lagrange multipliers on the feasibility, and $\tau_t^k\leq 1$ constraint respectively.

F.4.1 Initial capital taxes

The first order condition with respect to τ_0^k is given by

$$U_C\left(C_0,N_0;\varphi\right)\frac{\left(F_K\left(K_0,N_0\right)-\delta\right)}{1+\tau^c}\sum_i\pi_i\mu_i a_{i,0}-\kappa=0,$$

where κ is the multiplier on $\tau_0^k\leq 1$. So, if $\sum_i\pi_i\mu_i a_{i,0}=0$, τ_0^k is indeterminate since it is equivalent to a lump-sum tax. On the other hand, if $\sum_i\pi_i\mu_i a_{i,0}>0$, then a higher τ_0^k reduces inequality while being undistortive, so it is optimal to set $\tau_0^k=1$.

F.4.2 From first order conditions to taxes

Define $R_t^*\equiv 1+r_t-\delta$ and

$$\eta_{-1}\equiv\frac{R_0}{\left(1+\tau^c\right)}\sum_i\pi_i\mu_i a_{i,0}.$$

The first order conditions are[8]

$$[C_t]:W_C\left(C_t,N_t;\varphi,\mu,\lambda\right)-\nu_t+U_{CC}\left(C_t,N_t;\varphi\right)\left(\eta_t-\eta_{t-1}\right)=0,\quad\forall\ t\geq 0,\tag{F.24}$$

[8]The term

$$-\left(\left(1-\tau_0^k\right)F_{KN}\left(K_0,N_0\right)\right)U_C\left(C_0,N_0;\varphi\right)\left(\frac{\sum_i\pi_i\mu_i a_{i,0}}{1+\tau^c}\right)$$

in the derivative with respect to N_0 is equal to zero since, either $\sum_i\pi_i\mu_i a_{i,0}=0$ or $\tau_0^k=1$.

$$[N_t] : W_N\left(C_t, N_t; \varphi, \mu, \lambda\right) + \nu_t F_N\left(K_t, N_t\right) + U_{CN}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right) = 0, \quad \forall\, t \geq 0, \tag{F.25}$$

$$[K_{t+1}] : -\nu_t + \left[F_K\left(K_{t+1}, N_{t+1}\right) + (1-\delta)\right]\beta\nu_{t+1} = 0, \quad \forall\, t \geq 0, \tag{F.26}$$

$$[T] : \sum_i \pi_i \mu_i = 0. \tag{F.27}$$

From ($F.25$) and ($F.24$) we obtain

$$F_N\left(K_t, N_t\right) = -\frac{W_N\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CN}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}{W_C\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CC}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}, \quad \forall\, t \geq 0, \tag{F.28}$$

and using the intertemporal condition ($F.26$) we get

$$R^*_{t+1} = \frac{1}{\beta}\frac{W_C\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CC}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}{W_C\left(C_{t+1}, N_{t+1}; \varphi, \mu, \lambda\right) + U_{CC}\left(C_{t+1}, N_{t+1}; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}, \quad \forall\, t \geq 0, \tag{F.29}$$

These two equations can be used to back out the optimal taxes on labor and capital income.

Plugging ($F.28$) into ($F.21$) implies

$$\frac{U_N\left(C_t, N_t; \varphi\right)}{U_C\left(C_t, N_t; \varphi\right)} = \frac{W_N\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CN}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}{W_C\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CC}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}\frac{\left(1-\tau^h_t\right)}{\left(1+\tau^c\right)},$$

which can be rearranged into

$$\frac{\tau^h_t + \tau^c}{1+\tau^c} = 1 - \frac{U_N\left(C_t, N_t; \varphi\right)}{U_C\left(C_t, N_t; \varphi\right)}\frac{W_C\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CC}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}{W_N\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CN}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}. \tag{F.30}$$

In any competitive equilibrium ($F.22$) holds, which together with $p_t = R_t p_{t+1}$implies

$$\frac{U_C\left(C_{t+1}, N_{t+1}; \varphi\right)}{U_C\left(C_t, N_t; \varphi\right)}\beta R_{t+1} = 1.$$

Substituting this into ($F.29$), it follows that

$$\frac{R_{t+1}}{R^*_{t+1}} = \frac{W_C\left(C_{t+1}, N_{t+1}; \varphi, \mu, \lambda\right) + U_{CC}\left(C_{t+1}, N_{t+1}; \varphi\right)\left(\eta_{t+1} - \eta_t\right)}{W_C\left(C_t, N_t; \varphi, \mu, \lambda\right) + U_{CC}\left(C_t, N_t; \varphi\right)\left(\eta_t - \eta_{t-1}\right)}\frac{U_C\left(C_t, N_t; \varphi\right)}{U_C\left(C_{t+1}, N_{t+1}; \varphi\right)}. \tag{F.31}$$

F.4.3 Explicit formulas for U and its derivatives

From ($F.17$), it follows that

$$\begin{aligned}
U_C\left(C_t, N_t; \varphi\right) &= \gamma\Omega\left(C_t^\gamma\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma} C_t^{-1}, \\
U_N\left(C_t, N_t; \varphi\right) &= -\left(1-\gamma\right)\Omega\left(C_t^\gamma\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma}\left(1-N_t\right)^{-1}, \\
U_{CC}\left(C_t, N_t; \varphi\right) &= -\left(1+\gamma\left(\sigma-1\right)\right)\gamma\Omega\left(C_t^\gamma\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma} C_t^{-2}, \\
U_{NN}\left(C_t, N_t; \varphi\right) &= -\left(1-\gamma\right)\left(1-\left(1-\sigma\right)\left(1-\gamma\right)\right)\Omega\left(C_t^\gamma\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma}\left(1-N_t\right)^{-2},
\end{aligned}$$

$$U_{CN}\left(C_t, N_t; \varphi\right) = -\gamma\left(1-\sigma\right)\left(1-\gamma\right)\Omega\left(C_t^{\gamma}\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma} C_t^{-1}\left(1-N_t\right)^{-1}.$$

F.4.4 Explicit formulas for W and its derivatives

It follows from the derivatives of U and equations ($F.15$) and ($F.14$) that

$$\begin{aligned} U_C\left(C_t, N_t; \varphi\right) c_{i,t}\left(C_t, N_t; \varphi\right) &= \gamma\Omega\left(C_t^{\gamma}\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma}\omega_i^c, \\ U_N\left(C_t, N_t; \varphi\right) e_i h_{i,t}\left(C_t, N_t; \varphi\right) &= -\left(1-\gamma\right)\Omega\left(C_t^{\gamma}\left(1-N_t\right)^{1-\gamma}\right)^{-\sigma} C_t^{\gamma}\left(1-N_t\right)^{-\gamma} e_i\left(1-\omega_i^h\left(1-N_t\right)\right). \end{aligned} \tag{F.32}$$

Substituting these into the definition of $W\left(C_t, N_t; \varphi, \mu, \lambda\right)$ we get

$$\begin{aligned} W\left(C_t, N_t; \varphi, \mu, \lambda\right) = &\sum_i \pi_i\varphi_i\left(\left(\omega_i^c\right)^{\gamma}\left(\omega_i^h\right)^{1-\gamma}\right)^{1-\sigma}\frac{\left(C_t^{\gamma}\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma}}{1-\sigma} + \sum_i \pi_i\mu_i\Big[\omega_i^c\gamma\Omega\left(C_t^{\gamma}\left(1-N_t\right)^{1-\gamma}\right)^{1-\sigma} - \\ &\left(1-\gamma\right)\Omega\left(C_t^{\gamma}\left(1-N_t\right)^{1-\gamma}\right)^{-\sigma} C_t^{\gamma}\left(1-N_t\right)^{-\gamma} e_i\left(1-\omega_i^h\left(1-N_t\right)\right)\Big]. \end{aligned} \tag{F.33}$$

Collecting terms and simplifying we obtain

$$W\left(C_t, N_t; \varphi, \mu, \lambda\right) = \Phi U\left(C_t, N_t; \varphi\right) + \Psi U_N\left(C_t, N_t; \varphi\right), \tag{F.34}$$

where

$$\Phi \equiv 1 + \left(1-\sigma\right)\sum_i \pi_i\mu_i\left(\gamma\omega_i^c + \left(1-\gamma\right)e_i\omega_i^h\right), \tag{F.35}$$

$$\Psi \equiv \sum_i \pi_i\mu_i e_i. \tag{F.36}$$

Also, to be used later, define

$$\Theta = \frac{\Phi}{\Psi}. \tag{F.37}$$

Taking derivatives we obtain

$$\begin{aligned} W_C\left(C_t, N_t; \varphi, \mu, \lambda\right) &= \Phi U_C\left(C_t, N_t; \varphi\right) + \Psi U_{CN}\left(C_t, N_t; \varphi\right), \\ W_N\left(C_t, N_t; \varphi, \mu, \lambda\right) &= \Phi U_N\left(C_t, N_t; \varphi\right) + \Psi U_{NN}\left(C_t, N_t; \varphi\right). \end{aligned}$$

F.4.5 Substituting the derivatives into the tax formulas

Substituting the derivatives into equation (F.30) we get

$$\frac{\tau_t^h + \tau^c}{1+\tau^c} = \frac{\Psi\left(1-N_t\right)^{-1} + C_t^{-1}\left(\eta_t - \eta_{t-1}\right)}{\Phi + \left(1-\left(1-\sigma\right)\left(1-\gamma\right)\right)\Psi\left(1-N_t\right)^{-1} + \gamma\left(1-\sigma\right)C_t^{-1}\left(\eta_t - \eta_{t-1}\right)}. \tag{F.38}$$

And substituting the derivatives into (F.31) yields

$$\frac{R_{t+1}}{R^*_{t+1}} = \frac{\Phi - (1-\sigma)(1-\gamma)\Psi(1-N_{t+1})^{-1} - (1+\gamma(\sigma-1))C_{t+1}^{-1}(\eta_{t+1}-\eta_t)}{\Phi - (1-\sigma)(1-\gamma)\Psi(1-N_t)^{-1} - (1+\gamma(\sigma-1))C_t^{-1}(\eta_t-\eta_{t-1})}. \tag{F.39}$$

In what follows, we proceed in steps: we first consider an economy with only asset heterogeneity, then an economy with only productivity heterogeneity, and, finally the economy with both types of heterogeneity.

F.5 Only Asset Heterogeneity

Lemma 6 *If $e_i = \bar{e}$ for all $i \in I$, then $\Psi = 0$ and $\Phi > 0$.*

Proof. If $e_i = \bar{e}$ for all $i \in I$, then it follows from the definition of Ψ that

$$\Psi = \sum_i \pi_i \mu_i e_i = \bar{e} \sum_i \pi_i \mu_i = 0,$$

where the last equality follows from equation (*F.27*). Next, notice that

$$u\left(c^m_{i,t}(C_t, N_t; \varphi), h^m_{i,t}(C_t, N_t; \varphi)\right) = \left((\omega^c_i)^\gamma \left(\omega^h_i\right)^{1-\gamma}\right)^{1-\sigma} \frac{\left(C_t^\gamma (1-N_t)^{1-\gamma}\right)^{1-\sigma}}{1-\sigma},$$

and, therefore, the solution to the problem must satisfy $C_t^\gamma (1-N_t)^{1-\gamma} > 0$ for any finite $t \geq 0$. On the other hand, since $\Psi = 0$, it follows from equation (*F.34*) that

$$W(C_t, N_t; \varphi, \mu, \lambda) = \Phi\Omega \frac{\left(C_t^\gamma (1-N_t)^{1-\gamma}\right)^{1-\sigma}}{1-\sigma}.$$

Fix some finite t, if $\Phi \leq 0$ then reducing $C_t^\gamma (1-N_t)^{1-\gamma}$ to 0 is feasible and, since $\Omega > 0$, would weakly increase welfare which is a contradiction. ■

Proposition 7 *There exists a finite $t^* \geq 0$ such that the optimal tax system is given by $\tau^k_t = 1$ for $0 \leq t < t^*$ and $\tau^k_t = 0$ for all $t > t^*$; and , τ^h_t follows*

$$\frac{R_{t+1}}{R^*_{t+1}} = \frac{1-\tau^h_{t+1}}{1-\tau^h_t} \frac{1+\tau^c+\gamma(\sigma-1)\left(\tau^c+\tau^h_t\right)}{1+\tau^c+\gamma(\sigma-1)\left(\tau^c+\tau^h_{t+1}\right)}, \tag{F.40}$$

for $0 \leq t \leq t^*$, and $\tau^h_t = -\tau^c$ for all $t > t^*$.

Proof. Suppose $\eta_t = 0$ for all $t \geq 0$. Evaluating (*F.39*) for period 0 we get

$$\frac{R_1}{R^*_1} = \frac{\Phi - (1-\sigma)(1-\gamma)\Psi(1-N_1)^{-1}}{\Phi - (1-\sigma)(1-\gamma)\Psi(1-N_0)^{-1} + (1+\gamma(\sigma-1))C_0^{-1}\eta_{-1}},$$

which, since $\Psi = 0$ (from Lemma 6), implies that

$$\widetilde{\Phi} C_0 \frac{R^*_1}{R_1} = \widetilde{\Phi} C_0 + \eta_{-1},$$

where

$$\widetilde{\Phi} \equiv \frac{\Phi}{1+\gamma\left(\sigma-1\right)}.$$

It follows that, if

$$\eta_{-1} < \widetilde{\Phi} C_0 \left(R_1^* - 1\right),$$

then,

$$\widetilde{\Phi} C_0 \frac{R_1^*}{R_1} < \widetilde{\Phi} C_0 R_1^* \;\Rightarrow\; R_1 > 1 \;\Rightarrow\; \tau_1^k < 1.$$

Moreover, if

$$\eta_{-1} > -\widetilde{\Phi} C_0,$$

then,

$$\widetilde{\Phi} C_0 \frac{R_1^*}{R_1} > 0 \;\Rightarrow\; R_1 \text{ is finite} \;\Rightarrow\; \tau_1^k \text{ is finite.}$$

Hence, in particular,

$$-\widetilde{\Phi} C_0 \equiv P_1 < \eta_{-1} < M_1 \equiv \widetilde{\Phi} C_0 \left(R_1^* - 1\right) \;\Rightarrow\; \tau_1^k < 1 \text{ and } \tau_1^k \text{ is finite.}$$

For $t \geq 1$, since $\eta_t = 0$ for all $t \geq 0$ and $\Psi = 0$, it follows from ($F.39$) that

$$\frac{R_{t+1}}{R_{t+1}^*} = 1,$$

hence, $\tau_t^k = 0$ for all $t \geq 2$. Thus, if $P_1 < \eta_{-1} < M_1$, then the upper bound constraint on the τ_t^k is never binding and, in fact, $\eta_t = 0$ for all $t \geq 0$. Now, pick some finite t^* and suppose $\eta_t > 0$ for $t \leq t^* - 2$ and $\eta_t = 0$ for all $t \geq t^* - 1$. Then, evaluating ($F.39$) at $t = t^* - 1$ gives

$$\widetilde{\Phi} C_{t^*-1} \frac{R_{t^*}^*}{R_{t^*}} = \widetilde{\Phi} C_{t^*-1} + \eta_{t^*-2},$$

and, analogously to above, it follows that

$$-\widetilde{\Phi} C_{t^*-1} < \eta_{t^*-2} < \widetilde{\Phi} C_{t^*-1} \left(R_{t^*}^* - 1\right) \;\Rightarrow\; \tau_{t^*}^k < 1 \text{ and } \tau_{t^*}^k \text{ is finite.}$$

For $t \leq t^* - 2$, ($F.39$) evaluated with $\tau_t^k = 1$ yields

$$\frac{1}{R_{t+1}^*} = \frac{\widetilde{\Phi} - C_{t+1}^{-1}\left(\eta_{t+1} - \eta_t\right)}{\widetilde{\Phi} - C_t^{-1}\left(\eta_t - \eta_{t-1}\right)}, \quad 0 \leq t \leq t^* - 2.$$

Let

$$X_t \equiv \eta_{t-1} - \eta_t,$$

then

$$X_t = \widetilde{\Phi} C_t \left(R_{t+1}^* - 1\right) + \frac{C_t}{C_{t+1}} R_{t+1}^* X_{t+1}, \quad 0 \leq t \leq t^* - 2,$$

which is a first-order difference equation on X_t with terminal condition

$$X_{t^*-1} = \eta_{t^*-2}.$$

Denote

$$A_t \equiv \widetilde{\Phi} C_t \left(R^*_{t+1} - 1\right), \quad \text{and} \quad B_t \equiv \frac{C_t}{C_{t+1}} R^*_{t+1},$$

so that we have the following dynamic system

$$\begin{aligned} X_t &= A_t + B_t X_{t+1}, \quad 0 \leq t \leq t^* - 2, \\ \eta_{t-1} &= \eta_t + X_t, \quad 0 \leq t \leq t^* - 1, \end{aligned}$$

where the second equation comes from the definition of X_t. In what follows the idea is to transform the bounds imposed on η_{t^*-2} on bounds on η_{-1}. Iterate on both equations to get

$$\eta_{-1} = \eta_{t^*-1} + \sum_{\tau=1}^{t^*} X_{\tau-1},$$

and

$$\sum_{\tau=1}^{t^*} X_{\tau-1} = \sum_{\tau=1}^{t^*} \left(\sum_{s=\tau-1}^{t^*-2} \left(\prod_{j=\tau-1}^{s-1} B_j \right) A_s + \left(\prod_{j=\tau-1}^{t^*-2} B_j \right) X_{t^*-1} \right).$$

Hence, it follows that

$$\eta_{-1} = \eta_{t^*-1} + \sum_{\tau=1}^{t^*} \left(\sum_{s=\tau-1}^{t^*-2} \left(\prod_{j=\tau-1}^{s-1} B_j \right) A_s + \left(\prod_{j=\tau-1}^{t^*-2} B_j \right) X_{t^*-1} \right). \tag{F.41}$$

Now using the definitions of A_t and B_t, the terminal condition, and the fact that $\eta_{t^*-1} = 0$ we obtain

$$\eta_{-1} = \sum_{\tau=1}^{t^*} \sum_{s=\tau-1}^{t^*-2} \left(\prod_{j=\tau-1}^{s-1} \frac{C_j}{C_{j+1}} R^*_{j+1} \right) \widetilde{\Phi} C_s \left(R^*_{s+1} - 1\right) + \sum_{\tau=1}^{t^*} \left(\prod_{j=\tau-1}^{t^*-2} \frac{C_j}{C_{j+1}} R^*_{j+1} \right) \eta_{t^*-2},$$

which relates the η_{-1} and η_{t^*-2}. Hence, imposing the upper bound

$$\eta_{t^*-2} < \widetilde{\Phi} C_{t^*-1} \left(R^*_{t^*} - 1\right),$$

is equivalent to imposing

$$\begin{aligned} \eta_{-1} &< \sum_{\tau=1}^{t^*} \sum_{s=\tau-1}^{t^*-2} \left(\prod_{j=\tau-1}^{s-1} \frac{C_j}{C_{j+1}} R^*_{j+1} \right) \widetilde{\Phi} C_s \left(R^*_{s+1} - 1\right) + \sum_{\tau=1}^{t^*} \left(\prod_{j=\tau-1}^{t^*-2} \frac{C_j}{C_{j+1}} R^*_{j+1} \right) \widetilde{\Phi} C_{t^*-1} \left(R^*_{t^*} - 1\right) \\ &= \sum_{\tau=1}^{t^*} \left(\prod_{j=\tau}^{t^*} R^*_j - 1 \right) \widetilde{\Phi} C_{\tau-1}, \end{aligned}$$

and the lower bound

$$\eta_{t^*-2} > -\widetilde{\Phi} C_{t^*-1}$$

is equivalent to

$$\eta_{-1} > \sum_{\tau=1}^{t^*} \sum_{s=\tau-1}^{t^*-2} \left(\prod_{j=\tau-1}^{s-1} \frac{C_j}{C_{j+1}} R_{j+1}^* \right) \widetilde{\Phi} C_s \left(R_{s+1}^* - 1 \right) - \sum_{\tau=1}^{t^*} \left(\prod_{j=\tau-1}^{t^*-2} \frac{C_j}{C_{j+1}} R_{j+1}^* \right) \widetilde{\Phi} C_{t^*-1}$$
$$= -\sum_{\tau=1}^{t^*} \widetilde{\Phi} C_{\tau-1}.$$

Therefore, we obtain that

$$-\sum_{\tau=1}^{t^*} \widetilde{\Phi} C_{\tau-1} \equiv P_{t^*} < \eta_{-1} < M_{t^*} \equiv \sum_{\tau=1}^{t^*} \left(\prod_{j=\tau}^{t^*} R_j^* - 1 \right) \widetilde{\Phi} C_{\tau-1} \;\Rightarrow\; \tau_{t^*}^k < 1 \text{ and } \tau_{t^*}^k \text{ is finite.}$$

For $t \geq t^*$, the fact that $\eta_t = 0$ for all $t \geq t^* - 1$ and $\Psi = 0$ implies $\tau_t^k = 0$ for all $t \geq t^* + 1$. Thus, if $P_{t^*} < \eta_{-1} < M_{t^*}$, then the upper bound constraint on τ_t^k is binding only for $t \leq t^* - 1$ and, in fact, $\eta_t > 0$ for $t \leq t^* - 2$ and $\eta_t = 0$ for all $t \geq t^* - 1$.

Finally, notice that

$$\lim_{t \longrightarrow \infty} P_t = -\infty, \quad \text{and} \quad \lim_{t \longrightarrow \infty} M_t = \infty,$$

and since the η_{-1} is finite the result for τ_t^k follows. To establish the result for τ_t^h, first notice that equation ($F.38$), using the fact that $\Psi = 0$, we have

$$\tau_t^h = -\tau^c + \frac{(1+\tau^c) C_t^{-1} (\eta_t - \eta_{t-1})}{\Phi + \gamma (1-\sigma) C_t^{-1} (\eta_t - \eta_{t-1})}.$$

Thus, since $\eta_t = 0$ for all $t \geq t^* - 1$, we have that

$$\tau_t^h = -\tau_c, \quad \text{for } t \geq t^*.$$

To show that equation ($F.40$) is satisfied for $t \leq t^* - 1$, first solve for $C_t^{-1} (\eta_t - \eta_{t-1})$ from equation ($F.38$),

$$C_t^{-1} (\eta_t - \eta_{t-1}) = \frac{-\Phi}{\gamma (1-\sigma) - \frac{1+\tau^c}{\tau_t^h + \tau^c}},$$

and substitute it into equation ($F.39$). ■

F.6 Only Productivity Heterogeneity

Proposition 8 *Assuming capital income taxes are bounded only by the positivity of gross interest rates, the optimal labor income tax, τ_t^h, can be written as a function of N_t given by*

$$\tau_t^h (N_t) = \frac{1+\tau^c}{(1-N_t)\Theta + \gamma + \sigma (1-\gamma)} - \tau^c, \quad \textit{for} \;\; t \geq 0, \tag{F.42}$$

with sensitivity

$$N_t \frac{d\tau_t^h(N_t)}{dN_t} = \frac{N_t}{1-N_t}\left(\tau_t^h+\tau^c\right)\left(1-(\gamma+\sigma(1-\gamma))\frac{\left(\tau_t^h+\tau^c\right)}{(1+\tau^c)}\right), \quad \text{for } t \geq 0. \tag{F.43}$$

It is optimal to set the capital-income tax rate according to

$$\frac{R_{t+1}}{R_{t+1}^*} = \frac{1-N_t}{1-N_{t+1}} \frac{1-\tau_{t+1}^h}{1-\tau_t^h} \frac{\tau_t^h+\tau^c}{\tau_{t+1}^h+\tau^c}, \quad \text{for } t \geq 0. \tag{F.44}$$

Proof. In this economy there is no heterogeneity in initial levels of assets, i.e. $a_{i,0} = a_0$ for all $i \in I$. It follows that

$$\eta_{-1} = \frac{R_0}{(1+\tau^c)} \sum_i \pi_i \mu_i a_{i,0} = \frac{R_0}{(1+\tau^c)} a_0 \sum_i \pi_i \mu_i = 0,$$

where the last equality follows from ($F.27$). Since, we do not impose an upper bound on the capital income tax the constraint $U_C(C_t, N_t; \varphi) \geq \beta U_C(C_{t+1}, N_{t+1}; \varphi)$ is dropped and ($F.38$) becomes

$$\frac{\tau_t^h+\tau^c}{1+\tau^c} = \frac{\Psi(1-N_t)^{-1}}{\Phi+(1-(1-\sigma)(1-\gamma))\Psi(1-N_t)^{-1}}. \tag{F.45}$$

Rearranging we get ($F.42$). The sensitivity of the labor income tax is derived by differentiating the formula above with respect to N_t, i.e.

$$\frac{d\tau_t^h(N_t)}{dN_t} = \frac{(1+\tau^c)\frac{\Phi}{\Psi}}{\left((1-N_t)\frac{\Phi}{\Psi}+\gamma+\sigma(1-\gamma)\right)^2}.$$

Then, substituting Φ/Ψ from ($F.42$) , which yields

$$\frac{\Phi}{\Psi} = \frac{1}{(1-N_t)}\left(\frac{1}{\frac{\tau_t^h+\tau^c}{1+\tau^c}}-(\gamma+\sigma(1-\gamma))\right),$$

implies ($F.43$). In the absence of the upper bound on the capital income tax, equation ($F.39$) becomes

$$\frac{R_{t+1}}{R_{t+1}^*} = \frac{\frac{\Phi}{\Psi}-(1-\sigma)(1-\gamma)(1-N_{t+1})^{-1}}{\frac{\Phi}{\Psi}-(1-\sigma)(1-\gamma)(1-N_t)^{-1}},$$

which, again substituting Φ/Ψ, implies equation ($F.44$) completing the proof. ■

F.7 Asset and Productivity Heterogeneity

We start by establishing the following lemma, using an analogous argument to the one used in Lemma 6.

Lemma 7 *It must be that* $\Phi-(1-\sigma)(1-\gamma)\Psi(1-N_t)^{-1} > 0$.

Proof. Using the fact that

$$U_N(C_t, N_t; \varphi) = -(1-\sigma)(1-\gamma)(1-N_t)^{-1} U(C_t, N_t; \varphi),$$

it follows from equation ($F.34$) that

$$W\left(C_t, N_t; \varphi, \mu, \lambda\right) = [\Phi - (1-\sigma)(1-\gamma)(1-N_t)^{-1}\Psi]\Omega\frac{\left(C_t^{\gamma}(1-N_t)^{1-\gamma}\right)^{1-\sigma}}{1-\sigma}..$$

Fix some finite t, if $\Phi - (1-\sigma)(1-\gamma)\Psi(1-N_t)^{-1} \leq 0$ then reducing $C_t^{\gamma}(1-N_t)^{1-\gamma}$ to 0 is feasible and, since $\Omega > 0$, would weakly increase welfare which is a contradiction. ■

Finally, we can prove the main result of this section, using a combination of the arguments used to prove the last two propositions.

Proposition 9 *There exists a finite integer $t^* \geq 0$ such that the optimal tax system is given by $\tau_t^k = 1$, for $0 \leq t < t^*$, τ_t^k follows equation*

$$\frac{R_{t+1}}{R_{t+1}^*} = \frac{1-N_t}{1-N_{t+1}}\frac{1-\tau_{t+1}^h}{1-\tau_t^h}\frac{\tau_t^h+\tau^c}{\tau_{t+1}^h+\tau^c}, \tag{F.46}$$

for $t > t^*$, τ_t^h evolves according to

$$\frac{R_{t+1}}{R_{t+1}^*} = \frac{\Theta + \sigma(1-N_{t+1})^{-1}}{\Theta + \sigma(1-N_t)^{-1}}\frac{1-\tau_{t+1}^h}{1-\tau_t^h}\frac{1+\tau^c+\gamma(\sigma-1)\left(\tau^c+\tau_t^h\right)}{1+\tau^c+\gamma(\sigma-1)\left(\tau^c+\tau_{t+1}^h\right)}, \tag{F.47}$$

for $0 \leq t \leq t^*$, and τ_t^h is determined by

$$\tau_t^h(N_t) = \frac{(1+\tau^c)}{(1-N_t)\Theta + \gamma + \sigma(1-\gamma)} - \tau^c, \tag{F.48}$$

for all $t > t^*$.

Proof. The existence of t^* such that $\eta_t > 0$, for $t \leq t^* - 2$ and $\eta_t = 0$, for all $t \geq t^* - 1$, can be shown using an argument analogous to the one used in the proof of Proposition 7, the only difference being that Ψ is no longer equal to 0. The corresponding P_{t^*} and M_{t^*} are given by

$$P_{t^*} \equiv -\sum_{\tau=1}^{t^*}\widehat{\Phi}_{\tau-1}C_{\tau-1}, \quad \text{and} \quad M_{t^*} \equiv \sum_{\tau=1}^{t^*}\left(\widehat{\Phi}_{t^*}\prod_{j=\tau}^{t^*}R_j^* - \widehat{\Phi}_1\right)C_{\tau-1},$$

where

$$\widehat{\Phi}_t \equiv \frac{\Phi - (1-\sigma)(1-\gamma)\Psi(1-N_t)^{-1}}{1+\gamma(\sigma-1)}.$$

The numerator is strictly positive as a result of Lemma 7. Since $\eta_t = 0$, for all $t \geq t^* - 1$, the same argument used in Proposition 8 implies that equations ($F.46$) and ($F.48$) must be satisfied, for $t \geq t^*$. To show that equation ($F.47$) is satisfied for $t \leq t^*$, again analogously to the proof of Proposition 7, first solve for $C_t^{-1}(\eta_t - \eta_{t-1})$ from equation ($F.38$),

$$C_t^{-1}(\eta_t - \eta_{t-1}) = \frac{\left(\frac{1+\tau^c}{\tau_t^h+\tau^c} - (\sigma + \gamma(1-\sigma))\right)\Psi(1-N_t)^{-1} - \Phi}{\gamma(1-\sigma) - \frac{1+\tau^c}{\tau_t^h+\tau^c}},$$

and substitute it into equation ($F.39$). ■

F.8 Relationship with Straub and Werning (2020)

The reason why capital income taxes should converge to zero, that is, the reason why the point made by Straub and Werning (2020) does not directly apply to our environment is because the planner can use lump-sum taxes. When lump-sum taxes are not available, the planner might choose to tax capital income because it needs to obtain revenue and it is less distortive than other instruments. When it is available, since lump-sum taxes are not distortive, capital income taxes are only chosen by the planner in order to provide redistribution; lump-sum taxes are always a more efficient alternative to obtain revenue.

To see this more clearly, consider, for simplicity, the environment above without heterogeneity and without lump-sum transfers. First notice that, in this case, it follows that $\mu_i = \mu$ for all i and from (F.16), (F.35) and (F.36) that

$$\omega_i^c = \frac{\left(\frac{(e_i)^{(1-\gamma)(1-\sigma)}}{\varphi_i}\right)^{-\frac{1}{\sigma}}}{\sum_j \pi_j \left(\frac{(e_j)^{(1-\gamma)(1-\sigma)}}{\varphi_j}\right)^{-\frac{1}{\sigma}}} = 1, \qquad \omega_i^h = \frac{\left(\frac{(e_i)^{1-\gamma(1-\sigma)}}{\varphi_i}\right)^{-\frac{1}{\sigma}}}{\sum_j \pi_j \left(\frac{(e_j)^{1-\gamma(1-\sigma)}}{\varphi_j}\right)^{-\frac{1}{\sigma}}} = 1,$$

$$\Phi = 1 + (1-\sigma) \sum_i \pi_i \mu_i \left(\gamma \omega_i^c + (1-\gamma) e_i \omega_i^h\right) = 1 + (1-\sigma)\mu, \quad \text{and} \quad \Psi = \sum_i \pi_i \mu_i e_i = \mu.$$

Then, the first order condition of the Ramsey problem with respect to C_t, equation (F.24), becomes

$$\nu_t = (1-(1-\sigma)\mu) U_C(C_t, N_t) + \mu U_{CN}(C_t, N_t) + (\eta_t - \eta_{t-1}) U_{CC}(C_t, N_t), \quad \forall\, t \geq 0,$$

where $\nu_t \geq 0$ is the multiplier on the resource constraint at time t, $\mu < 0$ is the multiplier on the incentive-compatibility constraint, and η_t is the multiplier on the upper bound of τ_t^k. Moreover, we have that

$$U_{CN}(C_t, N_t) = -(1-\sigma)(1-\gamma)(1-N_t)^{-1} U_C(C_t, N_t),$$

which allows us the rewrite the equation as

$$\nu_t = \left[1 - (1-\sigma)\mu\left(1 + \frac{1-\gamma}{1-N_t}\right)\right] U_C(C_t, N_t) + (\eta_t - \eta_{t-1}) U_{CC}(C_t, N_t), \quad \forall\, t \geq 0,$$

Now, the argument in Straub and Werning (2020) can be summarized, in this case, as follows: Suppose that $\sigma > 1$ and $\eta_t = 0$, $\forall\, t \geq t^*$, then it is possible to choose μ negative enough such that $\nu_t < 0$ which would yield a contradiction. It follows, therefore, that for μ negative enough the upper bound on τ_t^k would always be binding. To make μ more negative, one needs to increase the planner's need for revenue, for instance by increasing the amount of government expenditures or of initial debt. This leads to more distortionary taxation, and, if extreme enough, positive capital income taxes forever. When lump-sum taxes are available, however, it is easy to see that the corresponding first order condition in the planner's problem, equation (F.27), implies $\mu = 0$. In this case, $\eta_t = 0$, $\forall\, t \geq t^*$ implies $\nu_t = U_C(C_t, N_t) \geq 0$. Distortive taxes are no longer used to obtain revenue (so $\mu = 0$), but only if they allow the planner to provide redistribution more efficiently than the non-distortive lump-sum instrument.

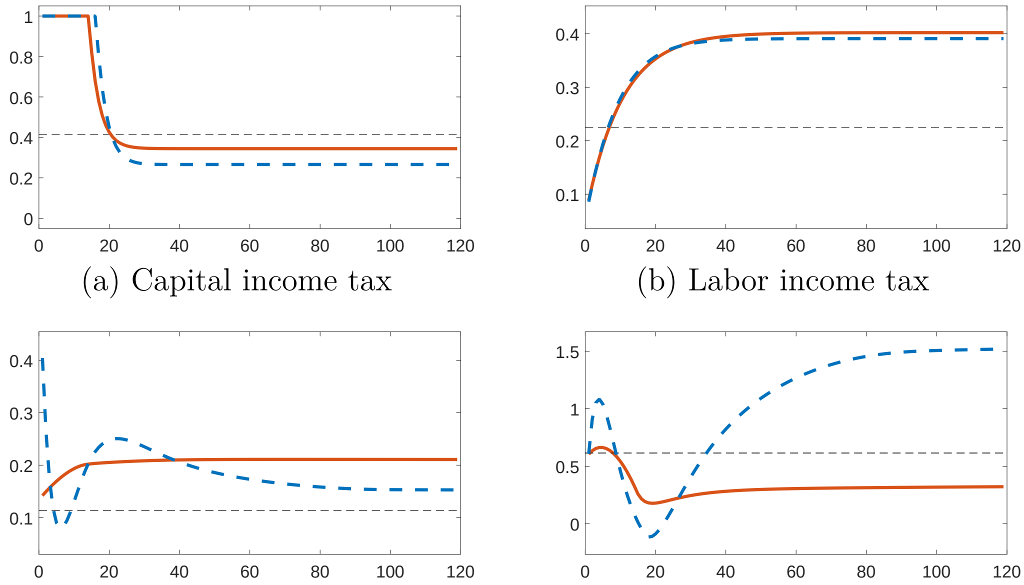


Figure 3: Optimal Fiscal Policy with 8 parameters

Notes: Dashed thin line: initial stationary equilibrium; Dashed thick line: optimal transition with 17 parameters (benchmark); Solid line: optimal transition with 8 parameters.

G Sensitivity Analysis

Figure 3 shows that the solution to the Ramsey problem with 8 parameters (α_0^k, β_0^k, λ^k, α_0^h, β_0^h, λ^h, β_0^T, and λ^T) produces a reasonable approximation for the benchmark solution, at least with respect to its basic features of capital and labor income taxes which we focus one her. The welfare gains with 8 parameters is of 3.4 percent relative to 3.5 percent in the benchmark results. In this appendix we use this approximation to explore to evaluate the robustness of the results with respect to changes in the planner's degree of inequality aversion, the labor-supply and intertemporal elasticities, and the introduction of preference shocks such that labor supply is independent of the productivity level.

G.1 Controlling the Degree of Inequality Aversion

For convenience we reintroduce here the welfare function used to obtain the results in this section. The utilitarian welfare function, which we consider in our benchmark results, places equal Pareto weights on every household. This implies a particular social preference with respect to the equality-versus-efficiency trade-off. Here, we consider a different welfare function that rationalizes different preferences about this trade-off,

$$W^{\hat{\sigma}} = \left(\int \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t, h_t) \right]^{\frac{1-\hat{\sigma}}{1-\sigma}} d\lambda_0 \right)^{\frac{1-\sigma}{1-\hat{\sigma}}},$$

where λ_0 is the initial distribution over individual states (a_0, e_0). Following Benabou (2002), we refer to $\hat{\sigma}$ as the planner's degree of inequality aversion.

By choosing different levels for $\hat{\sigma}$ we can place different weights on the equality versus efficiency trade-off, from the extreme of completely ignoring equality ($\hat{\sigma} = 0$ as in Section 6), passing through the utilitarian welfare function ($\hat{\sigma} = \sigma$), and in the limit reaching the Rawlsian welfare function ($\hat{\sigma} \to \infty$),

$$\lim_{\hat{\sigma} \to \infty} W^{\hat{\sigma}} = \min_{(a_0, e_0)} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t, h_t) \right].$$

Table 4 displays the results for different levels of $\hat{\sigma}$.

Table 4: Degree of Inequality Aversion (Benchmark: $\hat{\sigma} = 1.55$)

	t^*	τ^k	τ^h	Δ	Δ_L	Δ_I	Δ_R
$\hat{\sigma} = 0$	8	27.9	37.7	2.5	0.7	1.0	1.0
$\hat{\sigma} = 1$	11	29.6	40.3	3.3	0.3	1.3	1.7
$\hat{\sigma} = 1.55^*$	14	34.3	40.2	3.4	0.1	1.3	2.1
$\hat{\sigma} = 10$	14	34.4	39.7	3.0	-1.0	1.6	2.4

Note: When $\hat{\sigma} = 1.55 = \sigma$, the welfare function is utilitarian. The values for the fiscal instruments are the ones from the final steady state. For comparability, all results in the table have been computed using 8 parameters, and for the average welfare gains and its components we use the utilitarian welfare function.

Higher $\hat{\sigma}$ imply a stronger desire for redistribution. Accordingly, the higher $\hat{\sigma}$ is, the higher t^*, τ^k, and τ^h, since all, rebated via lump-sum transfers increase the amount of redistribution achieved by the policy. However, notice that, specially when the desire for redistribution is higher than the utilitarian, the long-run levels of τ^k and τ^h do not change significantly. For $\hat{\sigma} = 10$, labor taxes in the long run are actually slightly lower, but they are significantly higher in the short run which also explains the sizable negative level effect associated with that policy. Appendix O.8 presents the corresponding paths for the instruments and aggregates.

G.2 Labor-Supply and Intertemporal Elasticities

Labor-elasticity and intertemporal elasticity of substitution are important parameters in the optimal taxation literature. In this section we conduct the sensitivity analysis regarding these two elasticities.

The parameter $\phi \equiv \gamma(\sigma - 1)$ controls three important aspects of our benchmark experiment: relative risk aversion given by $\sigma\gamma + 1 - \gamma$, the households' intertemporal elasticity of substitution (IES), and the planner's degree of inequality aversion. In Table 5 we present model's statistics next to the corresponding targets for three IES values: 0.5, 0.65 (benchmark) and 0.8. For all three values we obtain a similar fit to the data. In Table 6 we present analogous results for different values of the aggregate Frisch elasticity: 0.35, 0.5 (benchmark) and 0.65. Again, we obtain a similar fit to the data for all three values of the Frisch elasticity.

Table 5: Alternative Calibrations of IES

	Target	**IES**		
		0.50	**0.65**	**0.80**
Aggregate Frisch elasticity (Ψ)	0.54	0.48	0.49	0.47
Average hours worked	0.32	0.33	0.33	0.34
Capital to output	2.50	2.49	2.49	2.48
Capital income share	0.38	0.38	0.38	0.38
Investment to output	0.26	0.26	0.26	0.26
Transfer to output (%)	11.40	11.40	11.40	11.40
Debt-to-output (%)	61.50	61.50	61.50	61.50
Share of workers (%)	76.70	81.39	79.30	79.14
Hhs with negative net-worth (%)	9.73	10.15	7.86	9.38
Correlation(earnings, wealth)	0.43	0.42	0.43	0.38
Variance of 1-year lab.inc. growth rate	2.33	2.37	2.32	2.36
Kelly skewness of 1-year lab.inc. growth rate	-0.12	-0.13	-0.13	-0.13
Moors kurtosis of 1-year lab.inc. growth rate	2.65	2.13	2.18	2.31
Share of self-emp. in population (%)	0.12	0.13	0.13	0.13
Share of wealth of self-emp. (%)	0.46	0.39	0.39	0.40
Share of earnings of self-emp. (%)	0.29	0.32	0.31	0.31

	Bottom (%)	**Quintiles**					**Top (%)**	**Gini**
	0-5	**1st**	**2nd**	**3rd**	**4th**	**5th**	**95-100**	
Wealth								
US Data	-0.2	-0.2	1.0	4.2	11.2	83.8	60.0	0.82
IES 0.50	-0.1	0.0	1.7	3.6	9.5	85.2	57.9	0.82
IES 0.65	-0.1	0.1	2.0	4.0	9.3	84.5	56.4	0.81
IES 0.80	-0.2	0.0	1.9	4.2	9.4	84.4	59.1	0.81
Earnings								
US Data	-0.2	-0.2	4.1	11.6	20.9	63.6	35.6	0.64
IES 0.50	0.0	0.0	5.7	10.8	20.4	63.1	35.1	0.62
IES 0.65	0.0	0.0	5.7	11.3	20.2	62.8	34.8	0.62
IES 0.80	0.0	0.0	5.0	10.7	20.2	64.2	35.6	0.64
Hours								
US Data	0.0	3.0	13.7	20.7	25.4	37.2	12.9	0.34
IES 0.50	0.0	0.2	14.0	23.1	26.9	35.8	9.8	0.35
IES 0.65	0.0	0.0	13.2	23.4	27.1	36.3	9.9	0.36
IES 0.80	0.0	0.0	12.5	23.1	27.3	37.1	10.3	0.37

Table 6: Alternative Calibrations of Frisch

	Target	Frisch		
		0.35	**0.50**	**0.65**
Intertemporal elasticity of substitution	0.65	0.65	0.65	0.65
Average hours worked	0.32	0.35	0.33	0.31
Capital to output	2.50	2.53	2.49	2.50
Capital income share	0.38	0.38	0.38	0.38
Investment to output	0.26	0.26	0.26	0.26
Transfer to output (%)	11.40	11.40	11.40	11.40
Debt-to-output (%)	61.50	61.50	61.50	61.50
Share of workers (%)	76.70	82.04	79.28	81.17
Hhs with negative net-worth (%)	9.73	8.21	7.86	9.49
Correlation(earnings, wealth)	0.43	0.47	0.43	0.46
Variance of 1-year lab.inc. growth rate	2.33	2.36	2.32	2.33
Kelly skewness of 1-year lab.inc. growth rate	-0.12	-0.13	-0.13	-0.14
Moors kurtosis of 1-year lab.inc. growth rate	2.65	2.19	2.15	2.06
Share of self-emp. in population (%)	0.12	0.13	0.13	0.12
Share of wealth of self-emp. (%)	0.46	0.38	0.39	0.38
Share of earnings of self-emp. (%)	0.29	0.29	0.31	0.32

	Bottom (%)	**Quintiles**					**Top (%)**	**Gini**
	0-5	**1st**	**2nd**	**3rd**	**4th**	**5th**	**95-100**	
				Wealth				
US Data	-0.2	-0.2	1.0	4.2	11.2	83.8	60.0	0.82
Frisch 0.35	-0.1	0.1	2.2	4.9	10.4	82.4	57.7	0.80
Frisch 0.50	-0.1	0.2	2.0	4.0	9.3	84.5	56.4	0.81
Frisch 0.65	-0.1	-0.0	1.3	3.0	10.0	85.7	48.7	0.80
				Earnings				
US Data	-0.2	-0.2	4.1	11.6	20.9	63.6	35.6	0.64
Frisch 0.35	0.0	0.0	5.3	10.9	17.0	66.8	34.8	0.64
Frisch 0.50	0.0	0.0	5.7	11.3	20.2	62.8	34.8	0.62
Frisch 0.65	0.0	0.2	6.1	10.2	19.0	64.4	38.7	0.63
				Hours				
US Data	0.0	3.0	13.7	20.7	25.4	37.2	12.9	0.34
Frisch 0.35	0.0	0.1	13.1	23.5	27.3	36.0	10.3	0.36
Frisch 0.50	0.0	0.0	13.2	23.4	27.1	36.3	9.9	0.36
Frisch 0.65	0.0	0.8	15.0	22.2	26.4	35.6	9.9	0.34

Table 7: Elasticities of Intertemporal Substitution and Frisch (Benchmark: IES= 0.65, Frisch= 0.5)

	t^*	τ^k	τ^h	Δ	Δ_L	Δ_I	Δ_R
IES = 0.5	71	22.8	39.9	5.9	0.8	0.9	4.1
IES = 0.8	10	25.5	37.9	2.9	0.4	1.2	1.3
Frisch = 0.35	14	32.5	42.3	3.7	-0.1	1.6	2.2
Frisch = 0.65	13	32.2	39.8	3.2	0.1	1.1	2.0
Benchmark (8 parameters)	14	34.3	40.2	3.4	0.1	1.3	2.1

Note: The values for the fiscal instruments are the ones from the final steady state. For comparability, all results in the table have been computed using 8 parameters.

The optimal policy results are most sensitive to changes in the IES, for the reason mentioned in the beginning of this section. When the IES increases from 0.65 to 0.8, the planner's inequality aversion is reduced and, accordingly, capital income taxes are kept at the upper bound for less periods (t^* goes from 14 to 10). Moreover, the higher IES and correspondingly lower risk aversion implies that long-run capital income taxes lead to, at the same time, higher distortions and less insurance benefits. It follows that the optimal long-run capital income tax is lower. With a lower IES of 0.5 we see the most dramatic effects as the associated increase in the planner's degree of inequality aversion generates a shift to obtain most welfare gains via redistribution. That is achieved by keeping capital income taxes at the upper bound for 71 years, the distortionary effects being mitigated by the lower IES. Appendix O.9 presents the corresponding paths for the instruments and aggregates.

Intuitively, a higher Frisch elasticity implies a lower optimal labor income tax and a higher associated level effect, though the results are significantly less sensitive to these changes relative to changes in the IES. Note that these results are in line with the propositions established in Section 2. Appendix O.10 presents the corresponding paths for the instruments and aggregates.

G.3 Adding parameters to Chebyshev Approximation

To arrive at the 17 parameters used in our benchmark experiments we started with a very small set of parameters and gradually increased the number until the optimal paths for fiscal instruments stopped moving in a meaningful way. Appendix O.7 presents the figures of the corresponding paths for the instruments and explains exactly how these parameters were used. Table 8 presents the corresponding welfare gains. Our benchmark experiment has 17 parameters. The experiment with 20 parameters was run to check that further increases would not affect the paths or welfare any further, we do not take this to be the benchmark because it is too computationally expensive and we want to be able to run comparable experiments for maximization of efficiency and the experiment with a capital levy for instance. We think both the magnitudes of the differences in welfare gains and in the optimal paths are small enough to indicate that the 17-node solution to be a good enough approximation of the actual optimum.

Table 8: Adding parameters

	2	3	8	11	14	16	17*	20
Average welfare gain (Δ)	1.647	2.793	3.398	3.453	3.455	3.514	3.517	3.520

G.4 Final Period for Movements in Instruments

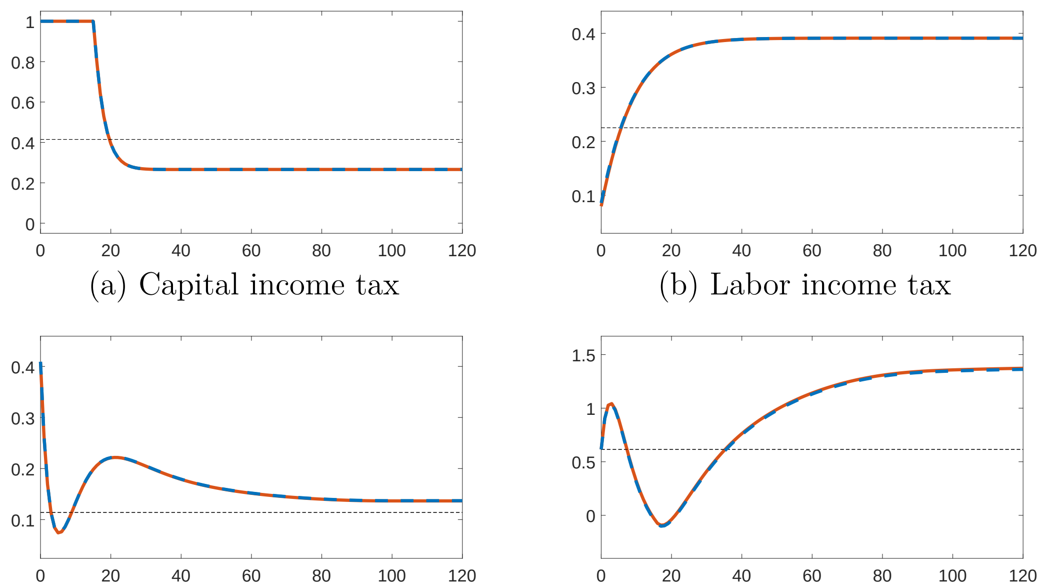
We allow capital and labor income taxes and lump-sum transfers to move for 100 years in our benchmark experiment. To make sure this choice is not affecting our results we increased this number, in increments of 10, until 150 and reoptimized at each step. Table 9 shows that average welfare gains of the optimal policy follow an inverse U-shape in the final year reaching the maximum at 120. The trade-off being that a higher final period increases the flexibility of the paths in the long-run at the expense of less flexibility in the short run. The magnitude of the differences, however, are small enough that we think our choice of 100 does not affect the results in any significant way. Moreover, since one of the choice variables in the new approximation method is a convergence rate for each fiscal instrument, the point at which the instruments become constant is endogenous and it is actually chosen to be significantly lower than 100 for all of our benchmark results, see Appendix O.1. It is also worth noticing that since in the Ramsey experiment the full path of taxes are announced, the state variables start to converge even before the taxes stop moving.

Table 9: Final period of movements in instruments

	100	110	120	130	140	150
Average welfare gain (Δ)	3.517	3.518	3.519	3.518	3.516	3.513

G.5 Terminal Period of the Transition

In this section, we present a robustness check regarding the terminal period of the transition. In our Benchmark experiment, we set the length of the transition to 250 periods. To inspect the sensitivity of the main results with respect to this assumption, we reran our optimization algorithm while doubling the terminal period, setting it to 500. The comparison of the optimal policy with the benchmark can be seen in Figure 4. Aside from a slight difference in the final debt level, the policies are essentially indistinguishable.



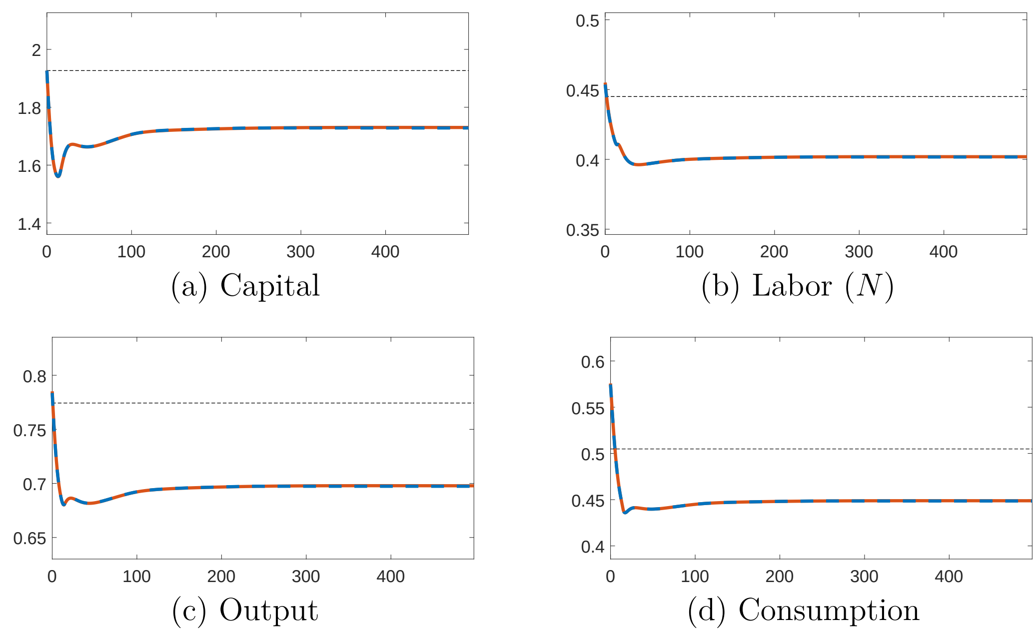
(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 4: Optimal Fiscal Policy

Note: Black dashed lines: initial stationary equilibrium; Blue dashed curves: optimal transition with terminal period of 250 (benchmark); Red solid curves: optimal transition with terminal period of 500.

Importantly, the welfare gains from the Benchmark experiment and the transition with the terminal period set at 500 are also almost identical: 3.5179 vs. 3.5176 percent respectively. There are two main reasons for this: (1) all aggregates associated with the optimal policy have mostly converged by period 250, as can be seen in Figure 5; and (2) the small differences that appear after period 250 have insignificant welfare implications since $\sum_{t=250}^{\infty} \beta^t / \sum_{t=0}^{\infty} \beta^t \approx 7 \times 10^{-6}$ for our benchmark $\beta = 0.9538$.



(a) Capital

(b) Labor (N)

(c) Output

(d) Consumption

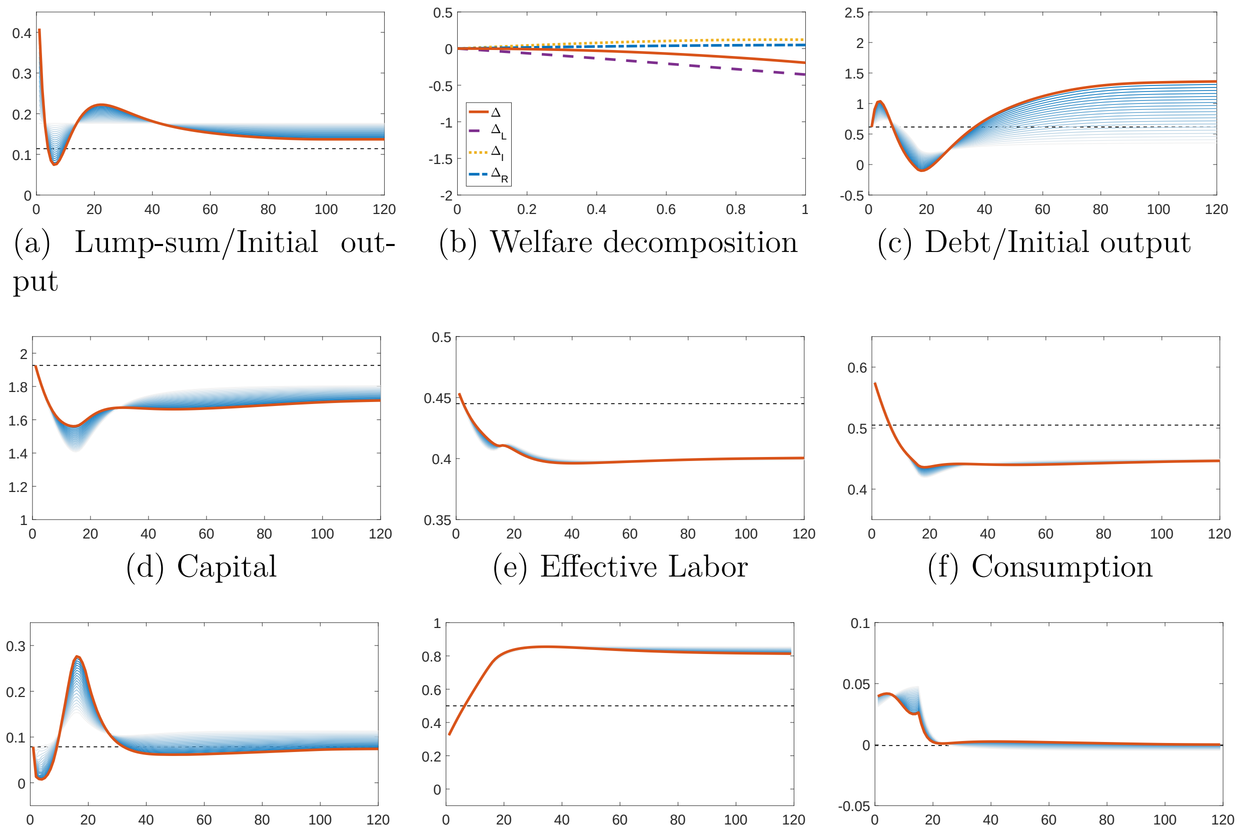
Figure 5: Aggregates

Note: Black dashed lines: initial stationary equilibrium; Blue dashed curves: optimal transition with terminal period of 250 (benchmark) extended by a constant in the last 250 periods; Red solid curves: optimal transition with terminal period of 500.

H Understanding the Lump-Sum Path

H.1 More details on the variation towards constant lump-sum transfers

Figure 6 presents more detail on the results from Figure 9 in the paper—the y-axes in Figure 6 have been set to facilitate comparison with the next robustness exercise presented in Figure 7 in the next subsection.



(a) Lump-sum/Initial output (b) Welfare decomposition (c) Debt/Initial output

(d) Capital (e) Effective Labor (f) Consumption

(g) Prop. w/ Neg. Assets (h) Intratemporal Wedge (i) Intertemporal Wedge

Figure 6: Varying Lump-Sum Transfers Towards a Constant

Notes: (a) Black dashed line: initial stationary equilibrium; Red and blue solid curves: optimal transition and perturbations of it; (b) the x-axis represents the movement in number of periods capital income taxes are kept in the upper bound from the optimum, y-axis shows change in the welfare gains in percent points.

Panel 6g shows that the path of the proportion of households with negative assets is smoothed out by moving to a constant path for transfers. Panel 6b shows that this actually leads to redistributive and insurance welfare gains, however, the losses from the level effect more than outweigh these gains. So, we are left with understanding where these losses come from. Recall that the level effect captures the welfare of the average household in the economy. So, it is instructive to look at aggregates. Panel 6d shows that the path of aggregate capital decreases more in the short run as the path of transfers becomes flatter which is in line with the argument that the back-loading of lump-sum transfers mitigates distortions to savings decisions associated with high capital income taxes in the first regime. Panel 6f shows that the consumption of the average household is also less smooth when transfers become constant.

To be more precise about the effects on distortions, Panels 6h and 6i show what happens to wedges to the Euler equations of the average household. We define these wedges as follows:

$$\text{Intratemporal Wedge}_t = \frac{u_c\left(C_t, H_t\right)}{u_h\left(C_t, H_t\right)} w_t - 1,$$

$$\text{Intertemporal Wedge}_t = \frac{\beta\left(1 + r_{t+1}\right) u_c\left(C_{t+1}, H_{t+1}\right)}{u_c\left(C_t, H_t\right)} - 1,$$

where C_t and H_t denote the consumption and hours worked of the average household. These wedges capture exactly the welfare losses accounted for in the level effect. Notice that, while the intratemporal wedge is not significantly affected by changes to the path of lump-sum, the intertemporal wedge is. Having lump-sum transfers that increase over time between periods 8 and 20, as in the optimal path, leads to a significant reduction in the intertemporal wedge in those periods. To understand this, notice that a positive intertemporal wedge means that the average household is not saving enough or borrowing too much. Therefore, an increasing path of lump-sum transfers mitigates this distortion in the periods preceding it: households save more of the initial lump-sum transfers they receive in order to avoid being borrowing constrained (which, as it turns out, many are still not able to avoid).

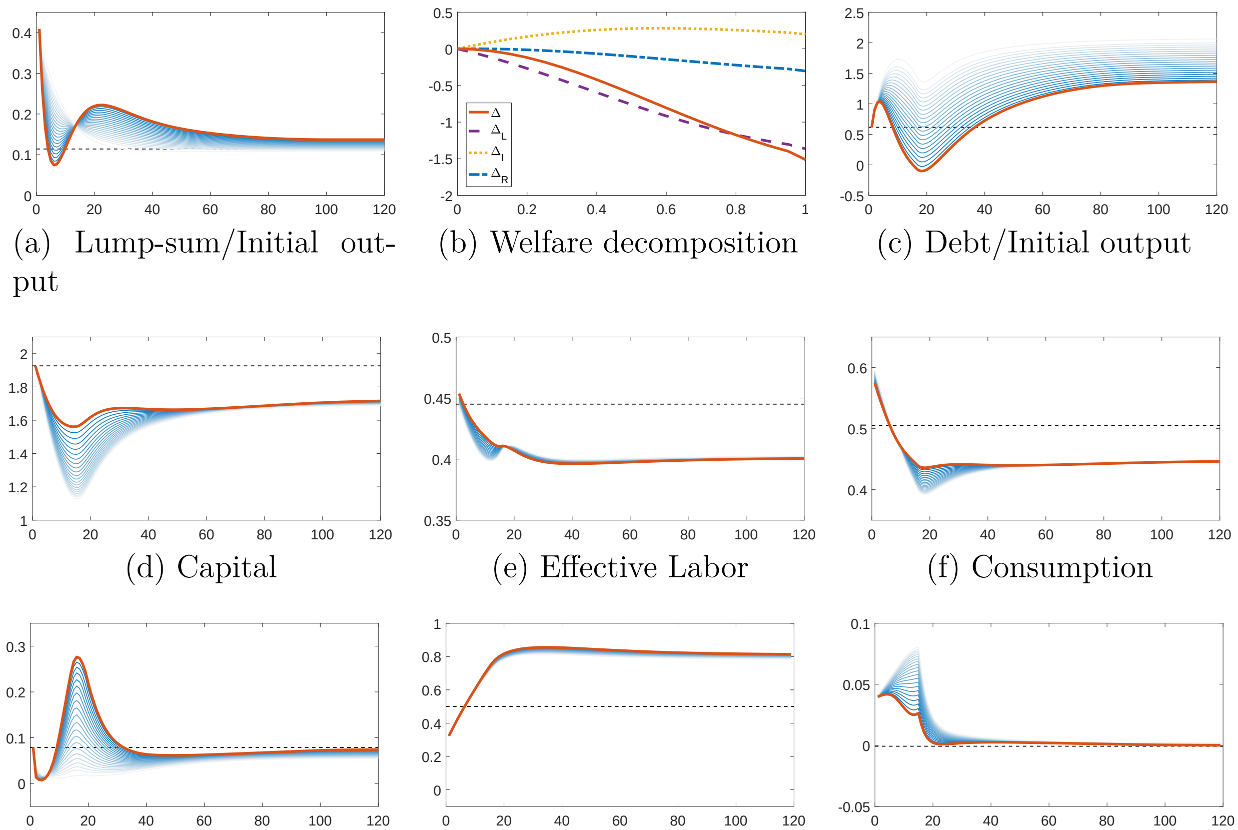
H.2 Variation towards monotonic lump-sum transfers

In this section we consider an alternative variation that is slightly harder to motivate, but that makes the effects shown in the previous section even clearer. We created a path for lump-sum transfers by setting the difference between the initial and the final value to be the same as in the optimal path, and by having the path converge monotonically over time from the initial to the final value. Figure 7 shows what happens as the benchmark optimal path of lump-sum transfers is gradually moved to this monotonic path.

There are two details we should mention. We balance the intertemporal budget constraint of the government with the average level of lump-sum transfers, which explains why the long-run level in Panel 7a is different from the optimal path: more front-loading must be balanced by lower transfers in the future. Moreover, we must choose some convergence rate for the monotonic path. We thought that a reasonable rate would be the convergence rate of the optimal lump-sum path in the 8-parameter solution, so that is the one we use.

Notice, in Panel 7b, that the welfare losses come again mostly from the level effect. The main difference relative to the results from Panel 6b are the fact that the welfare effects are now an order of magnitude larger. This is helpful, since it indicates that the causes for these welfare losses should be even clearer in this experiment.

In line with the reasoning put forward above, it is easy to see from Panels 7e and 7f that the aggregate labor and consumption are less smooth as lump-sum converges to the monotonic path. This is a result of the fact that the additional front-loading of lump-sum transfers implied by the monotonic path implies an additional accumulation of government debt seen in Panel 7c. This increase in debt, in the initial 17 periods, compounds with the capital income taxes at the upper bound which already reduce household savings leading to a significantly larger reduction to aggregate capital. The distortions to the intertemporal wedge in Panel 7i are also consistent with this. Finally, in line with the referee's reasoning, 7g shows that with the monotonic path the proportion of households close to the constraint no longer displays the sharp increase until period 17 generated by the optimal path. This, again, despite the obvious benefits, actually leads to more distortions to the intertemporal margin, on average, since households no longer save to avoid being constrained (an incentive that allows for the mitigation of savings distortions in the optimal path).



(a) Lump-sum/Initial output (b) Welfare decomposition (c) Debt/Initial output

(d) Capital (e) Effective Labor (f) Consumption

(g) Prop. w/ Neg. Assets (h) Intratemporal Wedge (i) Intertemporal Wedge

Figure 7: Varying Lump-Sum Transfers Towards a Monotonically Decreasing Path

Notes: (a) Black dashed line: initial stationary equilibrium; Red and blue solid curves: optimal transition and perturbations of it; (b) the x-axis represents the movement in number of periods capital income taxes are kept in the upper bound from the optimum, y-axis shows change in the welfare gains in percent points.

H.3 Back-loading lump-sum transfers

In this section, we consider the following experiment that clarifies the fact that the path of lump-sum transfers and debt can have important welfare implications. We start from the "Constant lump-sum" experiment in which lump-sum transfers are required to stay constant after an initial jump in period 0 and we reoptimize other instruments given this constraint (see Table 4). We then force lump-sum transfers to remain at their pre-reform levels for T periods, jumping in period T to the level that balances the government's present-value budget constraint. We keep the paths capital and labor income taxes fixed as we change T.

Figure 8 shows the results for for $T \in \{0, 1, ..., 30\}$. The figure displays the paths of: lump-sum transfers, government debt, the proportion of households with negative assets, and average welfare. The experiment with $T = 0$, which generates welfare gains of 3.4 percent, is shown in the lightest shade of gray, with darker shades of gray being associated with higher levels of T. It is clear from the figures that backloading lump-sum transfers (by increasing T) leads to: (1) more accumulation of assets by the government; (2) a higher proportion of households with negative assets who are, therefore, more likely to be borrowing constrained; and (3) a reduction in average welfare gains down to almost 0 when $T = 30$. The intuition is straight forward, backloading lump-sum transfers pushes households against their borrowing constraints which reduces welfare. Finally, for $T = 30$, we ran an additional experiment in which we reoptimize the paths of capital an labor income taxes subject to the additional restriction to the path of lump-sum transfers. This experiment leads to welfare gains of 2.5 percent, which we represent as the red dot in the last panel of Figure 8. Allowing for the other taxes to adjust significantly mitigates the welfare effects of the constraint to the path of lump-sum transfers (generating gains of 2.5 percent instead of 0.2 percent when taxes are kept fixed). Nevertheless, the reduction from 3.4 to 2.5 is still sizable and we think it is safe to assume this difference would only increase further for higher levels of T. Indeed, in the limit as $T \to \infty$ we would reach the "fixed lump-sum" experiment which generates welfare gains of 2.1 percent.

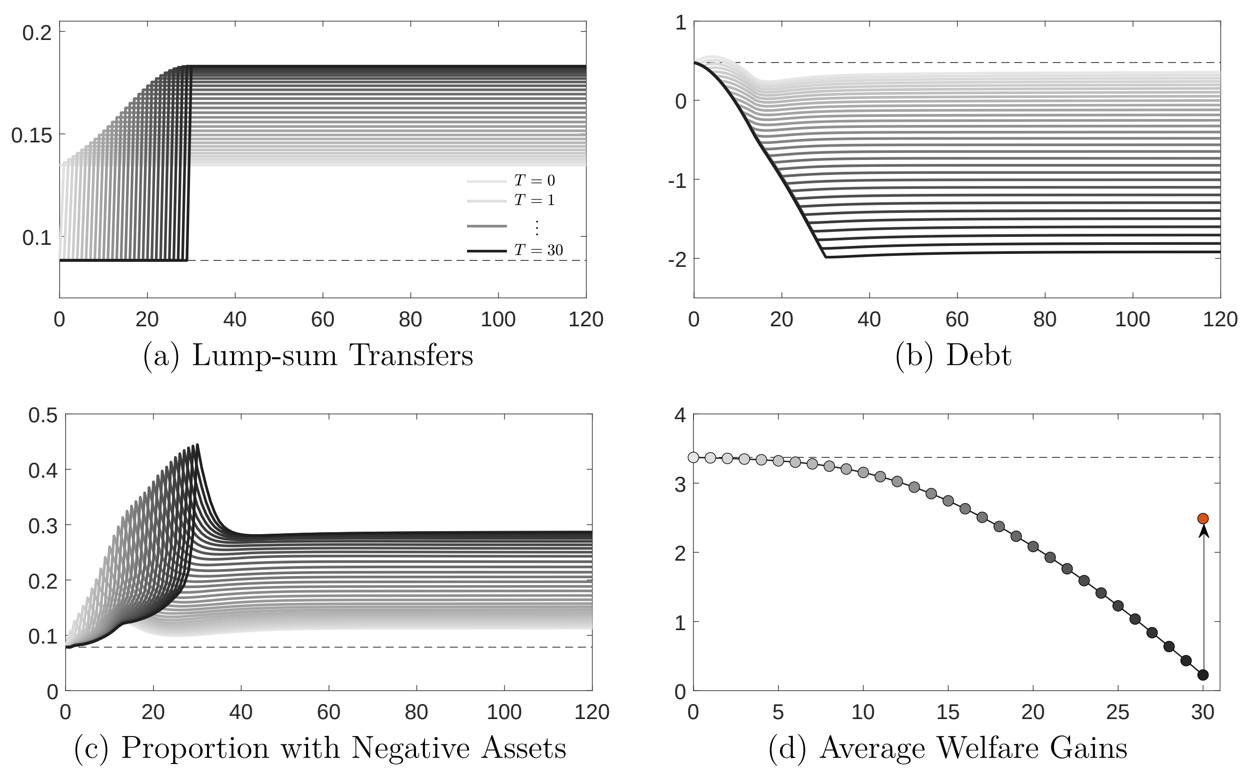


Figure 8: Backloading Lump-sum Transfers

Notes: Black dashed line: initial stationary equilibrium; Lightest gray solid line: constant lump-sum experiment; Darker shades of gray are associated with more periods of lump-sum fixed at the initial level following the legend in the first panel. Red dot in last panel is welfare when capital and labor income taxes are reoptimized given the restriction to lump-sum transfers.

I Initial Capital Levy

Since the utilitarian planner wants to front load capital income taxes, we conduct another experiment in which we remove the upper bound to capital income taxes, see Figure 9. Unsurprisingly, we find that the planner expropriates all asset holdings in period 0. Surprisingly, however, this does not lead to lower capital income taxes in the future periods, on the contrary, long-run capital income taxes are higher than in the benchmark experiment.

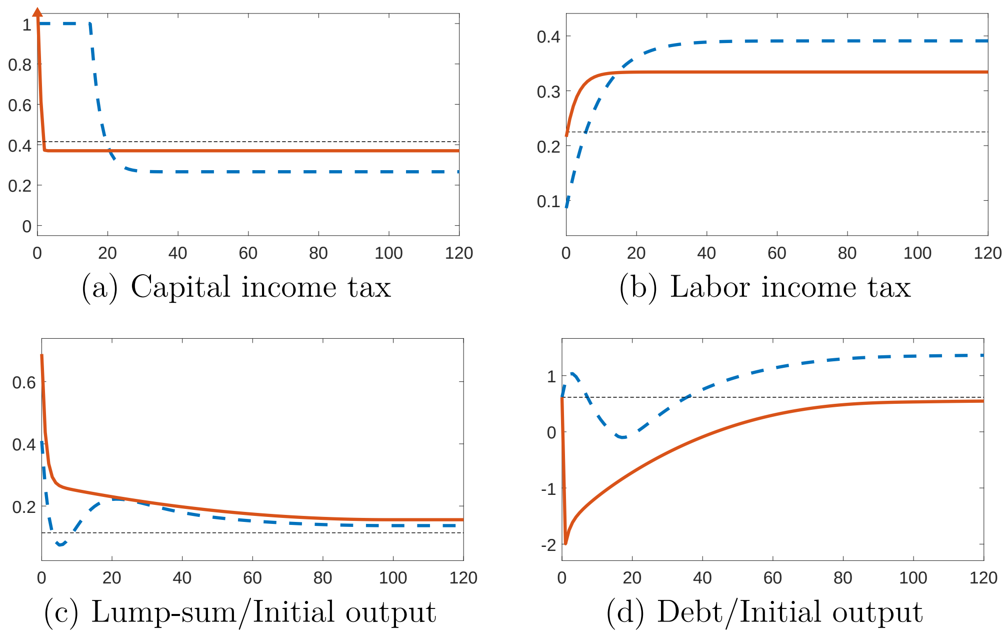


Figure 9: Optimal Fiscal Policy: Levy on Initial Capital Income

Note: Thin dashed line: initial stationary equilibrium; Solid line: path that maximizes the utilitarian welfare function allowing for capital income taxes to move in period 0 (though the tax level at $t = 0$ is not plotted since it is equal to $(1 + r_0)/r_0 = 21.96$); Thick dashed line: benchmark results.

With the capital levy, precautionary savings are expropriated in period 0. Households immediately begin to rebuild their buffer stock and these efforts are not significantly diminished by the high capital income taxes since they are focused on the stocks. These savings decision are, therefore, relatively inelastic making capital income taxes less distortive. Moreover, on impact the government obtains a lot of revenue obtaining a sizable asset position. As a result, capital is crowded in and the downward distortions to capital accumulation associated with capital income taxes are, again, less relevant. On the other hand, capital income taxes are still beneficial to provide redistribution (mostly in the short run) and insurance (mostly in the long run). Importantly, even though capital income taxes are overall higher relative to the benchmark, the equilibrium capital stock is still higher throughout the transition—Appendix O.4 presents as extensive list of figures comparing the two economies. The welfare gains are equivalent to a permanent 14.2 percent increase in consumption, 5.7 percent coming from the level effect, 1.4 percent from the insurance and 6.6 percent from redistribution.

J Constant Lump-Sum Transfers

In Table 4 of the paper, we present the welfare decomposition for experiments in which we keep each instrument fixed at their initial level. We replicate in the second row of Table 10 the results for lump-sum transfers.

Table 10: Welfare Decomposition: Effect of Lump-Sum Path

	Δ	Δ_L	Δ_I	Δ_R
Benchmark	3.5	0.2	1.2	2.1
Benchmark with constant lump-sum	3.3	-0.1	1.3	2.1
Reoptimizing subject to constant lump-sum	3.4	0.1	1.3	2.0

One important detail that should be mentioned is that the level of lump-sum transfers still need to balance the intertemporal budget of the government. So, the experiment here is exactly equivalent to the one presented in Figure 9 of the paper. The main take-away is that the losses associated with keeping transfers constant over time have to do with level effect, since those movement are helpful in mitigating distortions from high capital income taxes.

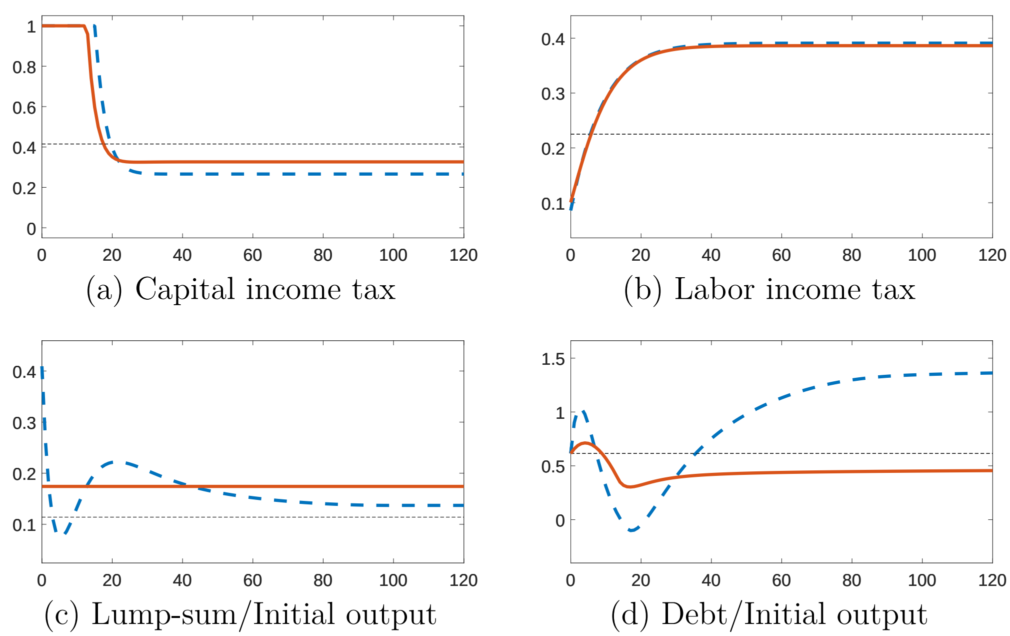


Figure 10: Optimal Fiscal Policy: Constant Lump-Sum Transfers

Notes: Thin dashed line: initial stationary equilibrium; Solid line: path that maximizes the utilitarian welfare function with the added restriction that lump-sum transfers are not allowed to vary over time after the initial change; Thick dashed line: benchmark results.

In the experiment from Table 4 and Figure 9, while lump-sum transfers are moved to a constant the income taxes are kept at their benchmark optimal paths. To investigate if the path for lump-sum transfers affects the optimal paths for capital and labor income taxes, we also ran another experiment in which we reoptimize subject to the constraint that lump-sum transfers must be kept constant over time (though its level can change on impact). The welfare decomposition for these results is shown in the third row of Table 10 and the take away from those results are essentially the same as for the second row. Figure 34 shows what happens to in particular to the paths of capital and labor income taxes. Labor income taxes are essentially unaffected. Capital income taxes are affected in two ways: (1) it stays in the upper bound for less periods; and (2) its long-run level is higher. The first change is exactly what one should expect given the intuition laid out above: without the mitigation of distortions to the savings decisions generated by the optimal non-monotonic path it is optimal to keep capital income taxes at the upper bound for less periods. The second effect has to do with the fact that a constant lump-sum path implies a significantly lower government debt in the long-run relative to the benchmark results. A lower debt crowds in capital justifying a relatively higher long-run capital income tax. Appendix O.5 presents an extensive list of complementary figures.

K Elasticity of Top 1 Percent

The earnings elasticity of the most productive households plays a role in some of the arguments we present below. So, we followed the procedure in Kindermann and Krueger (2021) to calculate this elasticity for the top 1 percent.

First, we calculated the Pareto distribution parameter that best approximates the earnings distribution in our model and obtained a value of $a = 1.806$. To calculate this number we use the results displayed in Figure 11 and, following Saez (2001), calculate a from $a/(a-1) = 2.24$.

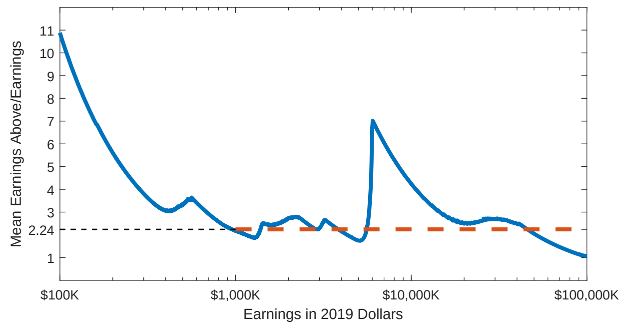


Figure 11: Pareto Fit

Note: On the y-axis we plot the mean earnings above each level divided by that level in our model. The Pareto distribution implies that this statistic is constant. The red-dashed line show the mean of this ratio for earning levels above one million 2019 dollars weighted by stationary distribution of earnings.

Second, starting from our initial stationary equilibrium we lowered the labor income tax (therefore increasing the net-of-tax rate) for all households in the top 1 percent of earnings by $\delta = 0.01$, and computed a transition towards a new stationary equilibrium—the threshold for the top 1 percent was updated in each period of the transition. Denoting the earnings of the top 1% on impact (the first period of transition) by z_h^1 and its benchmark level (the initial stationary equilibrium) by z_h^0, we can compute the elasticity

$$\epsilon(z_h) = \frac{z_h^1 - z_h^0}{\delta} \frac{1 - \tau^h}{z_h^0} = 0.16.$$

Combining this result with the estimates for the Pareto distribution parameter, we can calculate the peak of the Laffer curve using the formula

$$\tau_{\text{Laffer},t} = \frac{1}{1 + a\epsilon(z_h)} = 0.78.$$

The preferred value for this statistic in Kindermann and Krueger (2021) is of 0.73. So, even though we do not have a perfect match, given the fact that we did not target this statistic in any way, we think the result is reasonably accurate, and can be used as an external validation of our calibration strategy.

L The Importance of Transition

In this section, we illustrate the importance of taking transitional effects into account while computing the optimal fiscal policy. We contrast the maximizing steady-state welfare approach with one-time, optimal policy change and further with our benchmark results. We present this comparison in Table 11.

Suppose the planner can choose stationary levels of all four fiscal policy instruments to maximize steady-state welfare. In particular, the planner can choose any level of government debt without incurring the transitional costs associated with it. It then chooses a debt-to-output ratio of −265 percent. At this level the amount of capital that is crowded in is close to the golden rule level, which implies zero interest rates (net of depreciation). Since capital income is zero, capital income taxes are not relevant which is why we do not display that number in Table 5. The average welfare gain associated with this policy is 14.8 percent. These are large welfare gains precisely because they ignore transitional effects, as if the economy has jumped immediately to a new steady state with a new distribution of assets, a much higher capital stock, and in which the government has a large amount of assets instead of debt.[9] An alternative experiment, which is closer to the one studied by Conesa et al. (2009), is to restrict the level of debt-to-output to remain at its initial level and choose only the other fiscal instruments. With this constraint, we find it is still optimal for the planner to focus on the level effect. Though the golden rule level of capital is not achieved, a negative capital income tax of −7.2 leads the capital level in that direction. The planner also sets relatively low labor income tax and transfer levels which are detrimental to insurance and redistribution, but reduce distortions. Ignoring transitional effects, the policy leads to an average welfare gain of 1.2 percent. However, accounting for its transitional effects the policy would actually lead to a welfare *loss* equivalent to an 3.5 percent permanent reduction in consumption. The difference between constant, optimal policy and our benchmark experiment, which are also presented in Table 11, can be found in Section 7.1 in the main body of the paper.

Table 11: Final Stationary Equilibrium: Effects of Time-Varying Policy

	τ^k	τ^h	T/Y	B/Y	K/Y	Δ	Δ_L	Δ_I	Δ_R
Initial equilibrium	41.5	22.5	11.4	61.5	2.49	–	–	–	–
Stat. equil.	–	36.4	18.8	−265.1	3.53	14.8	8.1	0.7	5.5
Stat. equil. no debt	−7.2	27.1	9.1	61.5	2.85	1.2	2.8	0.0	−1.5
Constant policy	67.5	27.9	19.7	53.9	2.02	1.7	−0.7	0.8	1.6
Benchmark	26.7	39.1	15.1	154.3	2.48	3.5	0.2	1.2	2.1

Note: All values, except for K/Y, are in percentage points.

[9] In Appendix N.1 we compare these results with the ones in Aiyagari and McGrattan (1998) in detail. The main reason for the differences in the results is that the calibration in that paper leads to significantly lower levels of inequality.

M Relationship with Acikgoz, Hagedorn, Holter, and Wang (2018)

The main results in our paper (sometimes denoted in this appendix by DP), presented Section 5, differ in many dimensions from the results obtained by Acikgoz, Hagedorn, Holter, and Wang (2018) (denoted by AHHW). There are two possible causes for these differences:

1. **Calibration.** AHHW work with preferences that are separable in consumption and labor,

$$u(c,h) = \frac{c^{1-\sigma}}{1-\sigma} - \chi \frac{h^{1+1/\phi}}{1+1/\phi}$$

while we use non-separable preferences that are consistent with a balanced growth path,

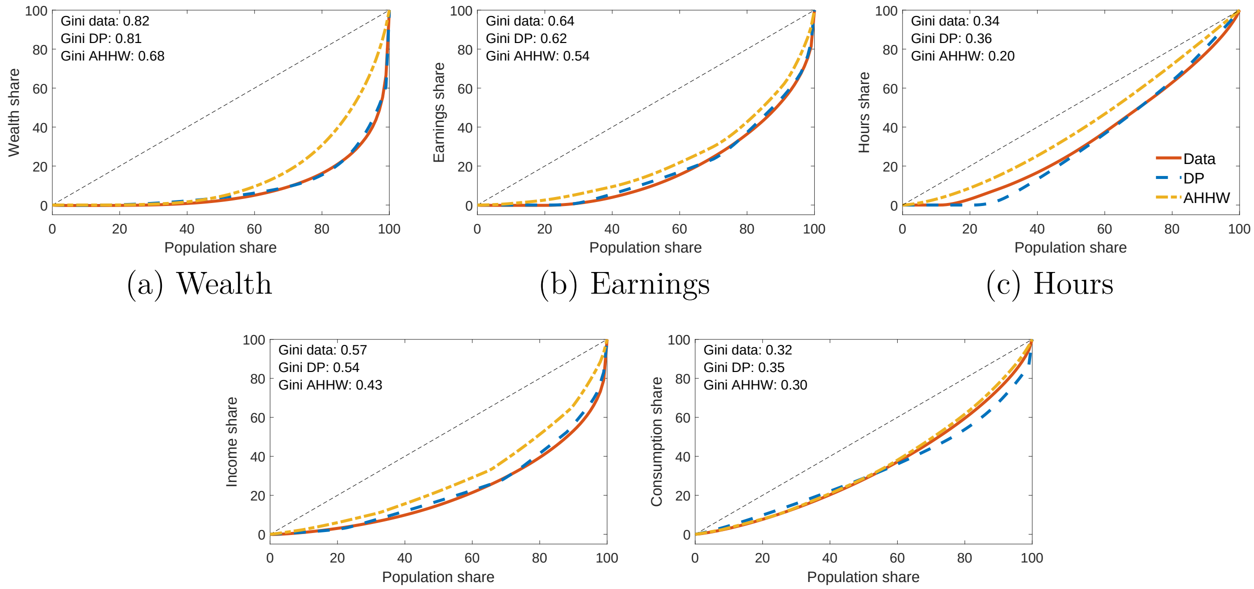
$$u(c,h) = \frac{(c^{\gamma}(1-h)^{1-\gamma})^{1-\sigma}}{1-\sigma}.$$

We also differ substantially in the calibration of the income process with our method targeting many more moments of the labor income process and of the earnings wealth and hours distributions. The full set of parameters used by AHHW are presented in Table 12, whereas our parameters are presented in the paper in Table 1. The different parameterizations lead to differences in every margin relevant for the Ramsey planner:

   - Elasticities: These are important determinants of how distortive taxes are. We use an intertemporal elasticity of substitution (IES) of 0.65 while AHHW use 0.5. We use a Frisch elasticity for the households of 0.49 and AHHW use 1.0.
   - Risk: The variance of before-government income implied by our calibration is 2.94,[10] which is very different from the corresponding number implied by the calibration in AHHW of 0.13. We think this is an important determinant of the optimal policy and that is why we target is directly. Moreover, the risk faced by the households is not fully summarized by variance of the growth rate of income, which is why we also target its skewness and kurtosis, AHHW abstract from this.
   - Inequality: we target many moments of the wealth, earnings and hours distributions, while AHHW include only one distributional target, the ratio of asset holdings at the 90 percentile over that of the 50 percentile. Figure 12 the fit of both models to inequality data. In general, relative to our calibration and the data we use the calibration from AHHW generates less wealth, hours, earnings and income inequality.

Finally, we think that working with BGP preferences has the advantage that they have been used in a number of important papers in the macro-public-finance literature, such as Aiyagari and McGrattan (1998), Conesa et al. (2009) and Röhrs and Winter (2017). This allows for a direct comparison of our results with the ones established in the literature. Also, given the better fit of the DP calibration to the

[10] This differs from the number reported in the paper, 2.32, because for comparability with AHHW, here we also include the self-employed.



(a) Wealth (b) Earnings (c) Hours

(d) Income (e) Consumption

Figure 12: Fit to Inequality Data

inequality data together with the fact that it replicates moments of the household labor income process, which are key determinants of the optimal policy, we think our calibration strategy yields more relevant policy implications.

2. **Solution method.** AHHW argue that in Ramsey steady-state policy in the SIM model is independent of initial conditions. They then solve for this Ramsey steady state first and then solve backwards for the optimal transition. Our numerical method, described in Section 3.2 of the paper, does not rely on a characterization of the Ramsey steady state. We parameterize the paths of fiscal instruments in the time domain and chose the parameters to maximize welfare along the transition path directly. We discuss our perspective on the pros and cons of both methods in Section M.4.

M.1 Long-run Results: Method Comparison

In what follows, we argue that the differences between the results in the two papers come *solely from the differences in calibration* rather than from the solution method. To make this point we proceed in two steps. First, we extend the AHHW method to our economy with balanced-growth-path preferences and compute the long-run optimal fiscal policy and allocation using the algorithm described in Appendix M.9. We, then, compare it with the long-run policy and allocation obtained using our method. Further, we apply our method to the AHHW economy, i.e. the model with separable preferences and their calibration.[11] We then compare the long-run results from this experiment with the results reported by AHHW in their paper. The results are presented in Table 13.

[11]We thank the authors of Acikgoz, Hagedorn, Holter, and Wang (2018) for generously providing details on their calibration and method.

Table 12: Parameters for Acikgoz et al. (2018) calibration

Description	Parameter	Value
Preferences and Technology		
Preference curvature	σ	2.000
Labor Disutility	χ	13.397
Frisch elasticity	ϕ	1.000
Discount factor	β	0.939
Capital share	α	0.360
Depreciation rate	δ	0.080
Borrowing constraint	$\underline{a}$	0.000
Fiscal Policy		
Capital income tax (%)	τ^k	0.360
Labor income tax (%)	τ^n	0.280
Consumption tax (%)	τ^c	0.000
Government expenditure to GDP	G/Y	0.072
Debt to GDP	B/Y	0.619
Labor productivity process		
Persistence of AR(1)	ρ_ε	0.933
Standard deviation of AR(1)	σ_ε	0.302
Number of grid points	n	7
Range of Stds in Tauchen	m	3.000

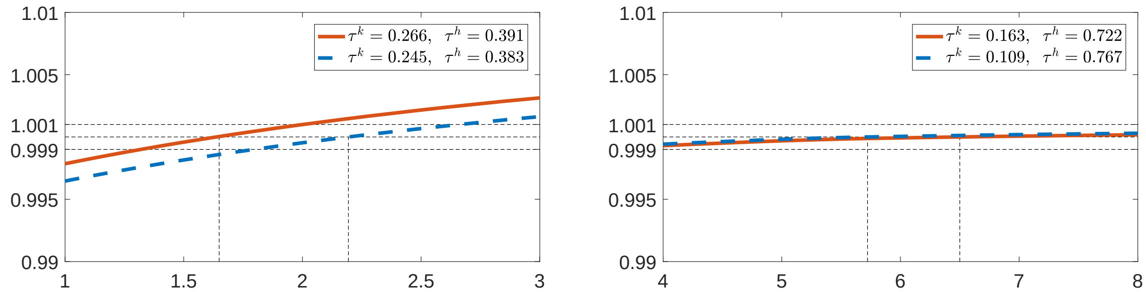
Table 13: Long-run Optimal Fiscal Policy: DP vs. AHHW Method

	τ^k	τ^h	T/Y	B/Y	K/Y	N	w	r	Welfare (%)
				DP Calibration					
DP Method	0.266	0.391	0.152	1.541	2.480	0.402	1.080	0.04817	3.518
AHHW Method	0.245	0.383	0.125	2.074	2.480	0.407	1.080	0.04819	3.501
				AHHW Calibration					
DP Method	0.163	0.722	0.025	6.508	2.434	0.284	1.056	0.06793	18.428
AHHW Method	0.109	0.767	0.084	5.597	2.473	0.271	1.065	0.06552	18.252

Note: The welfare gains for the AHHW method are obtained by imposing the long-run fiscal policy computed with that method and optimizing the transition towards it using our method. Appendices O.17 and O.18 present figures for the corresponding transition paths.

A couple of observations emerge from Table 13. For the DP calibration, both methods yield very similar long-run capital and labor income taxes. The long-run levels for capital, labor, and prices are also almost identical. The modified golden rule (MGR) implies an interest rate of 4.848 percent, so the deviations from it implied by both methods are of the same order. The only meaningful difference between the two methods is in the implied levels of long-run debt-to-output and transfers-to-output ratios. These differences have to do with the fact that in the long-run implied by the optimal policy there are not many households close to their borrowing constraints—see Figure 16g. As a result, the Ricardian equivalence nearly holds with respect to changes in the timing of lump-sum transfers. A lower level of lump-sum transfers combined with a higher level of debt (ceteris paribus) means that transfers have been front-loaded. So, the differences observed for these long-run transfer levels have to do not with the overall level of transfer but with their timing (Figure 68c), and since the Ricardian equivalence nearly holds this difference in timing is not very consequential to welfare. This is then problematic for both methods. For our method this is easy to see, the precision with which we can calculate each instrument is proportional to its effect on welfare. For the AHHW method, this near Ricardian equivalence lead to small effects of long-run debt-to-output on long-run interest rates. As a result, the modified golden rule, $\beta(1+r) = 1$, also does not pin down the long-run level of debt-to-output (and therefore transfer-to-output also) very precisely—see Figure 13a.

For the AHHW calibration the results obtained with the two methods differ more but are still similar. Relative to the results reported by AHHW we obtained higher capital income tax (0.163 vs. 0.109) and lower labor income tax (0.721 vs. 0.767). The long-run allocations and prices are also close. The largest discrepancies are again in the transfers-to-output ratio and debt-to-output ratios. The reasons for that are exactly the same as in the DP calibration case, except that the optimal policy implies even fewer households close to their borrowing constraints for this calibration—again, see Figure 16g. This magnifies the problem described in the previous



(a) DP Calibration (b) AHHW Calibration

Figure 13: $\beta(1 + r)$ vs. Debt-to-Output

Notes: Red solid curve: optimal taxes computed with our method; Blue dashed curve: optimal taxes computed with the AHHW method; Horizontal thin dashed lines: 0.999, 1.000, 1.001; Vertical thin dashed lines: debt-to-output that solves $\beta(1 + r) = 1$.

paragraph, which can be clearly appreciated by the flatness of the $\beta(1 + r)$ curves in Figure 13b.

M.1.1 Numerical Error Analysis

In this subsection, we provide an extensive analysis of all sources of numerical error and their implications for the long-run optimal policy. We begin by listing them for both methods.

In using the AHHW method to compute *long-run steady-state* optimal policy, numerical errors can arise from (see Algorithm 8 (Appendix M.9)):

1. Precision of approximation, via endogenous grid method, of the decision rules of households in Step 2 of the Algorithm 8.
2. Precision of approximation, via pdf iteration, of the stationary distribution over (a, e) in Step 2 of the Algorithm 8.
3. Precision of iterative procedure used to solve for the equilibrium prices in Step 2 of the Algorithm 8.
4. Precision of fixed point iteration on the policy function $q'(q, a, e)$ in Step 3 of the Algorithm 8.
5. Precision of approximation, via pdf iteration, of the stationary distribution $p(\lambda, a, e)$ in Step 6 of the Algorithm 8.
6. Precision of the procedure used to minimize the residuals is in Step 8 of the Algorithm 8.

In using our method to compute optimal policy *in transition and in the long-run steady state*, numerical errors can come from:

1. Precision of approximation, via endogenous grid method, of the decision rules of households in algorithm described in Step 1 of Algorithm 1 (Appendix D.1).
2. Precision of approximation, via pdf iteration, of the stationary distribution over (a, e) in Step 1 of the Algorithm 2 (Appendix D.2).

3. Precision of iterative procedure used to solve for the equilibrium prices in Step 1. of the Algorithm 2 (Appendix D.2).

4. Precision of transition computation, i.e. backward iteration of decision rules, forward iteration over distribution, in Step 6 of Algorithm 2 (Appendix D.2).

5. Precision of time-domain approximation of optimal paths by Chebyshev polynomials via equation D.5 (Appendix D.3).

6. Precision of the global optimization procedure described in Algorithm 3 (Appendix D.3).

Sources 1–3 are common to both methods, while sources 4–6 are different and not easily comparable. Also, it is important to notice that we are comparing here the sources of errors in our method used to solve for the optimal transition with the sources of errors in AHHW method used to solve for the optimal steady-state. Even though, we put a lot of effort into making the two methods as comparable as possible it is evident that they differ largely.

Stress test. The comparison between the long-run optimal policy computed using DP method and the AHHW method is an absolute stress test of our method. This is because, since in our method we directly maximize welfare, and the long-run is heavily discounted, simply due to the time preference of households, the long-run policy is the least precise outcome of our procedure. It would be fair to say that our method focuses relatively more in the short run (in proportion with the time preference of households) while the AHHW method focuses entirely on the long-run as that is the only thing we are computing with that their method (in Acikgoz et al. (2018) they also implement a backward iteration from the long-run policy to obtain optimal policy in transition, but we do not replicate that here). Thus, we view the fact that the optimal long-run policy is as close as it is with both methods as a great validation of the accuracy of our results. Given all the different sources of numerical error described above, and the fact that our algorithm focuses on the short run, the agreement between the long-run policy instrument we find is close to as good as one could hope for.

Policy instruments vs. residuals. We argue here that the policy instruments are a better way to judge the accuracy of our results relative to the residuals. First, here are the equations for the residuals as derived in M.8 :

$$[\bar{r}] : R_1 = 1 - \sum_{q,a,e} \left[\frac{\tilde{u}_c}{(1+\tau^c)} \frac{a}{\kappa} + q\tilde{u}_c + (q(1+\bar{r}) - q') \left(\tilde{u}_{cc} \frac{\partial \tilde{c}}{\partial \bar{r}} + \tilde{u}_{ch} \frac{\partial \tilde{h}}{\partial \bar{r}} \right) \right] \frac{p(q,a,e)}{A} - \frac{1}{A} \left[(w - \bar{w}) \frac{\partial \tilde{N}}{\partial \bar{r}} + \tau^c \frac{\partial \tilde{C}}{\partial \bar{r}} \right] \quad \text{(M.1)}$$

$$[\bar{w}] : R_2 = 1 - \sum_{q,a,e} \left[\frac{\tilde{u}_c}{(1+\tau^c)} \frac{e\tilde{h}}{\kappa} + (q(1+\bar{r}) - q') \left(\tilde{u}_{cc} \frac{\partial \tilde{c}}{\partial \bar{w}} + \tilde{u}_{ch} \frac{\partial \tilde{h}}{\partial \bar{w}} \right) \right] \frac{p(q,a,e)}{N} - \frac{1}{N} \left[(w - \bar{w}) \frac{\partial \tilde{N}}{\partial \bar{w}} + \tau^c \frac{\partial \tilde{C}}{\partial \bar{w}} \right] \quad \text{(M.2)}$$

$$[B] : R_3 = 1 - \beta(1+r). \quad \text{(M.3)}$$

In a nutshell, our implementation of the AHHW method chooses the long-run policy τ^k, τ^h, and r (where r determines B) to minimize these residuals. Importantly, however, we must deal with the fact that different policy instruments imply different household decisions, a different distribution of assets, and, therefore, a different joint distribution of Lagrange multipliers, assets, and productivities, $p(q, a, e)$.

Table 14: Residuals from the long-run, Ramsey steady state FOCs.

	R_1	R_2	R_3
DP Method	-0.4459	0.0383	0.0003
AHHW Method	0.0001	0.0008	0.0003

Note: DP Method: we evaluate the long-run policy $\tau^k = 0.266, \tau^h = 0.391, r = 0.04817$ found using DP method through the lens of the FOCs of the planner. AHHW Method: residuals for long-run policy $\tau^k = 0.245, \tau^h = 0.383, r = 0.04819$ found using Algorithm 8.

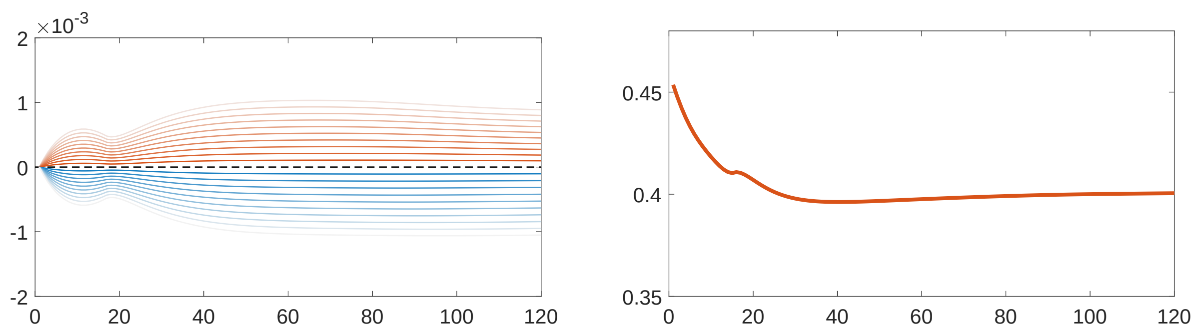
Correlated errors. Notice that the residuals R_1 and R_2 are affected, in particular, by the precision of the approximation of the three-dimensional distribution $p(q, a, e)$, which is specially difficult to compute precisely. Hence, these residuals reflect not only departures of policy instruments from the optimal ones but also the error associated with approximating the distribution $p(q, a, e)$. In Algorithm 8 (Appendix M.9) we now provide an extended discussion of how we construct this three dimensional grid, but any approximation would be prone to errors. Now, taking the grid over (q, a, e) as given, the AHHW method minimizes the residuals conditional on whatever approximation error is made in the computation of $p(q, a, e)$ and every other numerical error listed above. This makes it hard to interpret the residuals themselves, and one reason why we prefer looking directly at the policy instruments.

Comparison with Euler errors. There is fundamental difference between equations (M.1)–(M.3) and Euler errors. As mentioned above, the residuals R_1 and R_2 are affected by the precision of the approximation of $p(q, a, e)$, whereas to compute Euler errors we weight contingencies by their *exogenous* probabilities. The only endogenous object in the Euler equation is the policy function itself, and Euler errors are a good way to judge if the policy function satisfy that equation which otherwise involves only exogenous objects.

Next, suppose you know exactly what the policy function is, then one might already argue that it would be better to directly compare the policy function to the its true counterpart. Analogously, the objects characterized by equations (M.1)–(M.3) are the policy instruments τ^k, τ^h, and r, and if we could compare them to their true values that would be even more preferable since we bypass the endogeneity of $p(q, a, e)$ and the fact that every other numerical error affects the residuals of those equations. Now, there are two options: (1) the numerical errors in computing the optimal policy with the AHHW method are negligible and we can take the policy to be the true optimal policy in which case we think it is clearly preferable to compare policy; or (2) the numerical errors are not negligible in which case the residuals are hard to interpret for the reasons argued above. In the second case one should take both the policy obtained with our method and the one obtained with the AHHW method as numerical approximations and then we think the fact the methods yield policies as similar as they do, at least to some extent, validates the accuracy of those results.

Error accumulation. In this paragraph we argue our method is not prone to error accumulation as a result of forward iteration while solving for the optimal policy over transition. Our method globally approximates the time paths of instruments (using weighted families of Chebyshev polynomials). As we explain in Section 3.2, we pick the coefficient of these polynomials and convergence rate, λ^x, to maximize welfare. This implies that the path policy instruments is chose globally instead of having consecutive levels tied to each other through the planners FOCs. This independence frees us from the danger of error accumulation in the process of solving for the optimal policy. Of course this comes at a cost, which is that we have to make sure that the approximation is flexible enough to attain the global maximum. We provide robustness checks on this issue in Appendix G.3. One might also worry about error accumulation in computing welfare for any proposed policy path, as we have to iterate forward on government debt. To show this is not the case we conduct the following experiment.

We consider variations of the entire path of the optimal labor income taxes together with an appropriate adjustment to the path of lump-sum transfers (same in every period), so that the economy stays in equilibrium. We start with shifting down the labor income tax path by $1.0e-3$ and then smoothly, in steps of size $1.0e-4$, we move towards a path shifted up by $1.0e-3$, crossing the benchmark optimal path in the middle. We chose the magnitude of the step size of $1.0e-4$ so that it is two orders of magnitude larger than the precision with which we solve the transitional dynamics. The impact of this variation on debt-to-GDP ratio is presented in panel (A) of Figure 14. It presents differences between the optimal debt-to-GDP path and the paths in the economy subject to the tax variations. Red-colored paths are associated with paths of labor tax lower than the optimal one and blue-colored paths with those higher than the optimal one.



(a) Debt/GDP variation due to labor tax

(b) Labor tax base

Figure 14: The role of variation in labor income tax on government debt/GDP.

The time path of government debt following the tax variations reflects the time path of the labor tax base. Since, the labor tax base is falling over time, a uniform increase of labor income tax followed by uniform adjustment of the path of lump-sum transfers is reflected by a decreasing pattern of government debt, since the net effect is that government obtains more revenue in the short run and less in the long run. Second, once the tax base stabilizes over time around period 40 debt to GDP stabilizes too, and importantly it is not increasing.

Had it been increasing it would reflect the accumulation of the errors in forward iteration while computing transition. This is not the case, which is reassuring.

M.2 Transition: Comparison of Results from Both Calibrations

We have established that the differences in the long-run Ramsey allocations between AHHW and our paper result from significant differences between the calibrations, rather then numerical method used. We now proceed to compare the optimal transitions for both economies. We compute the optimal transition for AHHW economy implementing our method in exactly the same way as we describe in Section 3.2 of the paper. Figures 15 and 16 present the optimal fiscal policy instruments and some of the aggregates for the optimal transition in the AHHW economy and, for comparison, also the results for our economy. Appendix O.16 presents a more extensive list of figures.[12]

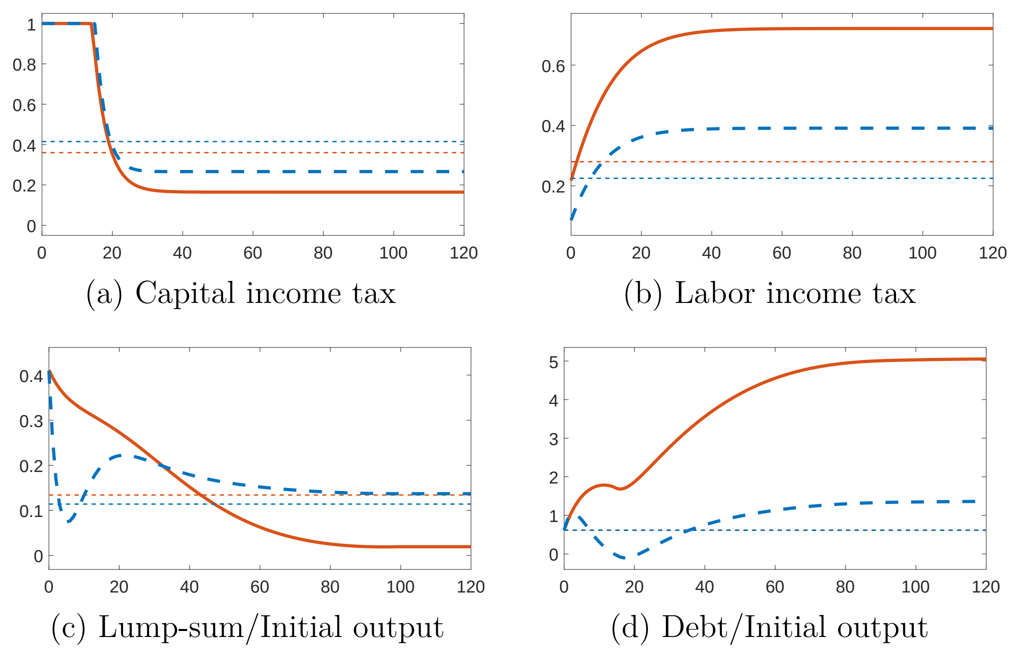


Figure 15: Optimal Fiscal Policy: Benchmark vs. AHHW Calibration

Notes: Red solid curve: optimal transition for calibration from Acikgoz et al. (2018); Blue dashed curve: optimal transition (benchmark); Thin dashed lines: corresponding values in initial stationary equilibrium.

Under both calibrations, the evolution of capital and labor income taxes over transition are qualitatively similar. Capital income taxes are front-loaded and settle at positive levels in the long run. Labor income taxes fall on impact and then increase monotonically towards their long-run levels. The long-run levels, however, of both instruments differ across calibrations, the AHHW economy settles at a lower capital income tax (by 10 percentage points) and higher labor income tax (by 32 percentage points). The lower long-run capital income taxes in AHHW can be understood by the lower level of risk faced by households in their economy: this reduces

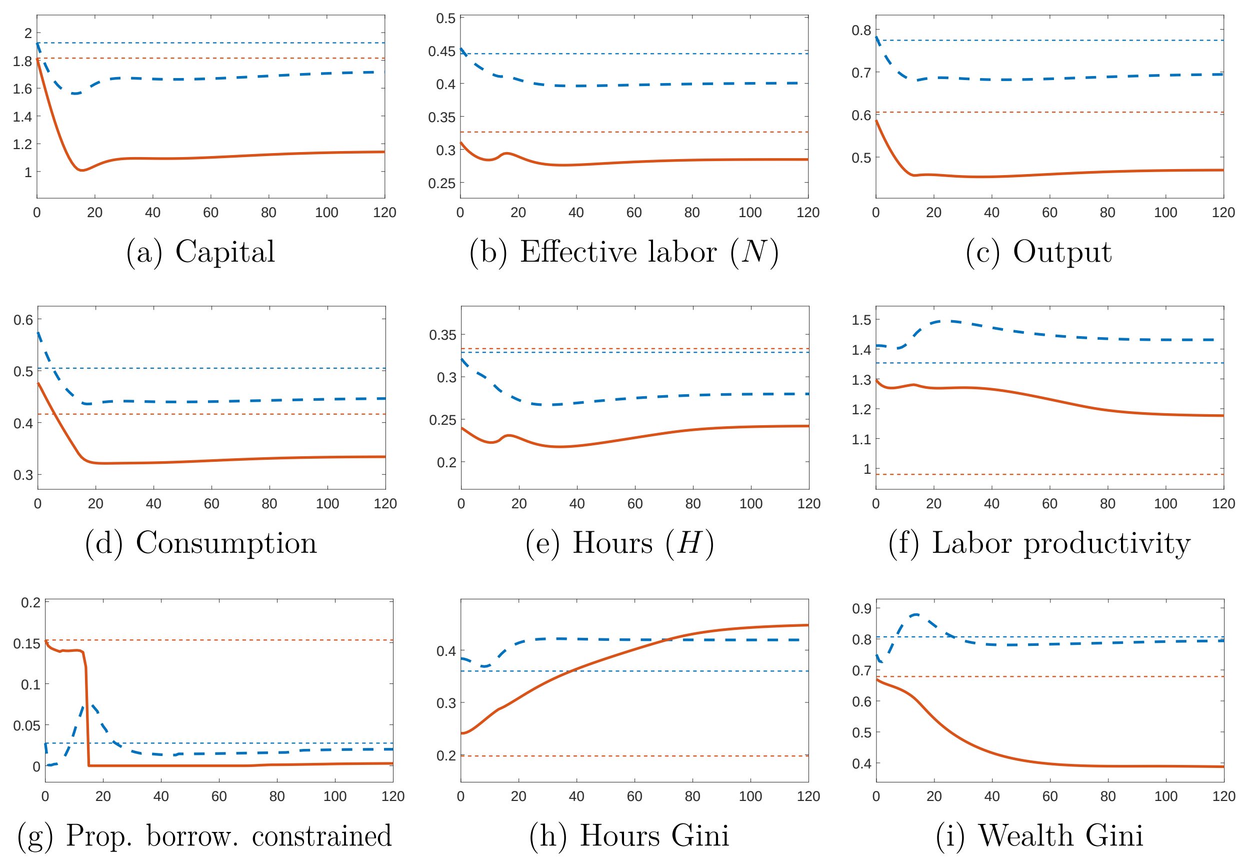
[12]Appendices O.17 and O.18 then present the figures for what happens when we impose the long-run results obtained with the AHHW method and otherwise use our method to solve for the transition, for both our calibration and theirs.

precautionary savings making savings decisions significantly more elastic (even with their slightly lower IES). More importantly, more insurance is obtained via the higher labor income taxes in the long run of the AHHW economy.

The substantially higher optimal labor income taxes in the AHHW economy are, to a large extent, a result of the strong wealth effects on labor supply implied by the separable utility function they use, see Section M.3 for quantification of the wealth effects in the two economies. The wealth effects on labor supply in Acikgoz et al. (2018) are strong enough that when the authors identify, with a decomposition experiment (see their Section 4.2.6), the effect of increasing labor income taxes from 28 to 76 percent, they find that it *increases* overall labor supply—it increases effective labor even more. This is a result of high productivity households, in particular, increasing their labor supply in response to lower after tax wages. It follows that the higher labor income taxes over the transition lead to a more than doubling of the Gini for hours—see Figure 16h—and to a significant increase in the average labor productivity—Figure 16f. So, labor income taxes in the AHHW economy is an especially effective instrument to provide redistribution and insurance. These effects are also present in the results for our calibration, but to a lower extent. Disciplining these wealth effects is, therefore, particularly important. Our calibration procedure does this by matching the distributions of hours, wealth, and earnings at the same time.

The patterns of lump-sum transfers and government debt differ across calibrations not only quantitatively but also qualitatively. In the AHHW economy, lump-sum transfers are more front-loaded, which implies a faster accumulation of government debt over transition. As a result, the debt-to-output ratio is more than four times higher in the final optimal steady state in the AHHW economy (1.54 vs. 6.51). To understand why this difference is so large, first consider the ways in which an increase in debt, resulting from the front-loading of lump-sum transfers (to keep thing simple), affects households: (1) front-loading lump-sum is desirable per se since households face borrowing constraints; (2) it crowds out capital which reduces output and, therefore, average consumption; (3) it reduces wages; and (4) it increases interest rates. Next, notice that all these effects are stronger the more households are close to their borrowing constraints. Figure 16g shows that, in the short run, the AHHW economy has significantly more households borrowing constrained. In period 17, the number of constrained households in the AHHW drops to zero and remains there indefinitely while for our calibration the number of constrained agents starts to drop though it never reaches zero.

Notice from Figure 15d that, before period 20, debt levels move in opposite directions: up for the AHHW economy and down for ours. After period 20, debt increases in both economies, but more in the AHHW economy. The first part is relatively easy to understand; the significantly higher level of constrained agents in the short run of the AHHW economy justifies more front-loading of lump-sum transfers, through effect (1). After period 20, both economies have very few households close to the constraint. Since both economies also display an optimal increasing path of debt following period 20, it must be beneficial to increase it. So, first, suppose that the net benefits from effects (1)–(4) in both economies are the same. Then, we would expect higher debt increases in the AHHW economy, since—again due to the lower number of constrained households—debt would have to be increased by more to achieve the same benefit. Finally, we attempt to compare the size of the different effects for the two economies, after period 20. Effect (1) is arguably stronger for our calibration since there are slightly more borrowing constrained households. Figure 16a shows that the crowding out of capital, effect (2), is not very relevant in either economy; capital actually increases after period 17 even with the large increases in debt in both economies—granted it could increase faster without the increase in debt. The same argument used above to explain why an increase in labor income taxes is more beneficial in the AHHW economy than in ours applies for the reduction in wages, effect (3). The fact that wealth inequality is substantially lower in the AHHW economy especially after period 20—see Figure 16i—makes the negative distributional effects of a higher interest rate, effect (4), less consequential in that economy than in ours. It is hard to determine in which economy the combined effect is stronger, so with this analysis we hope to convince the reader that it is reasonable to expect they are of similar magnitude and that is enough to justify the bigger increase in debt in the AHHW economy following period 20.



(a) Capital (b) Effective labor (N) (c) Output

(d) Consumption (e) Hours (H) (f) Labor productivity

(g) Prop. borrow. constrained (h) Hours Gini (i) Wealth Gini

Figure 16: Selected Aggregates: Benchmark vs. AHHW Calibration

Notes: Red solid curve: optimal transition for calibration from Acikgoz et al. (2018); Blue dashed curve: optimal transition (benchmark); Thin dashed lines: corresponding values in initial stationary equilibrium.

Though the optimal levels and some optimal patterns of fiscal instruments over time differ across calibrations, their impact on macroeconomic aggregates are similar. Figure 16 presents the evolution of these aggregates in both economies. Capital stock and effective labor fall over transition, implying a fall in output of similar magnitude across calibrations. In both economies the optimal policy implies front-loading of the aggregate consumption with initial increases of about 15 percent relative to the initial steady states. The optimal policies differ in terms of the response of hours worked, the AHHW economy experiences a much larger drop in labor supply relative to the initial steady-state. The labor productivity, on the other hand, rises more in the AHHW economy than in ours, which indicates more reallocation of labor towards productive agents. It follows that the complementarity between efficiency and redistribution that we highlight in the paper is even more pronounced in the AHHW economy.

M.3 Comparison of the Wealth Effects on Labor Supply

In this section, we illustrate the differences in the strength of wealth effects on labor supply between the DP and AHHW calibrations. To do so we compute and compare the following two statistics:

$$\Lambda_H = \frac{K+B}{H} \int_{A\times E} \frac{\partial h(a,e)}{\partial a} d\lambda\left(a,e\right),$$
$$\Lambda_N = \frac{K+B}{N} \int_{A\times E} \frac{\partial\left(eh(a,e)\right)}{\partial a} d\lambda\left(a,e\right),$$

where the notation follows the one in the paper. The first statistic, Λ_H, captures an aggregate wealth effect on hours worked following a marginal increase of asset positions at the household level. The second, Λ_N, captures an aggregate wealth effect on effective labor again following a marginal increase of asset positions at the household level. We normalize these aggregate wealth effects with the ratio of aggregate assets to aggregate hours worked for Λ_H, and with the ratio of aggregate assets to aggregate effective labor for Λ_N. This ensures comparability of the two statistics across economies featuring different levels of these aggregates. Both measures provide useful insight into the sensitivity of labor supply with respect to changes to the asset position of the households. Table 15 presents these measures computed for the DP and AHHW economies.

Table 15: Wealth effects: DP vs. AHHW economy

	Λ_H	Λ_N
DP Economy	-1.27	-0.45
AHHW Economy	-4.15	-0.94

Clearly the aggregate response of labor supply to changes in wealth is larger in the AHHW economy than in ours. In terms of hours worked, this response is more than three times larger in the AHHW economy, whereas in terms of effective labor it is two times larger. Importantly, the strength of these effects is largely driven by the distribution of assets and hours worked in the underlying economies. As we argue in Appendix M, our calibration fits these distributions better than the AHHW calibration. Given the lack of definitive direct empirical evidence on the strength of wealth effects, we think the simultaneous fit of earnings, asset, and hours distributions are an indirect way of disciplining the strength of wealth effects in our economy.

M.4 Remarks on the Methods

We think our method and the AHHW method of using first-order conditions (FOC) are complementary. Given the complexity of the methods, it can be hard to make sure the results are not affected by numerical error—or even by coding error. Applying both methods and finding consistent results is a great way of dealing with this. Nevertheless, if one has to pick one of the methods for solving a Ramsey problem we think the following aspects should be considered:

1. If one is interested only in long-run policy, if it is possible to establish that first-order conditions are necessary and sufficient, and that long-run policy is independent of initial conditions, and if the system of FOCs is well behaved, the FOC approach is likely to require significantly less computational power and should be preferred.

2. Based on the agreement of the results using both methods, we are confident the method in AHHW is indeed characterizing the optimal policy in the long run of our economy. However, we should note that, at the time of writing, sufficiency of the FOCs for the SIM model is yet to be formally established.

3. Even if the conditions of the first item are met, one should be careful with the known numerical issues associated with solving a system of, usually non-linear, equations implied by the FOCs. Small errors solving the system can hide big differences in the corresponding Ramsey policy, as is the case for the MGR in the AHHW calibration—see Figure 13b. If the system is not well behaved, it may actually be preferable to follow our method.

4. It can be hard to gauge from first-order conditions how quantitatively relevant they are. If deviations from these conditions lead to small changes in welfare, perhaps less weight should be put on using them to guide policy. Our approach forces us to focus on welfare relevant aspects of the optimal policy.

5. If one is interested in the short-run and medium-run policies and their effects, optimal transitions will eventually have to be computed. In that case, the relevant comparison is between our method and the backwards induction method of AHHW (which first computes the long-run policy then solves for previous period policy solving a system of FOCs for each period). The issue we see with the backwards induction method is the possibility of the numerical errors we discuss above propagating, as each step requires the solution of a potentially not-well-behaved system of equations. This is why we think our method—perhaps with the additional step of checking long-run FOCs—should be preferred, specially if the number of policy instruments is small enough to allow for robust global optimization. If there are too many policy instruments, one could restrict attention to policies that follow simple time patterns that can be described with few parameters, then using our method would still be feasible.

6. We do not think maximizing the objective function directly makes the algorithm more of a black box than when first-order conditions are used. We think neither method offers a direct economic rationale behind the numbers they produce for the optimal policy and one has to explain its properties independently of the approach used.

7. Finally, though consistency with first-order conditions is reassuring, if applied correctly, our method can be used to approximate the path of optimal policy in virtually any model in which one can compute transitions fast enough, even if first-order conditions are not tractable.

M.5 Extension of the AHHW Method to BGP preferences

This section extends the AHHW method of solving for the optimal policy in the long run for the balanced-growth-path preferences used in our paper. In the end, we present the numerical algorithm we use to solve the problem.

M.5.1 Environment

There is a measure one of households. Denote the household's history by $e^t = \left\{e^{t-1}, e_t\right\}$ with $e^0 = \{a_0, e_0\}$. Given a sequence of prices and taxes the household solves

$$V(a_0, e_0) = \max_{\{c_t(e^t), h_t(e^t), a_{t+1}(e^t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \sum_{e^t} \Pi\left(e^t\right) u\left(c_t\left(e^t\right), h_t\left(e^t\right)\right)$$

subject to

$$(1+\tau^c)\, c_t\left(e^t\right) + a_{t+1}\left(e^t\right) = \bar{w}_t e_t\left(e^t\right) h_t\left(e^t\right) + (1+\bar{r}_t)\, a_t\left(e^{t-1}\right) + T_t$$

$$a_{t+1}\left(e^t\right) \geq \underline{a}.$$

where

$$\bar{w}_t \equiv \left(1-\tau_t^h\right) w_t \quad \text{and} \quad \bar{r}_t \equiv \left(1-\tau_t^k\right) r_t.$$

Given prices, in each period, the representative firm solves

$$\max_{K_t, N_t} F(K_t, N_t) - w_t N_t - r_t K_t.$$

Government finances an exogenous stream of expenditure and lump-sum transfers with taxes on labor and capital or debt

$$G_t + T_t + r_t B_t = B_{t+1} - B_t + \tau^c C_t + \tau_t^h w_t N_t + \tau_t^k r_t (K_t + B_t).$$

M.5.2 Equilibrium

Definition 3 *Given K_0, B_0, an initial distribution $\Pi\left(e^0\right)$ and a policy $\pi \equiv \{\tau_t^k, \tau_t^h, T_t\}_{t=0}^{\infty}$, a* ***competitive equilibrium*** *is an allocation $\left\{\{c_t(e^t), h_t(e^t), a_{t+1}(e^t)\}_{e^t}, K_t, N_t, B_t\right\}_{t=0}^{\infty}$, a price system $P \equiv \{r_t, w_t\}_{t=0}^{\infty}$, such that for all t:*

1. *Given P and π, $\{c_t(e^t), h_t(e^t), a_{t+1}(e^t)\}_{e^t}$ solve the household's problem;*

2. *Factor prices are set competitively: $r_t = F_K(K_t, N_t)$, $w_t = F_N(K_t, N_t)$;*

3. *Government budget constraint holds and debt is bounded;*

4. *Markets clear,*

$$N_t = \sum_{e^t} \Pi\left(e^t\right) e_t h_t\left(e^t\right), \quad \textit{and} \quad K_t + B_t = \sum_{e^{t-1}} \Pi\left(e^{t-1}\right) a_t\left(e^{t-1}\right).$$

M.5.3 Characterization

First order conditions of the household's problem, assuming balanced-growth-path preferences:

$$\left(1+\tau^c\right) \tilde{c}_t\left(e^t\right) = \bar{w}_t e_t\left(e^t\right) \tilde{h}_t\left(e^t\right) + \left(1+\bar{r}_t\right) a_t\left(e^{t-1}\right) + T_t - a_{t+1}\left(e^t\right),$$

$$\tilde{h}_t\left(e^t\right) = \max\left(1 - \frac{1-\gamma}{\gamma} \frac{\left(1+\tau^c\right) \tilde{c}_t\left(e^t\right)}{\bar{w}_t e_t\left(e^t\right)}, 0\right),$$

$$\tilde{u}_c\left(e^t\right) \geq \beta\left(1+\bar{r}_{t+1}\right) \sum_{e^{t+1}} \Pi\left(e^{t+1} \mid e^t\right) \tilde{u}_c\left(e^{t+1}\right),$$

$$\left(a_{t+1}\left(e^t\right) - \underline{a}\right)\left(\tilde{u}_c\left(e^t\right) - \beta\left(1+\bar{r}_{t+1}\right) \sum_{e^{t+1}} \Pi\left(e^{t+1} \mid e^t\right) \tilde{u}_c\left(e^{t+1}\right)\right) = 0,$$

$$a_{t+1}\left(e^t\right) \geq \underline{a}.$$

Notice that the first two equations can be used to solve for:

$$\tilde{c}_t\left(e^t\right) = \max\left\{\frac{\gamma}{\left(1+\tau^c\right)}\left(\bar{w}_t e_t\left(e^t\right) + \left(1+\bar{r}_t\right) a_t\left(e^{t-1}\right) + T_t - a_{t+1}\left(e^t\right)\right), \right. \\ \left. \frac{1}{\left(1+\tau^c\right)}\left(\left(1+\bar{r}_t\right) a_t\left(e^{t-1}\right) + T_t - a_{t+1}\left(e^t\right)\right)\right\}, \tag{M.4}$$

$$\tilde{h}_t\left(e^t\right) = \max\left\{\gamma - (1-\gamma) \frac{\left(\left(1+\bar{r}_t\right) a_t\left(e^{t-1}\right) + T_t - a_{t+1}\left(e^t\right)\right)}{\bar{w}_t e_t\left(e^t\right)}, 0\right\}. \tag{M.5}$$

Factor prices are set competitively: $r_t = f_K(K_t, N_t)$, $w_t = f_N(K_t, N_t)$. Government budget constraint holds:

$$G_t + T_t + \left(1+\bar{r}_t\right) B_t + \bar{r}_t K_t + \bar{w}_t N_t = F\left(K_t, N_t\right) + \tau^c C_t + B_{t+1}.$$

Markets clear:

$$\tilde{N}_t = \sum_{e^t} \Pi\left(e^t\right) e_t \tilde{h}_t\left(e^t\right), \quad \tilde{K}_t = \sum_{e^{t-1}} \Pi\left(e^{t-1}\right) a_t\left(e^{t-1}\right) - B_t, \quad \text{and} \quad \tilde{C}_t = \sum_{e^t} \Pi\left(e^t\right) \tilde{c}_t\left(e^t\right). \tag{M.6}$$

M.6 Ramsey Problem

Given K_0, B_0, τ_0^k, τ_0^h, T_0, $\Pi\left(e^0\right)$ and a welfare function W, the **Ramsey problem** is to solve

$$\max_{\left\{\bar{w}_t, \bar{r}_t, T_t, B_{t+1}, a_{t+1}\left(e^t\right)\right\}} \sum_{t=0}^{\infty} \beta^t \sum_{e^t} \Pi\left(e^t\right) u\left(c_t\left(e^t\right), h_t\left(e^t\right)\right)$$

subject to

$$
\begin{aligned}
\tilde{u}_c\left(e^t\right) &\geq \beta\left(1+\bar{r}_{t+1}\right)\mathbb{E}\left[\tilde{u}_c\left(e^{t+1}\right) \mid e^t\right], \\
\left(a_{t+1}\left(e^t\right)-\underline{a}\right)&\Big(\tilde{u}_c\left(e^t\right)-\beta\left(1+\bar{r}_{t+1}\right)\mathbb{E}\left[\tilde{u}_c\left(e^{t+1}\right) \mid e^t\right]\Big)=0, \\
a_{t+1}\left(e^t\right) &\geq \underline{a}, \\
F\left(\tilde{K}_t,\tilde{N}_t\right)+B_{t+1}+\tau^c\tilde{C}_t &= G_t+T_t+\left(1+\bar{r}_t\right)B_t+\bar{r}_t\tilde{K}_t+\bar{w}_t\tilde{N}_t.
\end{aligned}
$$

We can set up the Lagrangian,

$$
\begin{aligned}
\mathcal{L}=\sum_{t=0}^{\infty}\beta^t\sum_{e^t}\Pi\left(e^t\right)\Big\{&\tilde{u}\left(e^t\right)+\theta_{t+1}\left(e^t\right)\left[\tilde{u}_c\left(e^t\right)-\beta\left(1+\bar{r}_{t+1}\right)\mathbb{E}\left[\tilde{u}_c\left(e^{t+1}\right) \mid e^t\right]\right] \\
&-\eta_{t+1}\left(e^t\right)\left[\left(a_{t+1}\left(e^t\right)-\underline{a}\right)\left(\tilde{u}_c\left(e^t\right)-\beta\left(1+\bar{r}_{t+1}\right)\mathbb{E}\left[\tilde{u}_c\left(e^{t+1}\right) \mid e^t\right]\right)\right]\Big\} \\
&+\sum_{t=0}^{\infty}\beta^t\kappa_t\Big[F\left(\tilde{K}_t,\tilde{N}_t\right)+B_{t+1}+\tau^c\tilde{C}_t-\Big(G_t+T_t+\left(1+\bar{r}_t\right)B_t+\bar{r}_t\tilde{K}_t+\bar{w}_t\tilde{N}_t\Big)\Big].
\end{aligned}
$$

Then,

$$
\begin{aligned}
\mathcal{L}=\sum_{t=0}^{\infty}\beta^t\sum_{e^t}\Pi\left(e^t\right)\Big\{&\tilde{u}\left(e^t\right)-\lambda_{t+1}\left(e^t\right)\left[\tilde{u}_c\left(e^t\right)-\beta\left(1+\bar{r}_{t+1}\right)\mathbb{E}\left[\tilde{u}_c\left(e^{t+1}\right) \mid e^t\right]\right]\Big\} \\
&+\sum_{t=0}^{\infty}\beta^t\kappa_t\Big[F\left(\tilde{K}_t,\tilde{N}_t\right)+B_{t+1}+\tau^c\tilde{C}_t-\Big(G_t+T_t+\left(1+\bar{r}_t\right)B_t+\bar{r}_t\tilde{K}_t+\bar{w}_t\tilde{N}_t\Big)\Big],
\end{aligned}
$$

with

$$
\lambda_{t+1}\left(e^t\right)\equiv\eta_{t+1}\left(e^t\right)\left(a_{t+1}\left(e^t\right)-\underline{a}\right)-\theta_{t+1}\left(e^t\right).
$$

And setting

$$
\lambda_0\left(e^{t-1}\right)\equiv 0,
$$

we get

$$
\begin{aligned}
\mathcal{L}=\sum_{t=0}^{\infty}\beta^t\sum_{e^t}\Pi\left(e^t\right)&\left[\tilde{u}\left(e^t\right)+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\tilde{u}_c\left(e^t\right)\right] \\
&+\sum_{t=0}^{\infty}\beta^t\kappa_t\left[F\left(\tilde{K}_t,\tilde{N}_t\right)+B_{t+1}-\Big(G_t+T_t+\left(1+\bar{r}_t\right)B_t+\bar{r}_t\tilde{K}_t+\bar{w}_t\tilde{N}_t-\tau_t^c\tilde{C}_t\Big)\right].
\end{aligned}
$$

M.7 First Order Conditions

Using (M.4), (M.5), and (M.6), we obtain

$$
\begin{aligned}
&[B_{t+1}]:\kappa_t=\beta\kappa_{t+1}\left(1+F_K\left(\tilde{K}_{t+1},\tilde{N}_{t+1}\right)\right) \\
&[T_t]:\sum_{e^t}\Pi\left(e^t\right)\left[\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial T_t}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial T_t}+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial T_t}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial T_t}\right)\right] \\
&\quad+\kappa_t\left[\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial T_t}-1+\tau^c\frac{\partial\tilde{C}_t}{\partial T_t}\right]=0 \\
&[\bar{r}_t]:\sum_{e^t}\Pi\left(e^t\right)\left[\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{r}_t}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{r}_t}+\lambda_t\left(e^{t-1}\right)\tilde{u}_c\left(e^t\right)+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{r}_t}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{r}_t}\right)\right]
\end{aligned}
$$

$$+\kappa_t\left[\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial\bar{r}_t}-A_t+\tau^c\frac{\partial\tilde{C}_t}{\partial\bar{r}_t}\right]=0$$

$$[\bar{w}_t]:\sum_{e^t}\Pi\left(e^t\right)\left[\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{w}_t}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{w}_t}+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{w}_t}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{w}_t}\right)\right]$$

$$+\kappa_t\left[\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial\bar{w}_t}-\tilde{N}_t+\tau^c\frac{\partial\tilde{C}_t}{\partial\bar{w}_t}\right]=0$$

$$[a_{t+1}\left(e^t\right)]:\beta^t\Pi\left(e^t\right)\left[\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}\right)\right]$$

$$+\beta^t\kappa_t\left(\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial a_{t+1}\left(e^t\right)}+\tau^c\frac{\partial\tilde{C}_t}{\partial a_{t+1}\left(e^t\right)}\right)+\beta^{t+1}\sum_{e^{t+1}}\Pi\left(e^{t+1}\right)\left[\tilde{u}_c\left(e^{t+1}\right)\frac{\partial\tilde{c}_{t+1}\left(e^{t+1}\right)}{\partial a_{t+1}\left(e^t\right)}+\tilde{u}_h\left(e^{t+1}\right)\frac{\partial\tilde{h}_{t+1}\left(e^{t+1}\right)}{\partial a_{t+1}\left(e^t\right)}\right.$$

$$\left.+\left(\lambda_{t+1}\left(e^t\right)\left(1+\bar{r}_{t+1}\right)-\lambda_{t+2}\left(e^{t+1}\right)\right)\left(\tilde{u}_{cc}\left(e^{t+1}\right)\frac{\partial\tilde{c}_{t+1}\left(e^{t+1}\right)}{\partial a_{t+1}\left(e^t\right)}+\tilde{u}_{ch}\left(e^{t+1}\right)\frac{\partial\tilde{h}_{t+1}\left(e^{t+1}\right)}{\partial a_{t+1}\left(e^t\right)}\right)\right]$$

$$+\beta^{t+1}\kappa_{t+1}\left(\left(F_N\left(\tilde{K}_{t+1},\tilde{N}_{t+1}\right)-\bar{w}_{t+1}\right)\frac{\partial\tilde{N}_{t+1}}{\partial a_{t+1}\left(e^t\right)}+\tau^c\frac{\partial\tilde{C}_{t+1}}{\partial a_{t+1}\left(e^t\right)}\right)+\beta^{t+1}\kappa_{t+1}\left[F_K\left(\tilde{K}_{t+1},\tilde{N}_{t+1}\right)-\bar{r}_{t+1}\right]\frac{\partial\tilde{K}_{t+1}}{\partial a_{t+1}\left(e^t\right)}=0.$$

Since T_t, $\bar{r}_t$, $\bar{w}_t$, $a_t\left(e^{t-1}\right)$, and $a_{t+1}\left(e^t\right)$ are choice variables, and again using (M.4), (M.5), we can implicitly differentiate this system to get, if $\tilde{h}_t\left(e^t\right)>0$,

$$\begin{aligned}
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial T_t}&=\frac{\gamma}{1+\tau^c},\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial T_t}=-\frac{(1-\gamma)}{\bar{w}_te_t\left(e^t\right)}\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{r}_t}&=\frac{\gamma}{1+\tau^c}a_t\left(e^{t-1}\right),\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{r}_t}=-\frac{(1-\gamma)}{\bar{w}_te_t\left(e^t\right)}a_t\left(e^{t-1}\right)\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{w}_t}&=\frac{\gamma}{1+\tau^c}e_t\left(e^t\right),\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{w}_t}=\frac{\gamma-\tilde{h}_t\left(e^t\right)}{\bar{w}_t}\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}&=\frac{\gamma}{1+\tau^c}\left(1+\bar{r}_t\right),\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}=-\frac{(1-\gamma)}{\bar{w}_te_t\left(e^t\right)}\left(1+\bar{r}_t\right)\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}&=-\frac{\gamma}{1+\tau^c},\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}=\frac{(1-\gamma)}{\bar{w}_te_t\left(e^t\right)}
\end{aligned}$$

and, otherwise,

$$\begin{aligned}
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial T_t}&=\frac{1}{1+\tau^c},\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial T_t}=0\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{r}_t}&=\frac{1}{1+\tau^c}a_t\left(e^{t-1}\right),\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{r}_t}=0\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{w}_t}&=0,\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{w}_t}=0\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}&=\frac{1}{1+\tau^c}\left(1+\bar{r}_t\right),\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}=0\\
\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}&=-\frac{1}{1+\tau^c},\quad\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}=0
\end{aligned}$$

so that[13]

$$\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial T_t}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial T_t}=\tilde{u}_c\left(e^t\right)\left(\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial T_t}-\frac{\bar{w}_te_t\left(e^t\right)}{(1+\tau^c)}\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial T_t}\right)=\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}$$

[13] That these equations hold for both cases above. For the second case, skip the middle equalities since the intratemporal condition does not hold.

$$\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{r}_t}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{r}_t}=\tilde{u}_c\left(e^t\right)\left(\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{r}_t}-\frac{\bar{w}_te_t\left(e^t\right)}{(1+\tau^c)}\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{r}_t}\right)=\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}a_t\left(e^{t-1}\right)$$

$$\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{w}_t}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{w}_t}=\tilde{u}_c\left(e^t\right)\left(\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{w}_t}-\frac{\bar{w}_te_t\left(e^t\right)}{(1+\tau^c)}\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{w}_t}\right)=\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}e_t\left(e^t\right)\tilde{h}_t\left(e^t\right)$$

$$\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}=\tilde{u}_c\left(e^t\right)\left(\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}-\frac{\bar{w}_te_t\left(e^t\right)}{(1+\tau^c)}\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_t\left(e^{t-1}\right)}\right)=\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}\left(1+\bar{r}_t\right)$$

$$\tilde{u}_c\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}+\tilde{u}_h\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}=\tilde{u}_c\left(e^t\right)\left(\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}-\frac{\bar{w}_te_t\left(e^t\right)}{(1+\tau^c)}\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}\right)=-\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}$$

and, therefore,

$$[B_{t+1}]:\kappa_t=\beta\kappa_{t+1}\left(1+F_K\left(\tilde{K}_{t+1},\tilde{N}_{t+1}\right)\right)$$

$$[T_t]:\sum_{e^t}\Pi\left(e^t\right)\left[\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial T_t}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial T_t}\right)\right]$$
$$+\kappa_t\left[\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial T_t}-1+\tau^c\frac{\partial\tilde{C}_t}{\partial T_t}\right]$$

$$[\bar{r}_t]:\sum_{e^t}\Pi\left(e^t\right)\left[\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}a_t\left(e^{t-1}\right)+\lambda_t\left(e^{t-1}\right)\tilde{u}_c\left(e^t\right)+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{r}_t}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{r}_t}\right)\right]$$
$$+\kappa_t\left[\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial\bar{r}_t}-A_t+\tau^c\frac{\partial\tilde{C}_t}{\partial\bar{r}_t}\right]$$

$$[\bar{w}_t]:\sum_{e^t}\Pi\left(e^t\right)\left[\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}e_t\left(e^t\right)\tilde{h}_t\left(e^t\right)+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial\bar{w}_t}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial\bar{w}_t}\right)\right]$$
$$+\kappa_t\left[\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial\bar{w}_t}-\tilde{N}_t+\tau^c\frac{\partial\tilde{C}_t}{\partial\bar{w}_t}\right]=0$$

$$[a_{t+1}\left(e^t\right)]:\beta^t\Pi\left(e^t\right)\left[-\frac{\tilde{u}_c\left(e^t\right)}{(1+\tau^c)}+\left(\lambda_t\left(e^{t-1}\right)\left(1+\bar{r}_t\right)-\lambda_{t+1}\left(e^t\right)\right)\left(\tilde{u}_{cc}\left(e^t\right)\frac{\partial\tilde{c}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}+\tilde{u}_{ch}\left(e^t\right)\frac{\partial\tilde{h}_t\left(e^t\right)}{\partial a_{t+1}\left(e^t\right)}\right)\right]$$
$$+\beta^t\kappa_t\left(\left(F_N\left(\tilde{K}_t,\tilde{N}_t\right)-\bar{w}_t\right)\frac{\partial\tilde{N}_t}{\partial a_{t+1}\left(e^t\right)}+\tau^c\frac{\partial\tilde{C}_t}{\partial a_{t+1}\left(e^t\right)}\right)$$
$$+\beta^{t+1}\sum_{e^{t+1}}\Pi\left(e^{t+1}\right)\left[\frac{\tilde{u}_c\left(e^{t+1}\right)}{(1+\tau^c)}\left(1+\bar{r}_{t+1}\right)+\left(\lambda_{t+1}\left(e^t\right)\left(1+\bar{r}_{t+1}\right)-\lambda_{t+2}\left(e^{t+1}\right)\right)\left(\tilde{u}_{cc}\left(e^{t+1}\right)\frac{\partial\tilde{c}_{t+1}\left(e^{t+1}\right)}{\partial a_{t+1}\left(e^t\right)}+\tilde{u}_{ch}\left(e^{t+1}\right)\frac{\partial\tilde{h}_{t+1}\left(e^{t+1}\right)}{\partial a_{t+1}\left(e^t\right)}\right)\right]$$
$$+\beta^{t+1}\kappa_{t+1}\left(\left(F_N\left(\tilde{K}_{t+1},\tilde{N}_{t+1}\right)-\bar{w}_{t+1}\right)\frac{\partial\tilde{N}_{t+1}}{\partial a_{t+1}\left(e^t\right)}+\tau^c\frac{\partial\tilde{C}_{t+1}}{\partial a_{t+1}\left(e^t\right)}\right)+\beta^{t+1}\kappa_{t+1}\left[F_K\left(\tilde{K}_{t+1},\tilde{N}_{t+1}\right)-\bar{r}_{t+1}\right]\frac{\partial\tilde{K}_{t+1}}{\partial a_{t+1}\left(e^t\right)}=0.$$

M.8 In Stationary Equilibrium

Optimality conditions:

$$[B_{t+1}]:1=\beta\left(1+r\right)$$

$$[T_t]:\kappa=\sum_{\lambda,a,e}\left[\frac{\tilde{u}_c}{(1+\tau^c)}+\left(\lambda\left(1+\bar{r}\right)-\lambda'\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial T}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial T}\right)\right]p\left(\lambda,a,e\right)+\kappa\left[\left(w-\bar{w}\right)\frac{\partial\tilde{N}}{\partial T}+\tau^c\frac{\partial\tilde{C}}{\partial T}\right]$$

$$[\bar{r}_t]:\kappa A=\sum_{\lambda,a,e}\left[\frac{\tilde{u}_c}{(1+\tau^c)}a+\lambda\tilde{u}_c+\left(\lambda\left(1+\bar{r}\right)-\lambda'\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial\bar{r}}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial\bar{r}}\right)\right]p\left(\lambda,a,e\right)+\kappa\left[\left(w-\bar{w}_t\right)\frac{\partial\tilde{N}}{\partial\bar{r}}+\tau^c\frac{\partial\tilde{C}}{\partial\bar{r}}\right]$$

$$[\bar{w}_t]:\kappa N=\sum_{\lambda,a,e}\left[\frac{\tilde{u}_c}{(1+\tau^c)}e\tilde{h}+\left(\lambda\left(1+\bar{r}\right)-\lambda'\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial\bar{w}}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial\bar{w}}\right)\right]p\left(\lambda,a,e\right)+\kappa\left[\left(w-\bar{w}\right)\frac{\partial\tilde{N}}{\partial\bar{w}}+\tau^c\frac{\partial\tilde{C}}{\partial\bar{w}}\right]$$

$$[a_{t+1}]: -\left(\lambda\left(1+\bar{r}\right)-\lambda'\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)-\kappa\left(\left(w-\bar{w}\right)e\frac{\partial\tilde{h}}{\partial a'}+\tau^{c}\frac{\partial\tilde{c}}{\partial a'}\right)=\beta\mathbb{E}\left[\left(\lambda'\left(1+\bar{r}\right)-\lambda''\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]$$

$$+\beta\kappa\left(\left(w-\bar{w}\right)\mathbb{E}\left[e'\frac{\partial\tilde{h}'}{\partial a'}|e\right]+\tau^{c}\mathbb{E}\left[\frac{\partial\tilde{c}'}{\partial a'}\right]\right)+\beta\kappa\left[r-\bar{r}\right].$$

Working with the last equation defining $q=\lambda/\kappa$, and using the households' Euler equation, we obtain

$$q'=\frac{-q\left(1+\bar{r}\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)-\beta\mathbb{E}\left[-q''\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]-\beta\left[r-\bar{r}\right]-\left(\left(w-\bar{w}\right)e\frac{\partial\tilde{h}}{\partial a'}+\tau^{c}\frac{\partial\tilde{c}}{\partial a'}\right)-\beta\left(\left(w-\bar{w}\right)\mathbb{E}\left[e'\frac{\partial\tilde{h}'}{\partial a'}|e\right]+\tau^{c}\mathbb{E}\left[\frac{\partial\tilde{c}'}{\partial a'}\right]\right)}{-\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[\left(1+\bar{r}\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}.$$

Notice that, if $q'(q)=b_0+b_1q$, then it follows that

$$q'=\frac{\beta\mathbb{E}\left[\left(b_0'+b_1'b_0\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]-\beta\left[r-\bar{r}\right]-\left(\left(w-\bar{w}\right)e\frac{\partial\tilde{h}}{\partial a'}+\tau^{c}\frac{\partial\tilde{c}}{\partial a'}\right)-\beta\left(\left(w-\bar{w}\right)\mathbb{E}\left[e'\frac{\partial\tilde{h}'}{\partial a'}|e\right]+\tau^{c}\mathbb{E}\left[\frac{\partial\tilde{c}'}{\partial a'}\right]\right)}{-\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[\left(1+\bar{r}\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}$$
$$+\frac{-\left(1+\bar{r}\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[b_1'b_1\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}{-\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[\left(1+\bar{r}\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}q,$$

and, therefore, from the Contraction Mapping Theorem it follows that q' is linear in q, and we have the following recursive formulas for its coefficient

$$b_0=\frac{\beta\mathbb{E}\left[\left(b_0'+b_1'b_0\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]-\beta\left[r-\bar{r}\right]-\left(\left(w-\bar{w}\right)e\frac{\partial\tilde{h}}{\partial a'}+\tau^{c}\frac{\partial\tilde{c}}{\partial a'}\right)-\beta\left(\left(w-\bar{w}\right)\mathbb{E}\left[e'\frac{\partial\tilde{h}'}{\partial a'}|e\right]+\tau^{c}\mathbb{E}\left[\frac{\partial\tilde{c}'}{\partial a'}\right]\right)}{-\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[\left(1+\bar{r}\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}$$
$$b_1=\frac{-\left(1+\bar{r}\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[b_1'b_1\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}{-\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[\left(1+\bar{r}\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]},$$

or

$$b_0=\frac{\beta\mathbb{E}\left[b_0'\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]-\beta\left[r-\bar{r}\right]-\left(\left(w-\bar{w}\right)e\frac{\partial\tilde{h}}{\partial a'}+\tau^{c}\frac{\partial\tilde{c}}{\partial a'}\right)-\beta\left(\left(w-\bar{w}\right)\mathbb{E}\left[e'\frac{\partial\tilde{h}'}{\partial a'}|e\right]+\tau^{c}\mathbb{E}\left[\frac{\partial\tilde{c}'}{\partial a'}\right]\right)}{-\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[\left(1+\bar{r}-b_1'\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}$$
$$b_1=\frac{-\left(1+\bar{r}\right)\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)}{-\left(\tilde{u}_{cc}\frac{\partial\tilde{c}}{\partial a'}+\tilde{u}_{ch}\frac{\partial\tilde{h}}{\partial a'}\right)+\beta\mathbb{E}\left[\left(1+\bar{r}-b_1'\right)\left(\tilde{u}'_{cc}\frac{\partial\tilde{c}'}{\partial a'}+\tilde{u}'_{ch}\frac{\partial\tilde{h}'}{\partial a'}\right)|e\right]}.$$

M.9 Algorithm to solve for the Ramsey steady-state

Algorithm 8 *The algorithm to solve for the Ramsey steady-state is as follows:*

1. *Guess* τ^k, τ^h, *and* r.

2. *Compute the associated stationary equilibrium. In particular, obtain the policy functions* $a'(a,e)$ *and* $h(a,e)$ *(which give* $\tilde{u}_c(a,e)$, $\tilde{u}_{cc}(a,e)$, *and* $\tilde{u}'_{cc}(a,e)$*), prices* r, *and* w, *and aggregates* $\tilde{N}$, $\tilde{K}$, $\tilde{C}$, A, *and* B.

3. *Use equation* $[a_{t+1}]$ *to find* $b_0(a,e)$ *and* $b_1(a,e)$ *in the iterative procedure from two relationships below:*

$$b_0 = \frac{\beta \mathbb{E}\left[b_0'\left(\tilde{u}_{cc}'\frac{\partial \tilde{c}'}{\partial a'} + \tilde{u}_{ch}'\frac{\partial \tilde{h}'}{\partial a'}\right)|e\right] - \beta\left[r - \bar{r}\right] - \left((w-\bar{w})\, e\frac{\partial \tilde{h}}{\partial a'} + \tau^c\frac{\partial \tilde{c}}{\partial a'}\right) - \beta\left((w-\bar{w})\,\mathbb{E}\left[e'\frac{\partial \tilde{h}'}{\partial a'}|e\right] + \tau^c\mathbb{E}\left[\frac{\partial \tilde{c}'}{\partial a'}\right]\right)}{-\left(\tilde{u}_{cc}\frac{\partial \tilde{c}}{\partial a'} + \tilde{u}_{ch}\frac{\partial \tilde{h}}{\partial a'}\right) + \beta\mathbb{E}\left[\left(1+\bar{r}-b_1'\right)\left(\tilde{u}_{cc}'\frac{\partial \tilde{c}'}{\partial a'} + \tilde{u}_{ch}'\frac{\partial \tilde{h}'}{\partial a'}\right)|e\right]}$$

$$b_1 = \frac{-\left(1+\bar{r}\right)\left(\tilde{u}_{cc}\frac{\partial \tilde{c}}{\partial a'} + \tilde{u}_{ch}\frac{\partial \tilde{h}}{\partial a'}\right)}{-\left(\tilde{u}_{cc}\frac{\partial \tilde{c}}{\partial a'} + \tilde{u}_{ch}\frac{\partial \tilde{h}}{\partial a'}\right) + \beta\mathbb{E}\left[\left(1+\bar{r}-b_1'\right)\left(\tilde{u}_{cc}'\frac{\partial \tilde{c}'}{\partial a'} + \tilde{u}_{ch}'\frac{\partial \tilde{h}'}{\partial a'}\right)|e\right]}.$$

4. *Construct a grid* G_q *on* q *as follows:*
   (a) *Set the number of grid point* n_q *and the bounds of the grid* $q_{min} = \min_{a,e} b_0(a,e)/(1-b_1(a,e))$ *and* $q_{max} = \min_{a,e} b_0(a,e)/(1-b_1(a,e))$.
   (b) *Construct the left part of the grid. Start with constructing a linear grid of size* $n_q/2$ *on the interval* $[0, |q_{min}|]$, *denote it by* $\mathbf{q_{aux}}$. *Then set the left part of the grid:* $\mathbf{q_L} = -\mathbf{q_{aux}}$. *Reverse the array* q_L.
   (c) *Construct the right part of the grid. Start with constructing a linear grid of size* $n_q/2$ *on the interval* $[0, q_{max}]$, *denote it by* $\mathbf{q_{aux}}$. *Then set the right part of the grid:* $\mathbf{q_R} = \mathbf{q_{aux}}$.
   (d) *Put together the two parts of the grid in one array and eliminate duplicate points i.e. set* $G_q = \mathbf{q_L} \cup \mathbf{q_R}$.
5. *Construct the policy function* $q'(q,a,e) = b_0(a,e) + b_1(a,e)q$ *on the grid* $G_q \times G_a \times G_e$ *using* $b_0(a,e)$ *and* $b_1(a,e)$.
6. *Using policy functions* $q'(q,a,e)$ *and* $a'(a,e)$ *compute the stationary distribution* $p(q,a,e)$.
7. *Use equation* $[T_t]$ *to find* κ,

$$\kappa = \frac{\sum_{q,a,e}\left[\frac{\tilde{u}_c}{(1+\tau^c)}\right] p\left(q,a,e\right)}{1 - \sum_{q,a,e}\left[\left(q\left(1+\bar{r}\right) - q'\right)\left(\tilde{u}_{cc}\frac{\partial \tilde{c}}{\partial T} + \tilde{u}_{ch}\frac{\partial \tilde{h}}{\partial T}\right)\right] p\left(q,a,e\right) - \left[\left(w-\bar{w}\right)\frac{\partial \tilde{N}}{\partial T} + \tau^c\frac{\partial \tilde{C}}{\partial T}\right]}.$$

8. *Compute the residuals of equations* $[\bar{r}_t]$, $[\bar{w}_t]$, *and* $[B_{t+1}]$ *that is,*

$$[\bar{r}_t] : R_1 = 1 - \sum_{q,a,e}\left[\frac{\tilde{u}_c}{(1+\tau^c)}\frac{a}{\kappa} + q\tilde{u}_c + \left(q\left(1+\bar{r}\right) - q'\right)\left(\tilde{u}_{cc}\frac{\partial \tilde{c}}{\partial \bar{r}} + \tilde{u}_{ch}\frac{\partial \tilde{h}}{\partial \bar{r}}\right)\right]\frac{p\left(q,a,e\right)}{A} - \frac{1}{A}\left[\left(w-\bar{w}\right)\frac{\partial \tilde{N}}{\partial \bar{r}} + \tau^c\frac{\partial \tilde{C}}{\partial \bar{r}}\right] \tag{M.7}$$

$$[\bar{w}_t] : R_2 = 1 - \sum_{q,a,e}\left[\frac{\tilde{u}_c}{(1+\tau^c)}\frac{e\tilde{h}}{\kappa} + \left(q\left(1+\bar{r}\right) - q'\right)\left(\tilde{u}_{cc}\frac{\partial \tilde{c}}{\partial \bar{w}} + \tilde{u}_{ch}\frac{\partial \tilde{h}}{\partial \bar{w}}\right)\right]\frac{p\left(q,a,e\right)}{N} - \frac{1}{N}\left[\left(w-\bar{w}\right)\frac{\partial \tilde{N}}{\partial \bar{w}} + \tau^c\frac{\partial \tilde{C}}{\partial \bar{w}}\right] \tag{M.8}$$

$$[B_{t+1}] : R_3 = 1 - \beta(1+r). \tag{M.9}$$

9. *If the residuals are small enough, then stop. Otherwise, update* τ^k, τ^h, *and* r *using some minimization procedure and go back to step 2.*

N Alternative Calibrations

In this appendix, we present three alternative calibration strategies which we use to discuss the dependence of the results on the set of statistics used to discipline the model and to address some of the concerns raised by the referees. We start with the calibration strategy used by Aiyagari and McGrattan (1998). Then, we move to what we call a No-Inequality-Targets calibration, in which we drop the targets associated with the cross-sectional distributions of wealth, earnings and hours. Finally, we discuss the Return Risk calibration in which we introduce the iid shock to the interest rate to account for the heterogeneity in the asset returns.

N.1 Aiyagari and McGrattan (1998) Calibration

We analyze two economies based on the calibration used in Aiyagari and McGrattan (1998). In the first one, we replicate exactly their numbers. In the second one, we drop the growth rate of technology and reparametrize the economy to hit the same targets. Table 16 presents the parameters for the former economy, replicating the original Aiyagari and McGrattan (1998) numbers and results. There are three parameters in the version without growth that differ which do not affect much the model moments. They are $\beta = 0.968$, $T = 0.035$ and $\tau = 0.391$. We use the exactly replica of Aiyagari and McGrattan (1998) to explain the differences in the optimal, long-run government debt levels between our benchmark calibration and their paper. We use the version without technology growth to compute the Ramsey policy and compare it to our main results.

Table 16: Parameters for Aiyagari and McGrattan (1998) calibration

Description	Parameter	Value
Preferences and Technology		
Consumption share	γ	0.328
Preference curvature	σ	1.500
Discount factor	β	0.988
Capital share	α	0.300
Depreciation rate	δ	0.075
Borrowing constraint	$\underline{a}$	0.000
Technology growth rate	g	0.018
Fiscal Policy		
Total income tax (%)	τ	0.376
Transfers	T	0.082
Debt to GDP	B/Y	0.667
Labor productivity process		
Persistence of AR(1)	ρ_ε	0.600
Standard deviation of AR(1)	σ_ε	0.300
Number of grid points	n	7
Range of Stds in Tauchen	m	3.000

Figures 17 and 18 present the fit of both versions of the Aiyagari and McGrattan (1998) calibration to the cross-sectional distributions and compare them to our benchmark calibration. They share few characteristics. While dispersion in hours worked is relatively close to the data, this calibration strategy largely underestimates the dispersion in wealth and consumption. It also produces a dispersion in earnings and income much lower than the one observed in the data. As we argue below these disparities are crucial for understanding the differences between our results and theirs on the optimal long-run debt and also for understanding the contrasting Ramsey policies that result.

N.1.1 Maximizing Steady State: Optimal Debt

The results from Section L in the paper contrasts sharply with the ones in Aiyagari and McGrattan (1998). They run a similar experiment with some important differences: (1) in their model the only tax available to the planner is a total income tax, (2) they have long run technological growth, and (3) their calibration strategy for the labor income process focuses on matching the auto-correlation and variance of labor income without targeting distributional moments. They find that the government, even though it could costlessly choose any level of debt-to-output, chooses a level very close to the actual level in the US data at the time, around 67

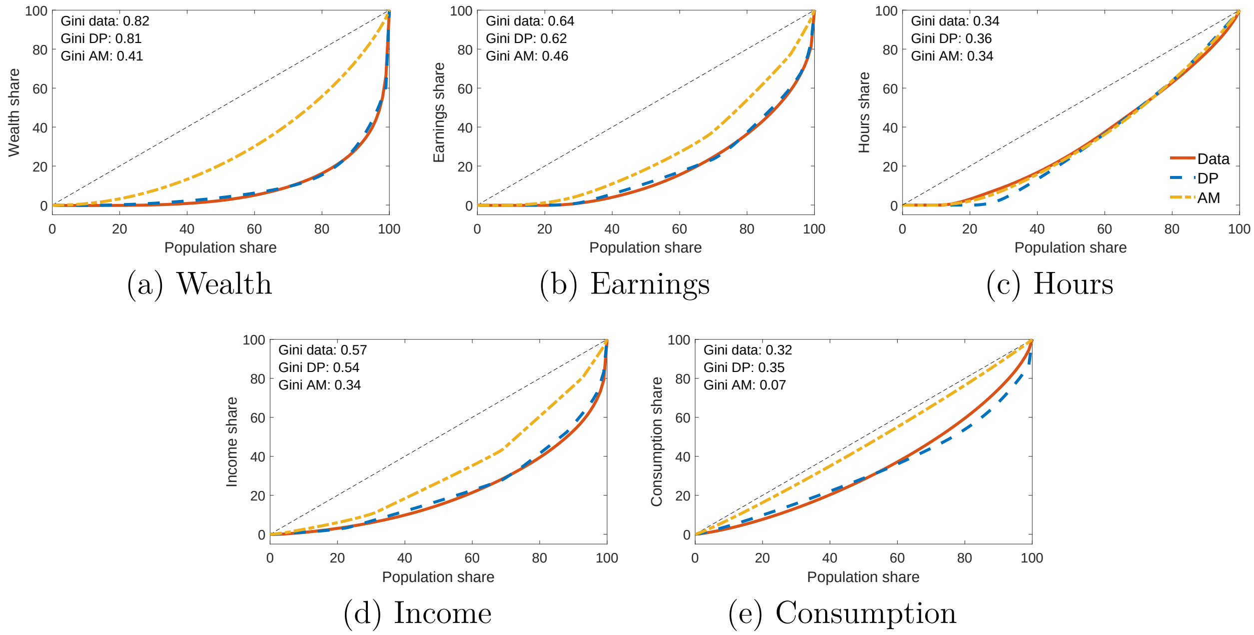


Figure 17: Aiyagari and McGrattan (1998) Calibration: Fit to Inequality Data

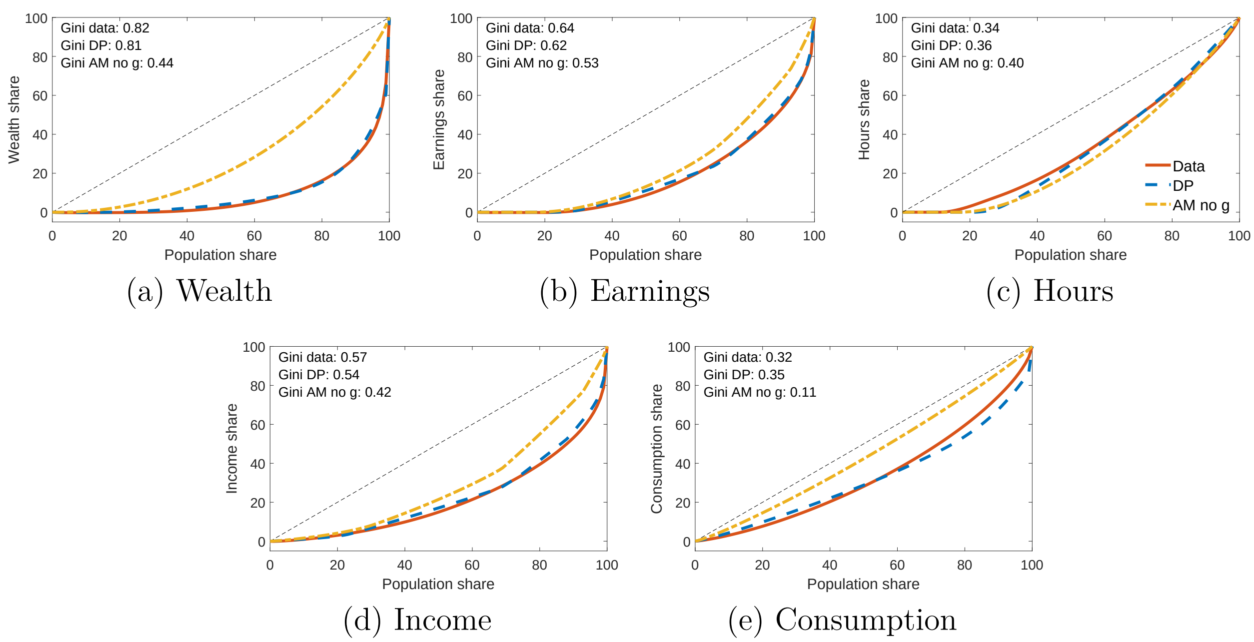


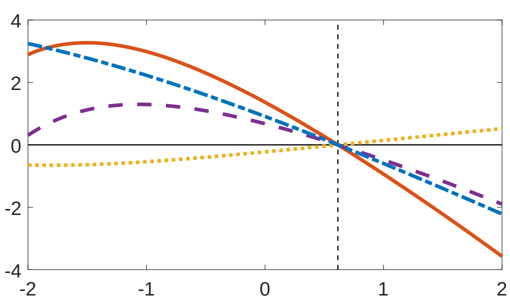
Figure 18: Aiyagari and McGrattan (1998) Calibration without Growth: Fit to Inequality Data

percent. In fact, they show that the welfare function is relatively flat with respect to the choice of debt-to-output. The main reason for the starkly different result is the difference in calibration.[14] To make this point, we modify our model in two ways: (1) we set capital and labor income taxes equal to each other, and (2) we introduce long-run technological growth. Then, we replicate the experiment in Aiyagari and McGrattan

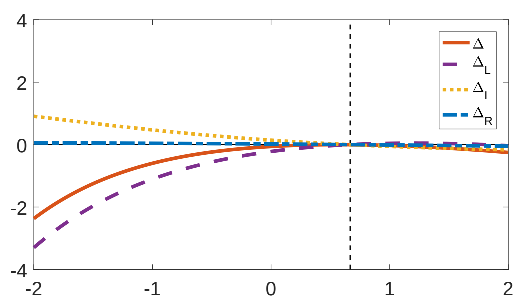
[14] Röhrs and Winter (2017) reach a similar conclusion, though they consider different fiscal instruments which we are purposefully abstracting from here.

(1998). Figure 19a displays the average welfare gains and its components for different levels of debt-to-gdp. For comparison, Figure 19b replicates the results for the exact calibration in Aiyagari and McGrattan (1998).

There are many interesting qualitative differences between the two figures: First, the average welfare gains in Aiyagari and McGrattan (1998) are flatter and peak at a positive debt-to-gdp level. Second, the level effect in Figure 19a follows quite closely the level of distortionary total-income taxes; higher levels of debt must be financed with higher taxes which reduce the level effect. When the government holds a lot of assets, however, at some point increasing it further decreases interest rates to such a degree that government asset income is reduced and taxes must increase. This is, to a large extent, what determines the optimal debt-to-gdp in Figure 19a. The mechanism highlighted by Aiyagari and McGrattan (1998) leads to a counteracting force: a higher level of government debt increases interest rates which incentivizes households to move away from their borrowing constraints. This effect reduces average distortions in the intertemporal margin and leads to a slightly increasing level for positive levels of debt-to-gdp in Figure 19b. Third, the insurance effect is increasing in debt-to-gdp in our calibration, which makes sense since it leads to higher interest rates and lower wages reducing the proportion of the households' income that is risky. As it turns out, however, in both economies the equilibrium lump-sum level is decreasing in the level of debt-to-gdp which reduces a part of the households' income that is certain. In Aiyagari and McGrattan (1998) it is this second effect that dominates and leads to a decreasing insurance effect. Finally, the redistribution effect is of a significantly higher magnitude with our calibration. This is simply because the calibration in Aiyagari and McGrattan (1998) leads to much less inequality (as seen in Figure 17) and, therefore, lower gains from reducing it. The common factor in all these differences is that the details of the calibration are crucial for virtually every aspect of the determinants of optimal policy.



(a) Benchmark with $\tau^k = \tau^h$ and growth



(b) Aiyagari and McGrattan (1998)

Figure 19: Welfare decomposition versus debt-to-gdp in steady state

Note: The variable in the x-axis is the debt-to-gdp in steady state; the thin dashed vertical line marks the level of debt-to-gdp in the initial stationary equilibrium, versus which the welfare changes are calculated.

N.1.2 Ramsey Policy in Aiyagari and McGrattan (1998) Economy

In Appendix O.13 we present an extensive list of figures for the Ramsey policy computed for the Aiyagari and McGrattan (1998) (AM) calibration and, for comparison, also our benchmark-calibration results—both approximations are computed with 8 parameters for comparability.

The results contrast starkly. The capital income tax stays at the upper bound for 3 years, much less than in our benchmark economy (14 years). This comes as no surprise though, given the lower level of wealth inequality in the AM calibration relative to the data. Also, long-run optimal capital income taxes differ substantially between the two calibrations, i.e. 6.4 vs. 34.3 percent. This difference has to do with larger need of the Ramsey planner to insure households in the long-run of our benchmark economy. Notice, in Figure 58t, that the variance of the growth rate of $c^\gamma(1-h)^{1-\gamma}$—which the measure of risk that matters for welfare—is very close between the two economies even though the long-run taxes system is very different.

Optimal labor income taxes follow similar time patterns for both economies, i.e. they are increasing over the period when the bound on the capital income taxes binds and constant afterwards. The major difference is in the levels of labor income taxes. The long-run labor income taxes for AM calibration is 7 percent versus 40 percent in our benchmark calibration. The calibration used in AM implies less before-tax income risk relative to our benchmark and hence it is less desirable to insure households over the optimal transition through taxing uncertain labor income.

The two calibration strategies imply opposing patterns for government debt and lump-sum transfer. The Ramsey planner in AM economy front-loads lump-sum more heavily at the expense of issuing massive amounts of government debt, reaching 567 percent of GDP in the long-run. The welfare gains associated with Ramsey policy in AM are of 8.9 percent, which is almost entirely driven by the level effect associated with the optimal policy. Table 17 presents the welfare decomposition and sheds more light on sources of welfare improvement.

Table 17: Alternative calibrations: long-run optimal policy and welfare decomposition

	t^*	τ^k	τ^h	T/Y	B/Y	K/Y	Δ	Δ_L	Δ_I	Δ_R
AM calibration	3	6.4	6.8	−27.7	567.2	2.81	8.9	10.8	−2.1	0.4
No Inequality Targets	13	27.2	62.0	23.3	155.0	1.89	26.9	3.1	11.1	10.7
Benchmark (8 parameters)	14	34.3	40.2	21.2	28.5	2.48	3.4	0.1	0.3	3.0

Note: All values, except for K/Y, are in percentage points.

N.2 No-Inequality-Targets (NIT) Calibration

Under this calibration strategy we consider the economy in which do not impose any cross-sectional distributions on the model, but rather stick only to the properties of the earnings process and macroeconomic aggregates. To make sure that we do not work with an under-identified system (less targets than parameters) we depart from our benchmark calibration strategy and we model the income process as a mixture of normal distributions as follows:

$$\log e' = \rho_e \log e + \varepsilon,$$

Table 18: Parameters for No-Inequality-Targets Calibration

Description	Parameter	Value
Preferences and Technology		
Consumption share	γ	0.886
Preference curvature	σ	1.608
Discount factor	β	0.905
Capital share	α	0.378
Depreciation rate	δ	0.104
Borrowing constraint	$\underline{a}$	-0.003
Fiscal Policy		
Capital income tax (%)	τ^k	0.415
Labor income tax (%)	τ^n	0.225
Consumption tax (%)	τ^c	0.047
Transfers	T	0.181
Government expenditure	G	0.615
Labor productivity process		
Persistence of mixture of Normals	ρ_ε	0.651
Probability of drawing from Normal 1	p	0.696
Mean of Normal 1	μ_1	1.803
Std of Normal 1	σ_1	0.356
Std of Normal 2	σ_2	0.218
Number of grid points	n	8
Range of Stds in Tauchen	m	2.000

where

$$\epsilon \sim \begin{cases} N\left(\mu_1, \sigma_1\right) & \text{with probability} \quad p, \\ N\left(\mu_2, \sigma_2\right) & \text{with probability} \quad 1-p. \end{cases}$$

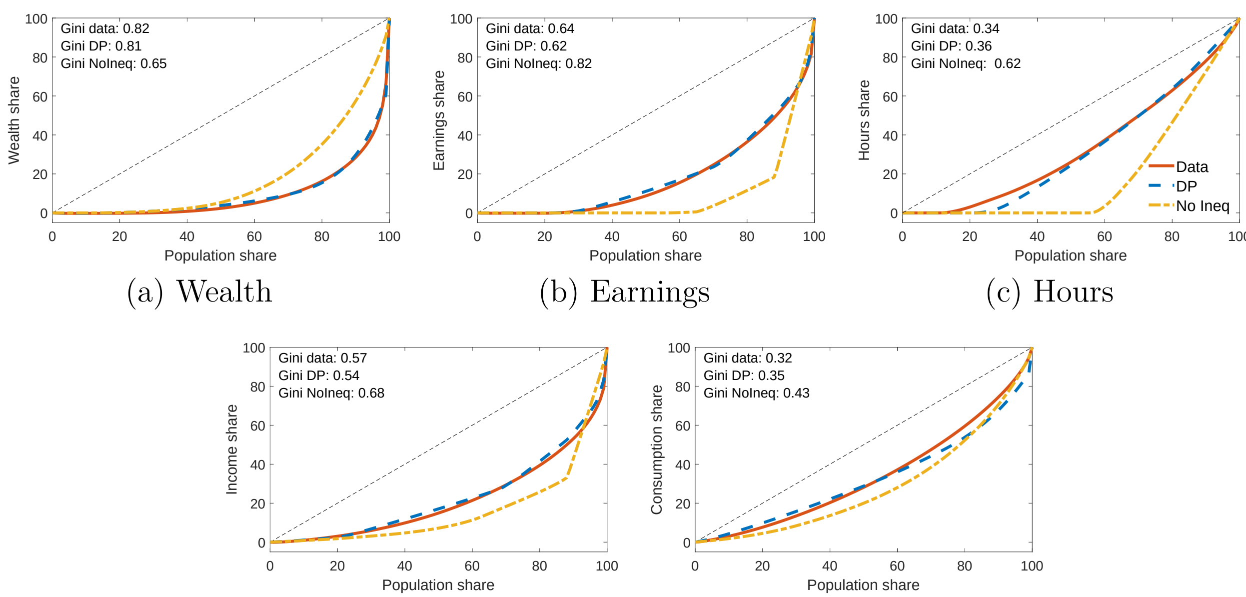
We use 8 points in the grid of productivities and set the range of standard deviations in the augmented Tauchen method to 2. This leaves the following parameters to discipline: $\{\rho_e, \mu_1, \mu_2, p, \sigma_1, \sigma_2\}$. We normalize the mean productivity to one by adjusting μ_2 accordingly. This leaves 5 parameters to discipline. Their values and the model moments used to discipline them are presented in Tables 18 and 19. This calibration strategy has more limitations than our benchmark strategy even considering only the subset of targets we focus on here. The model's fit to the macroeconomic aggregates and statistical properties of earnings is worse. In particular, under the no inequality calibration strategy the share of workers and the Moore kurtosis are significantly off relative to their data counterpart.

Figure 20 presents the fit of the No-Inequality-Targets (NIT) calibration to the cross-sectional distributions, and compares them to our benchmark calibration. Under the no inequality calibration the model overshoots substantially the inequality of hours worked and, as a result, the distribution of earnings. The Gini for hours

Table 19: No Inequality Targets: Target Statistics and Model Counterparts

(1) Macroeconomic Aggregates		
	Target	**Model**
Average hours worked	0.32	0.34
Capital to output	2.50	2.55
Capital income share	0.38	0.38
Investment to output	0.26	0.27
Transfer to output (%)	11.4	11.4
Debt-to-output (%)	61.5	61.5
Share of workers (%)	79.0	55.5
Fraction of hhs with negative net-worth (%)	9.7	9.9
(2) Statistical Properties of Earnings		
	Target	**Model**
Variance of 1-year change	2.33	2.35
Kelly Skewness of 1-year change	−0.12	−0.12
Moore Kurtosis of 1-year change	2.65	1.82

worked is twice as large as the one in the data. At the same time, this calibration strategy produces much lower wealth inequality relative to the data and the benchmark.



(a) Wealth (b) Earnings (c) Hours

(d) Income (e) Consumption

Figure 20: No-Inequality-Target Calibration: Fit to Inequality Data

N.2.1 Ramsey Policy in No-Inequality-Targets Economy

An extensive list of figures about the Ramsey policy computed for the NIT calibration is presented in Appendix O.14 together with those for our benchmark calibration computed with 8 parameters for comparability. The Ramsey planner in the NIT economy keeps the capital income tax at the upper bound for 13 periods and for 14 periods in our calibration. Long-run capital income taxes are the 27 percent for the NIT and 34 in ours. Given the significant differences between these calibrations, these paths are surprisingly similar.

The labor income tax in the NIT economy jumps on impact to 42 percent and increases to 63 percent in the long run. For comparison, labor income taxes are reduced on impact and increase towards 40 percent in the long run of our benchmark calibration. The substantially higher optimal labor income taxes in the NIT economy are to a large extent a result of the stronger wealth effects on labor supply. Since NIT economy underperforms in terms of wealth inequality the most productive agents are poorer in terms of wealth relative to our calibration. Following an increase in labor income taxes, the negative wealth effects on labor supply are stronger in the NIT economy and lead to an effectively more inelastic supply margin at the top of productivity distribution, actually increasing the labor supply of the most productive agents. As a result, it is optimal for the planner to tax labor at much higher rates in the NIT economy than in our calibration.

Similarly strong wealth effects also appear in the AHHW economy, which also underperforms in terms of wealth inequality and features higher optimal labor income taxes than what we find for our calibration, 77 percent versus 40 percent. So, for more details on this mechanism, we refer the reader to Appendix M. In the NIT economy lump-sum transfers more front-loaded, while the comparable 8-variable version of our results they are actually increasing over time. More aggressive front-loading of lump-sum transfers is possible due to large tax revenues driven by high labor income taxes. Also, as a result, the government debt increases more rapidly in the NIT economy.

N.3 Return Risk

Under this calibration strategy we add return risk to the model. As highlighted by Fagereng, Holm, Moll, and Natvik (2019), Fagereng, Guiso, Malacrino, and Pistaferri (2020) and Hubmer, Krusell, and Smith (2020), heterogeneity in returns on assets and their riskiness are important determinants of wealth dispersion. We extend our benchmark calibration to account for that, in a parsimonious way, by adding an i.i.d. shock to the interest rate faced by households. We model it with a three-point support that is independent of the households level of assets.[15] Hence, the interest rate faced by households becomes:

$$\tilde{r} = r + \eta,$$

where r is given by the marginal product of capital, and $\eta \in \{\eta_1, \eta_2, \eta_3\}$ with probabilities $\{\pi_i\}_{i=1}^{3}$. We discipline the parameters of the shock as follows. First, we set the mean of η to zero and assume the states have equal probability—this leaves one degree of freedom. Second, based on numbers provided by Hubmer et al. (2020) (in

[15] This allows us drop the additional market clearing condition associated with return risk, which requires equalizing aggregate capital income with the average over individual capital income—see equation (10) in Hubmer et al. (2020).

their Table 6), we calculate that the population-average standard deviation of return on assets is 0.098. Using this number we arrive at the following: $\eta_1 = -0.121, \eta_2 = 0, \eta_3 = 0.121$, with $\pi_1 = \pi_2 = \pi_3 = 1/3$. In Table 20, we illustrate the impact of this shock on the moments we target in our benchmark calibration. We also consider a lower level (0.5×0.098) and a higher level (2×0.098) of standard deviation for sensitivity analysis.

The bottom line from this analysis is that the return risk impacts mostly the distribution of wealth at the top and even that effect is quantitative limited. Under our preferred specification, i.e. medium return risk with standard deviation of 0.098, the shares of top 5 percent in wealth distribution rises by 1.7 percentage points, bringing the model closer to the data in this regard. But even if we impose a high return risk by doubling the standard deviation—which would be equivalent to imposing the standard deviation of the top 0.01 percent of wealth distribution on the entire population—the effects are still limited, with top 5 percent share growing by 4.6 percentage points. Figure 21 presents the model's fit to inequality data. The Lorenz curves for wealth, earnings, hours, and consumption are almost indistinguishable in the economy with return risk to our benchmark calibration. However, as would be expected, the income inequality increases, which shifts out the Lorenz curve for income under the Return-Risk calibration and increases the income Gini coefficient to 0.62 compared to 0.54 in our benchmark calibration and 0.57 in the data.

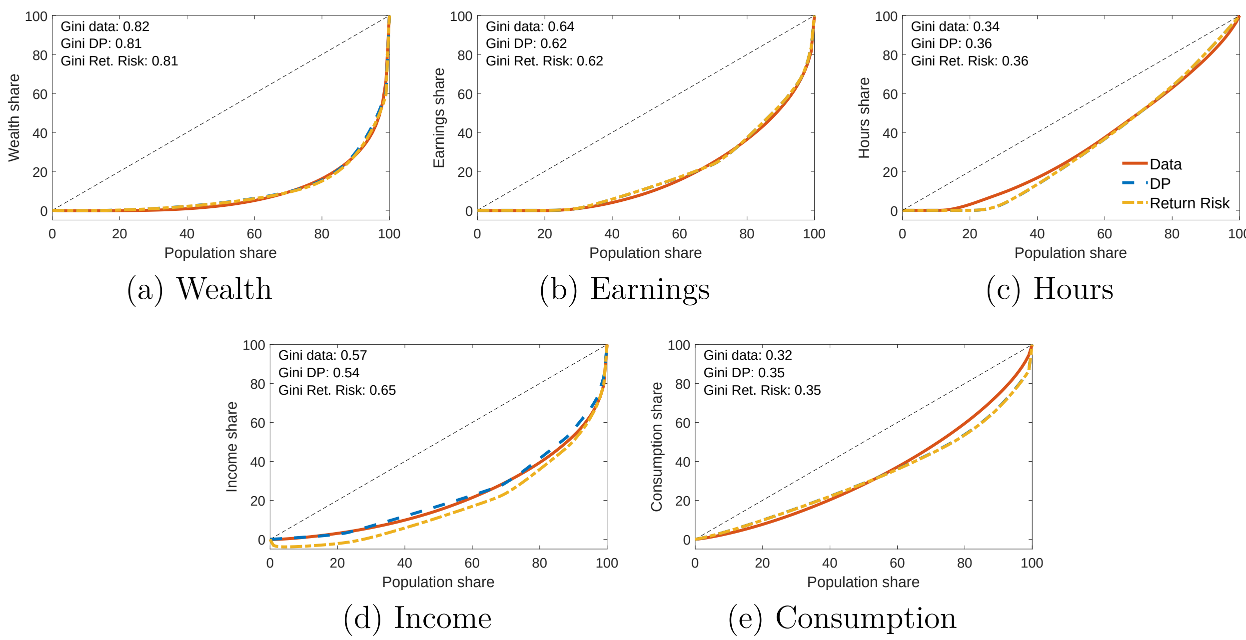


Figure 21: Return Risk Calibration: Fit to Inequality Data

N.3.1 Ramsey Policy in Return Risk Economy

In Appendix O.15 we present an extensive list of figures for the Ramsey policy computed for the return-risk calibration and, for comparison, also our benchmark-calibration results—both approximations are computed with 8 parameters for comparability. Given that the return risk does not change the fit of the model to the data significantly, it is not surprising that the optimal policy looks qualitatively and quantitatively very similar to the benchmark one. There are few minor differences. Return risk introduces another source of risk into the

Table 20: Return Risk: Target Statistics and Model Counterparts

(1) Macroeconomic Aggregates

	Target	Benchmark	Return Risk		
			Low	Medium	High
Average hours worked	0.32	0.33	0.33	0.33	0.33
Capital to output	2.50	2.49	2.49	2.48	2.46
Capital income share	0.38	0.38	0.38	0.38	0.38
Investment to output	0.26	0.26	0.26	0.26	0.26
Transfer to output (%)	11.4	11.4	11.4	11.4	11.4
Debt-to-output (%)	61.5	61.5	61.5	61.5	61.5
Hhs with negative net-worth (%)	9.7	7.9	7.9	8.0	8.3

(2) Cross-sectional Moments

	Bottom (%)	Quintiles					Top (%)	Gini
	0-5	1st	2nd	3rd	4th	5th	95-100	
				Wealth				
US Data	-0.2	-0.2	1.0	4.2	11.2	83.8	60.0	0.82
Benchmark	-0.1	0.1	2.0	4.0	9.3	84.5	56.4	0.81
Low Risk	-0.1	0.1	2.0	4.0	9.2	84.7	56.9	0.81
Medium Risk	-0.1	0.1	1.9	3.9	9.0	85.0	58.1	0.81
High Risk	-0.1	0.1	1.8	3.9	8.5	85.7	61.0	0.82
				Earnings				
US Data	-0.2	-0.2	4.1	11.6	20.9	63.6	35.6	0.64
Benchmark	0.0	0.0	5.7	11.3	20.2	62.8	34.8	0.62
Low Risk	0.0	0.0	5.7	11.3	20.2	62.8	34.8	0.62
Medium Risk	0.0	0.0	5.7	11.4	20.2	62.7	34.7	0.62
High Risk	0.0	0.0	5.8	11.4	20.1	62.7	34.7	0.62
				Hours				
US Data	0.0	3.0	13.7	20.7	25.4	37.2	12.9	0.34
Benchmark	0.0	0.0	13.2	23.4	27.1	36.3	9.9	0.36
Low Risk	0.0	0.0	13.3	23.4	27.1	36.2	9.9	0.36
Medium Risk	0.0	0.0	13.3	23.3	27.1	36.2	9.9	0.36
High Risk	0.0	0.0	13.4	23.3	27.2	36.1	9.9	0.36

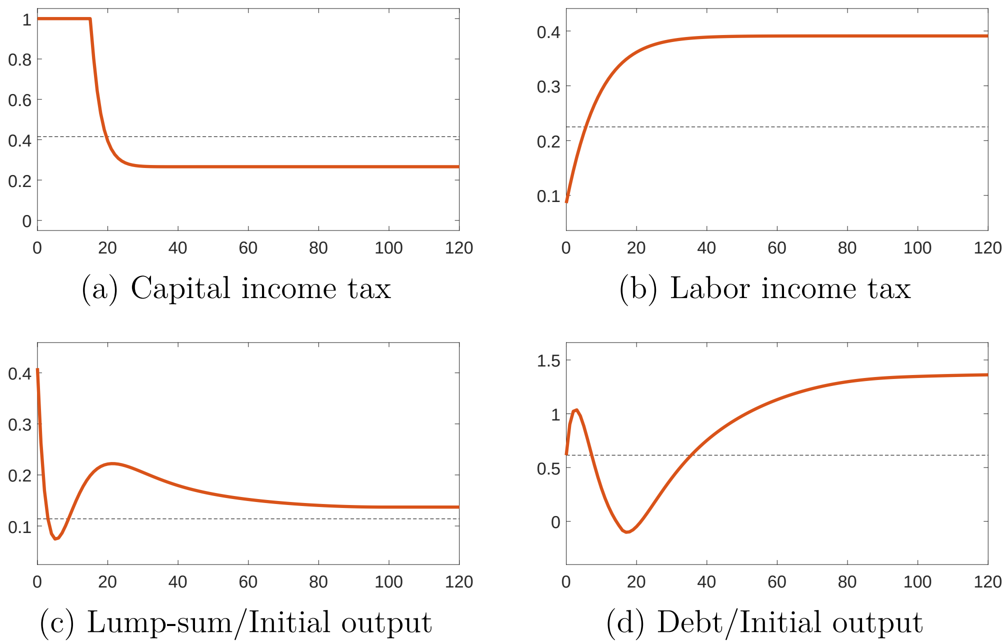
Note: Low Return Risk: $0.5 \times \sigma_r$, Medium Return Risk: $\sigma_r = 0.098$, High Return Risk: $2.0 \times \sigma_r$.

model making capital income taxes more effective at providing insurance. This, together with the fact that there is more inequality at the top of wealth distribution, explains the higher long-run optimal capital income tax: 42.2 percent versus 34.3 percent in the benchmark economy. The labor income tax follows a similar time pattern to the benchmark and in the long run the return-risk economy settles at rate 38.8 percent compared to 40.2 percent in the benchmark economy.

With the 8-parameter approximations we cannot precisely pinpoint the optimal debt-to-output path, but the higher capital income taxes in the long run translates to a lower debt-to-output ratio in the return-risk economy. As for the macro aggregates, only the evolution of capital differs noticeably between the return-risk and benchmark calibrations, the capital stock in the return-risk case is 3 percent lower in the long-run. This is a direct consequence higher long-run capital income taxes. Other than that the evolution of macroeconomic aggregates is qualitatively and quantitatively very similar in the return risk economy relative to the benchmark one. The optimal policy under the return-risk calibration yields 3.54 percent of welfare gains, which compares to 3.40 percent under the benchmark calibration with 8-parameter approximation.

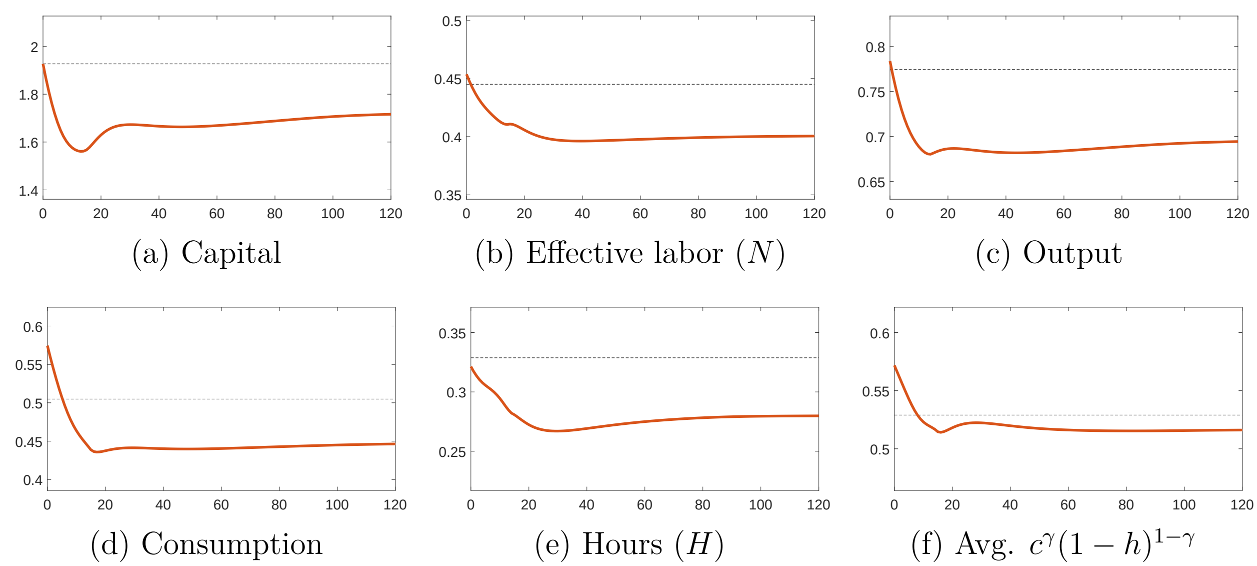
O Figures

O.1 Benchmark Results



(a) Capital income tax
(b) Labor income tax
(c) Lump-sum/Initial output
(d) Debt/Initial output

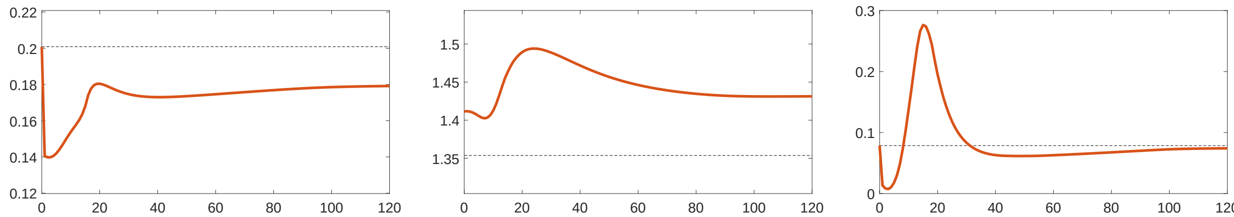
Figure 22: Optimal Fiscal Policy: Benchmark



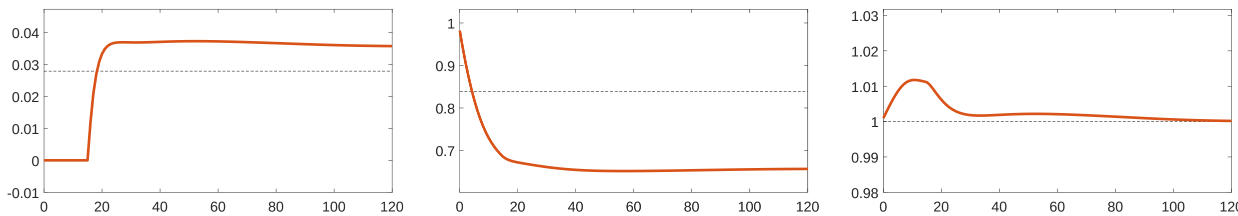
(a) Capital
(b) Effective labor (N)
(c) Output
(d) Consumption
(e) Hours (H)
(f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 23: Aggregates: Benchmark (1)

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition (benchmark).

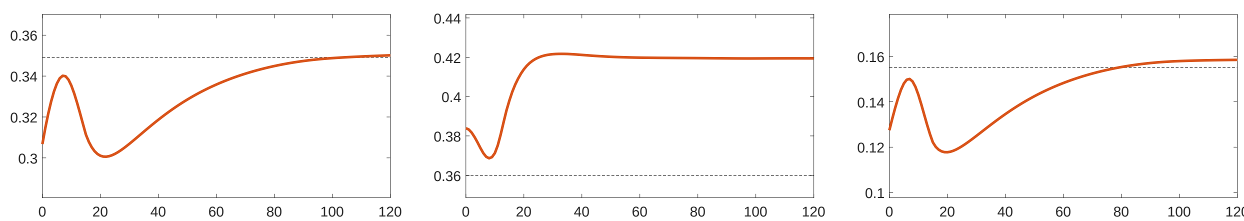


(g) Investment (h) Labor productivity (i) Prop. w/ negative assets

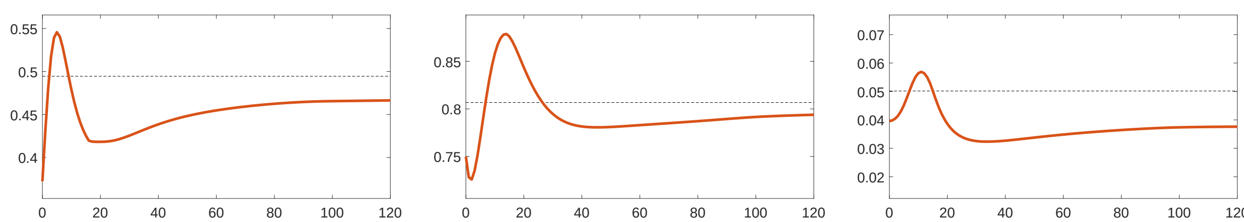


(j) After-tax int. rates (k) After-Tax Wages (l) Modified golden rule

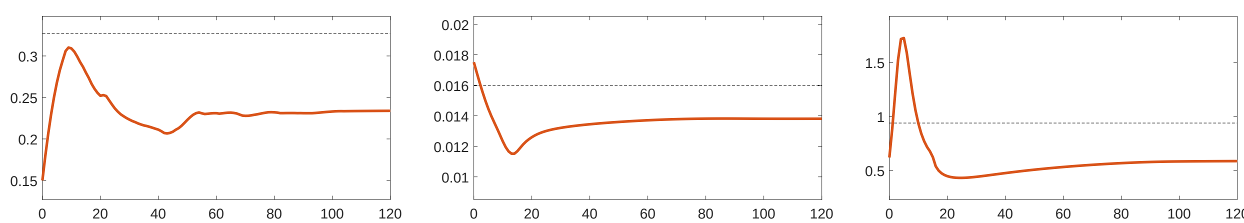
Figure 23: Aggregates: Benchmark (2)



(m) Consumption Gini (n) Hours Gini (o) Gini of $c^{\gamma}(1-h)^{1-\gamma}$



(p) After-tax income Gini (q) Wealth Gini (r) Var. cons. growth

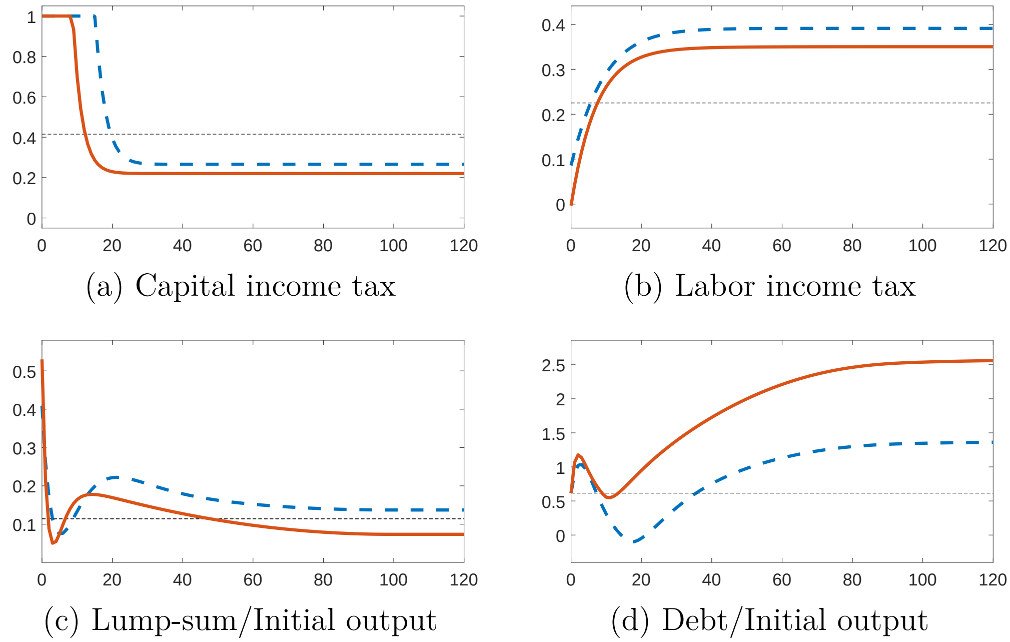


(s) Var. hours growth (t) Var. $c^{\gamma}(1-h)^{1-\gamma}$ growth (u) Var. a-t income growth

Figure 24: Inequality and Risk: Benchmark

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition (benchmark).

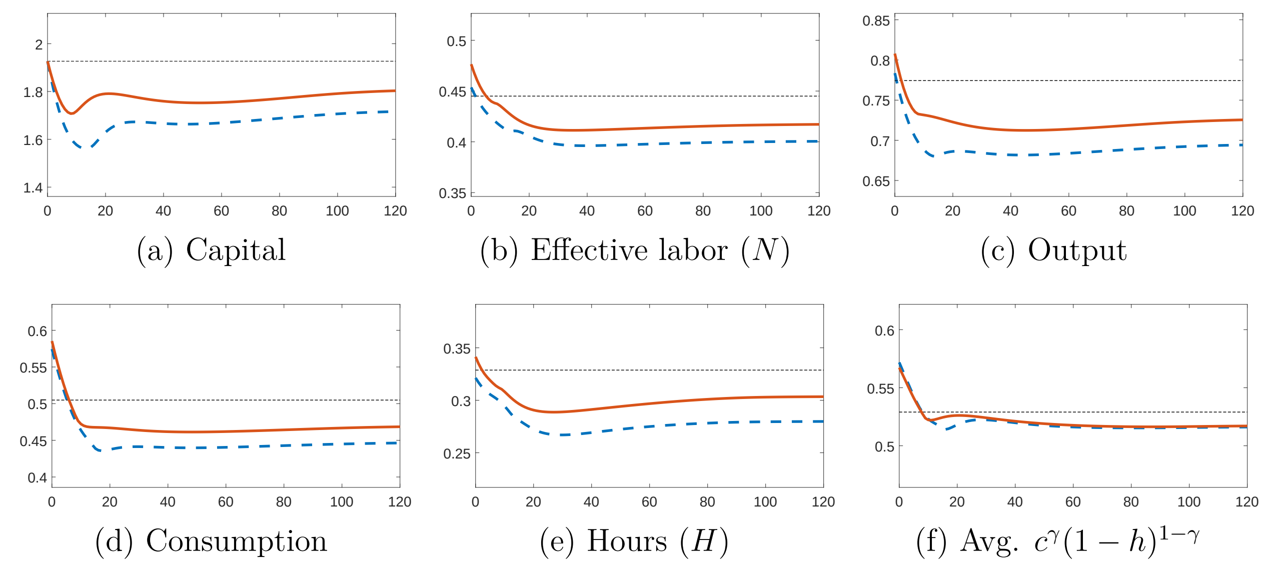
O.2 Maximizing Efficiency



(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 25: Optimal Fiscal Policy: Maximizing Efficiency

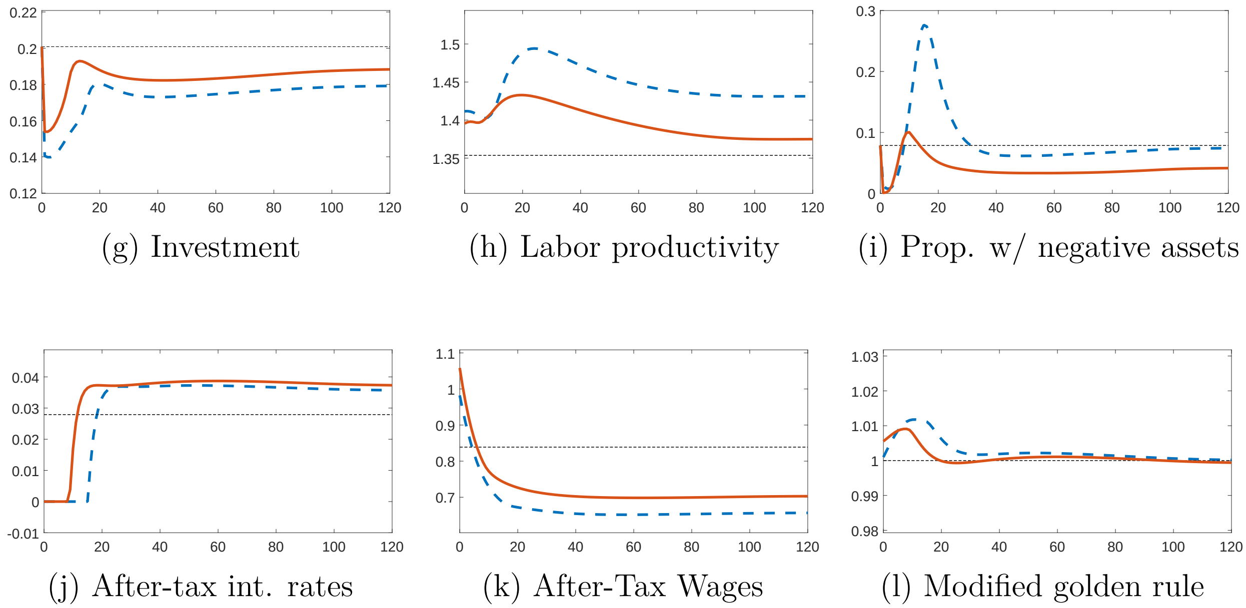


(a) Capital (b) Effective labor (N) (c) Output

(d) Consumption (e) Hours (H) (f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 26: Aggregates: Maximizing Efficiency (1)

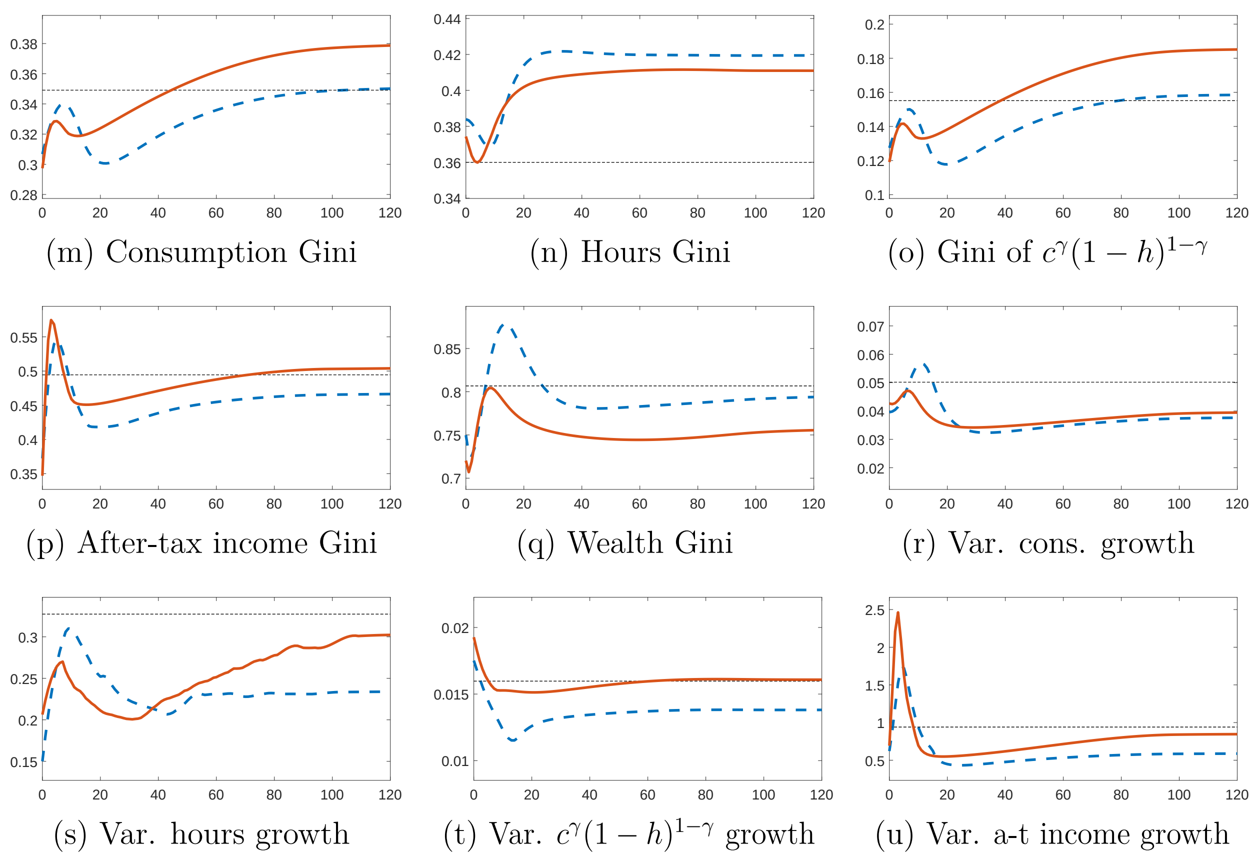
Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition maximizing efficiency; Blue dashed curve: optimal transition (benchmark).



(g) Investment (h) Labor productivity (i) Prop. w/ negative assets

(j) After-tax int. rates (k) After-Tax Wages (l) Modified golden rule

Figure 26: Aggregates: Maximizing Efficiency (2)



(m) Consumption Gini (n) Hours Gini (o) Gini of $c^{\gamma}(1-h)^{1-\gamma}$

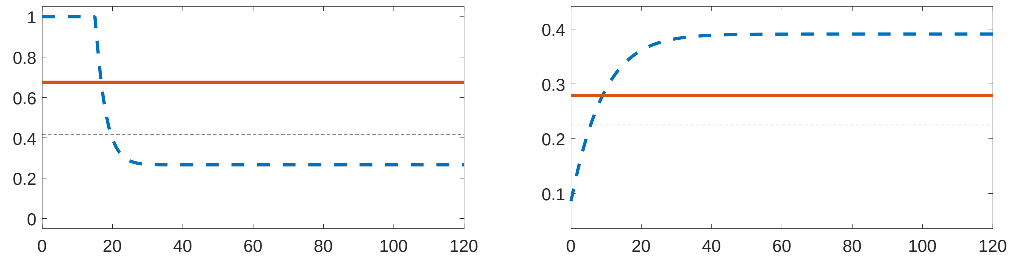
(p) After-tax income Gini (q) Wealth Gini (r) Var. cons. growth

(s) Var. hours growth (t) Var. $c^{\gamma}(1-h)^{1-\gamma}$ growth (u) Var. a-t income growth

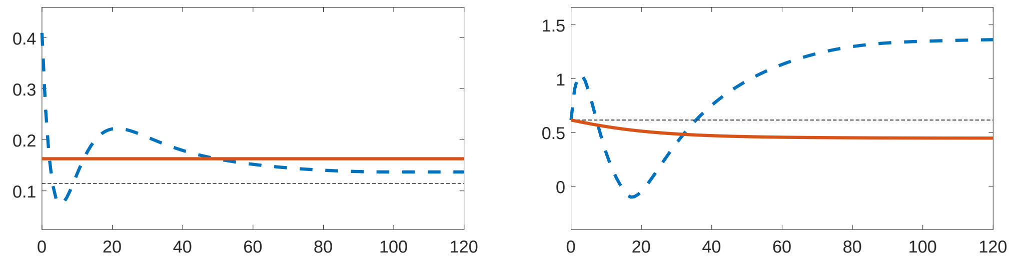
Figure 27: Inequality and Risk: Maximizing Efficiency

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition maximizing efficiency; Blue dashed curve: optimal transition (benchmark).

O.3 Constant Policy

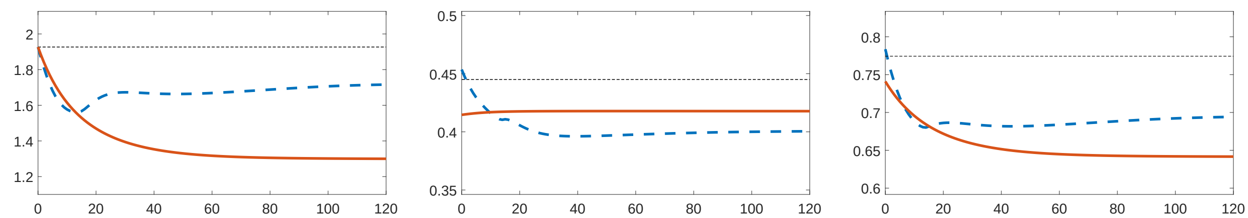


(a) Capital income tax (b) Labor income tax

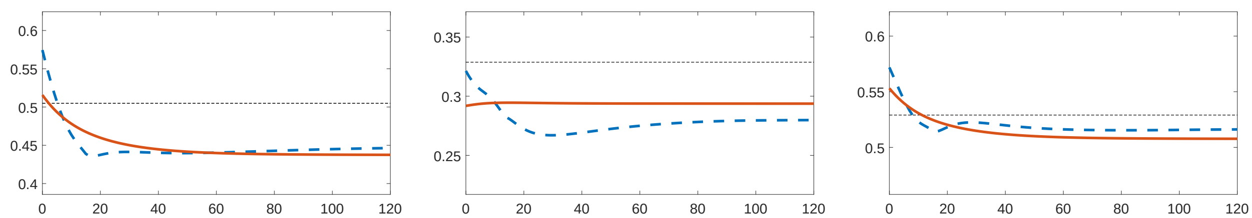


(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 28: Optimal Fiscal Policy: Constant Taxes



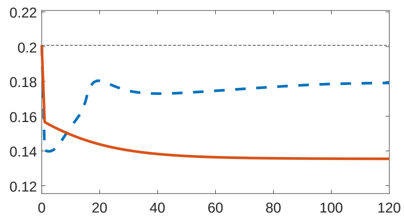
(a) Capital (b) Effective labor (N) (c) Output



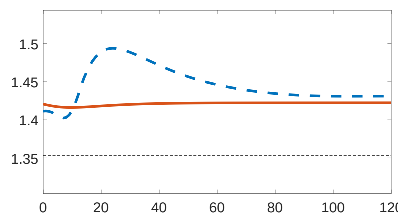
(d) Consumption (e) Hours (H) (f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 29: Aggregates: Constant Taxes (1)

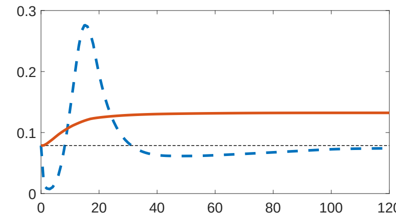
Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition with constant taxes; Blue dashed curve: optimal transition (benchmark).



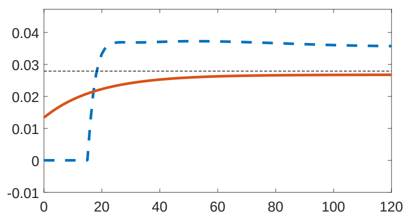
(g) Investment



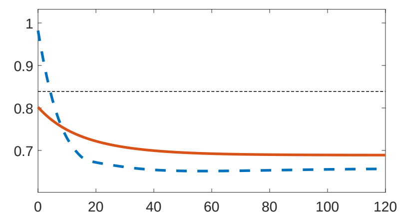
(h) Labor productivity



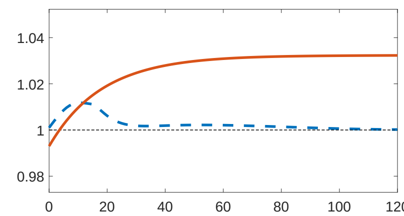
(i) Prop. w/ negative assets



(j) After-tax int. rates

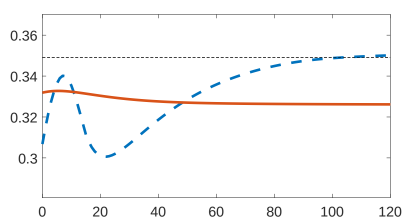


(k) After-Tax Wages

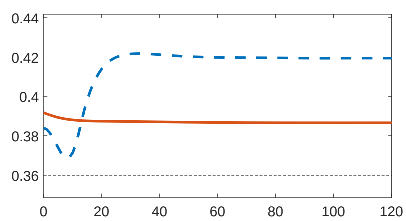


(l) Modified golden rule

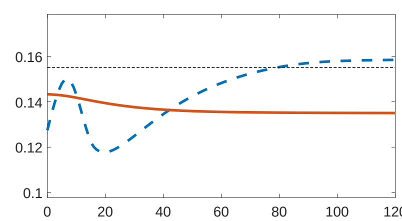
Figure 29: Aggregates: Constant Taxes (2)



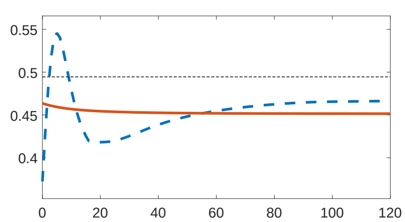
(m) Consumption Gini



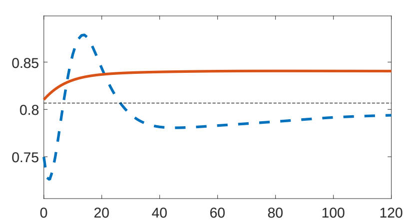
(n) Hours Gini



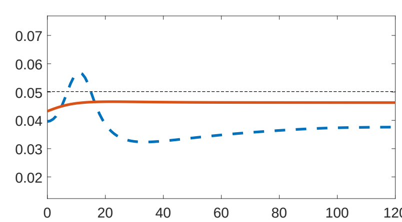
(o) Gini of $c^{\gamma}(1-h)^{1-\gamma}$



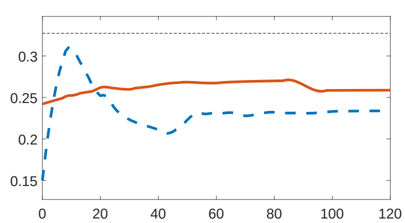
(p) After-tax income Gini



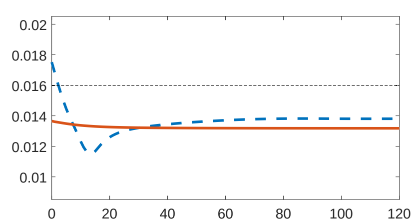
(q) Wealth Gini



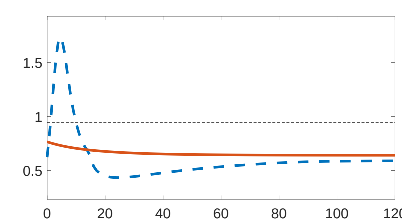
(r) Var. cons. growth



(s) Var. hours growth



(t) Var. $c^{\gamma}(1-h)^{1-\gamma}$ growth

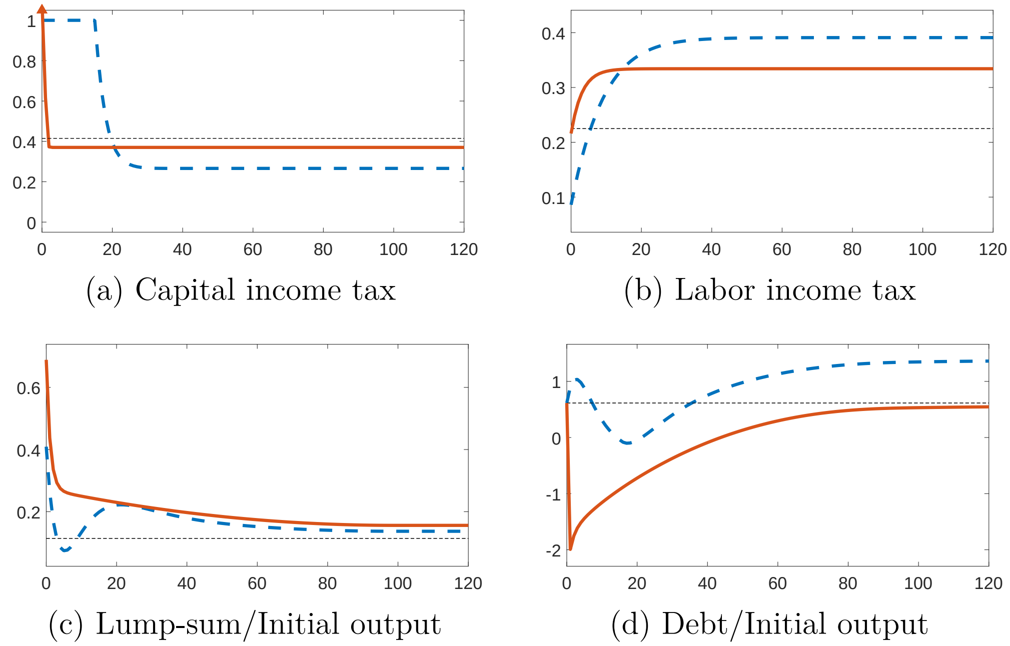


(u) Var. a-t income growth

Figure 30: Inequality and Risk: Constant Taxes

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition with constant taxes; Blue dashed curve: optimal transition (benchmark).

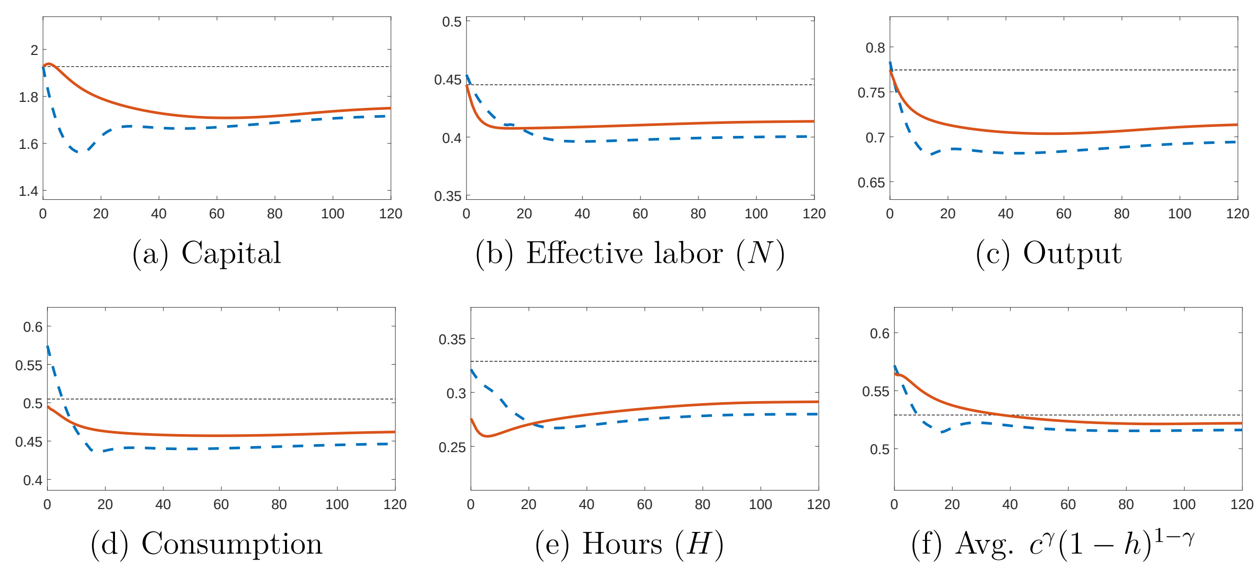
O.4 Levy on Initial Capital Income



(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 31: Optimal Fiscal Policy: Capital Levy



(a) Capital (b) Effective labor (N) (c) Output

(d) Consumption (e) Hours (H) (f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 32: Aggregates: Capital Levy (1)

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: path that maximizes the utilitarian welfare function allowing for capital income taxes to move in period 0 (though the tax level at $t = 0$ is not plotted since it is equal to $(1 + r_0)/r_0 = 21.96$); Blue dashed curve: optimal transition (benchmark).

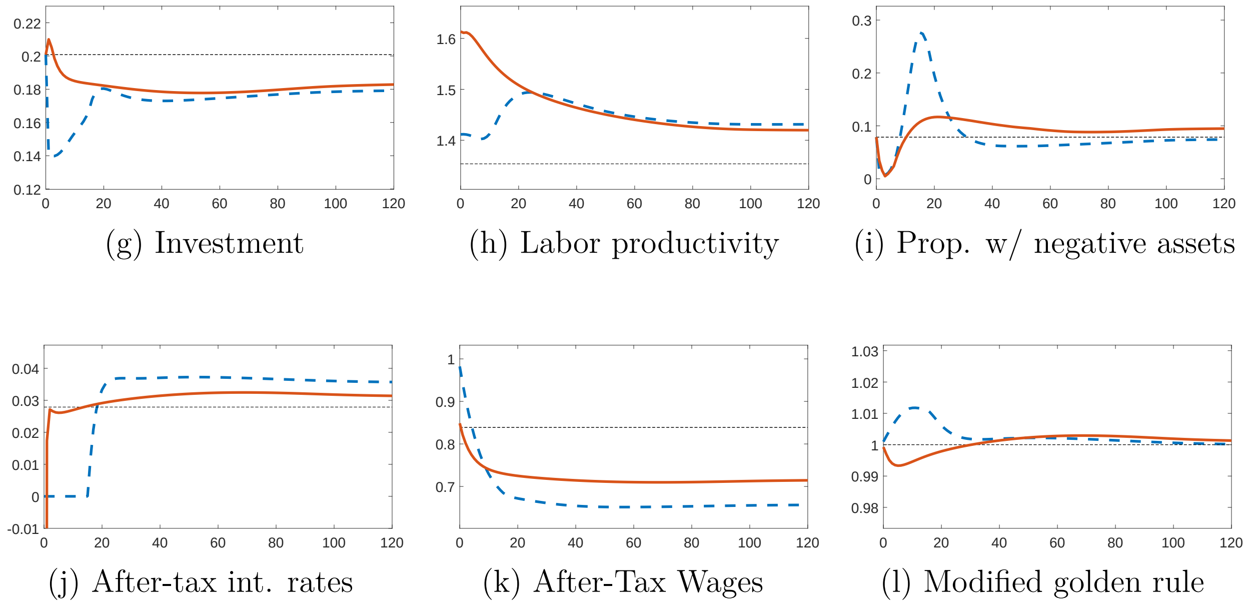


Figure 32: Aggregates: Capital Levy (2)

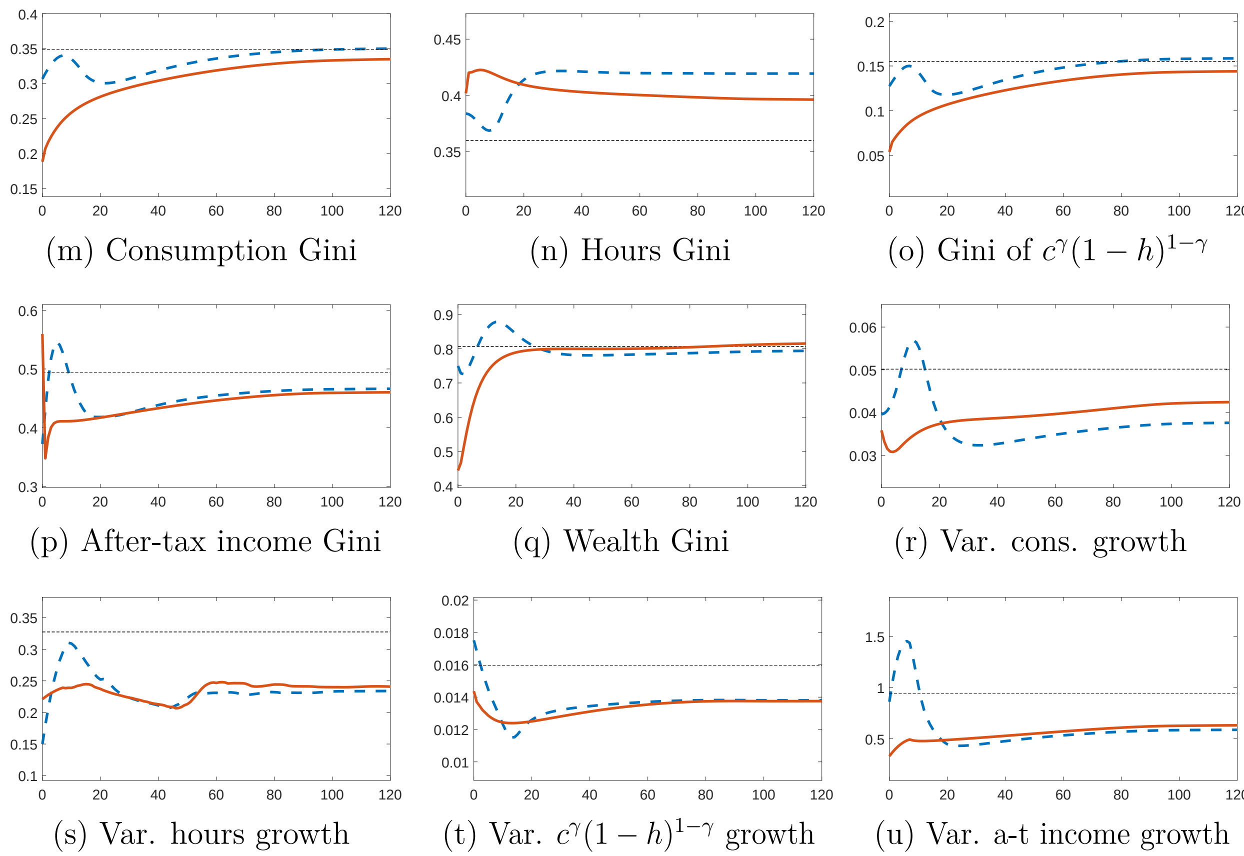
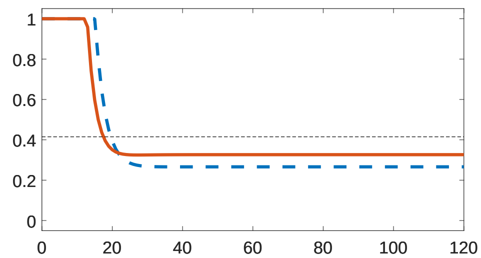


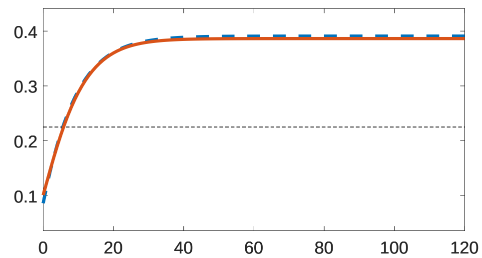
Figure 33: Inequality and Risk: Capital Levy

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: path that maximizes the utilitarian welfare function allowing for capital income taxes to move in period 0 (though the tax level at $t = 0$ is not plotted since it is equal to $(1 + r_0)/r_0 = 21.96$); Blue dashed curve: optimal transition (benchmark).

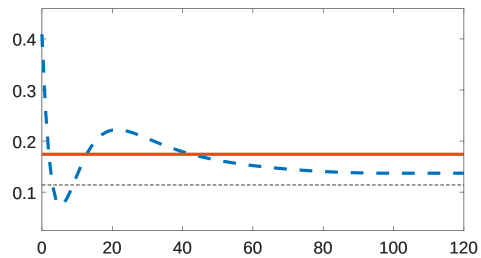
O.5 Constant Lump-Sum Transfers



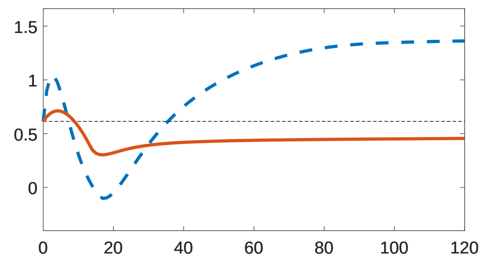
(a) Capital income tax



(b) Labor income tax

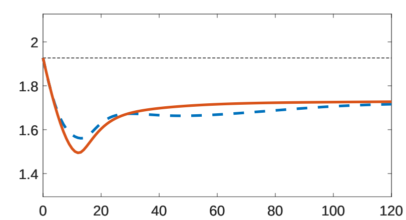


(c) Lump-sum/Initial output

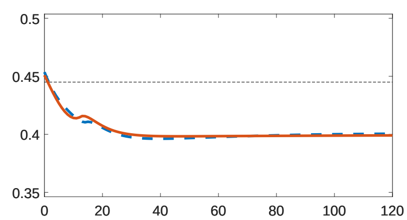


(d) Debt/Initial output

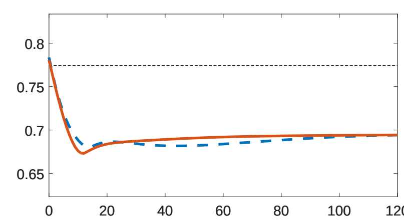
Figure 34: Optimal Fiscal Policy: Constant Lump-Sum Transfers



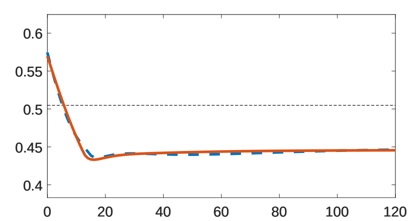
(a) Capital



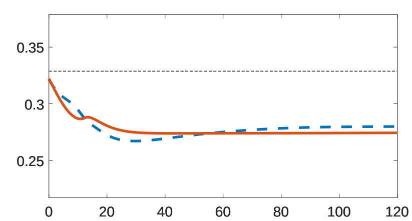
(b) Effective labor (N)



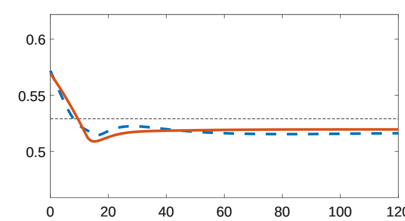
(c) Output



(d) Consumption



(e) Hours (H)



(f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 35: Aggregates: Constant Lump-Sum Transfers (1)

Notes: Thin dashed line: initial stationary equilibrium; Solid line: path that maximizes the utilitarian welfare function with the added restriction that lump-sum transfers are not allowed to vary over time after the initial change; Thick dashed line: benchmark results.

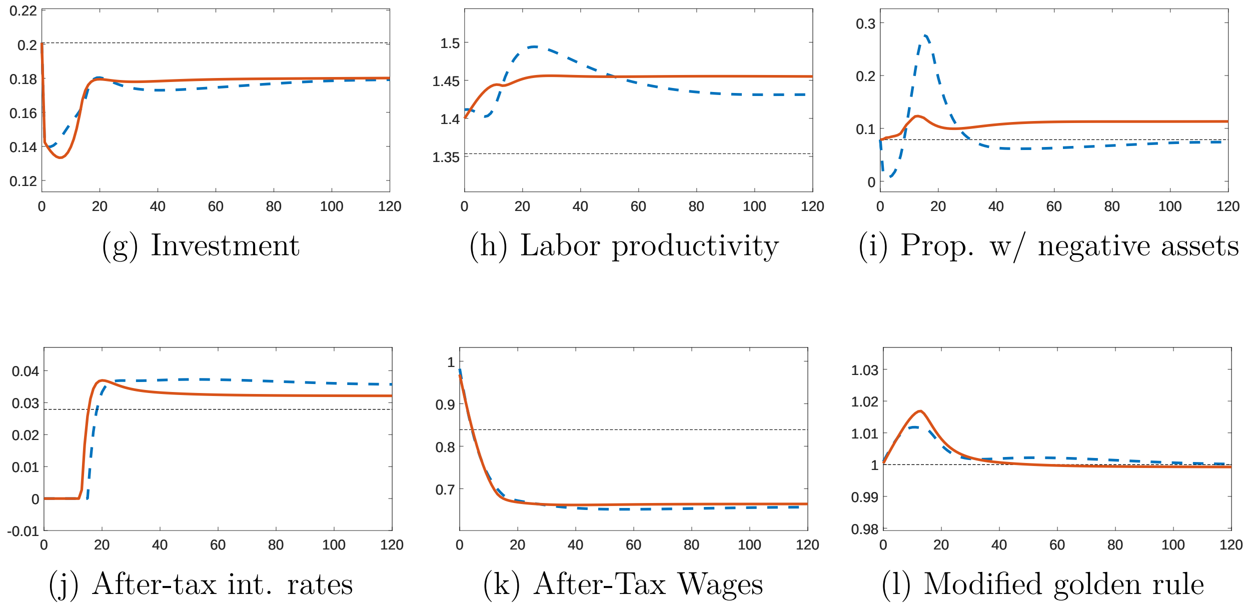


Figure 35: Aggregates: Constant Lump-Sum Transfers (2)

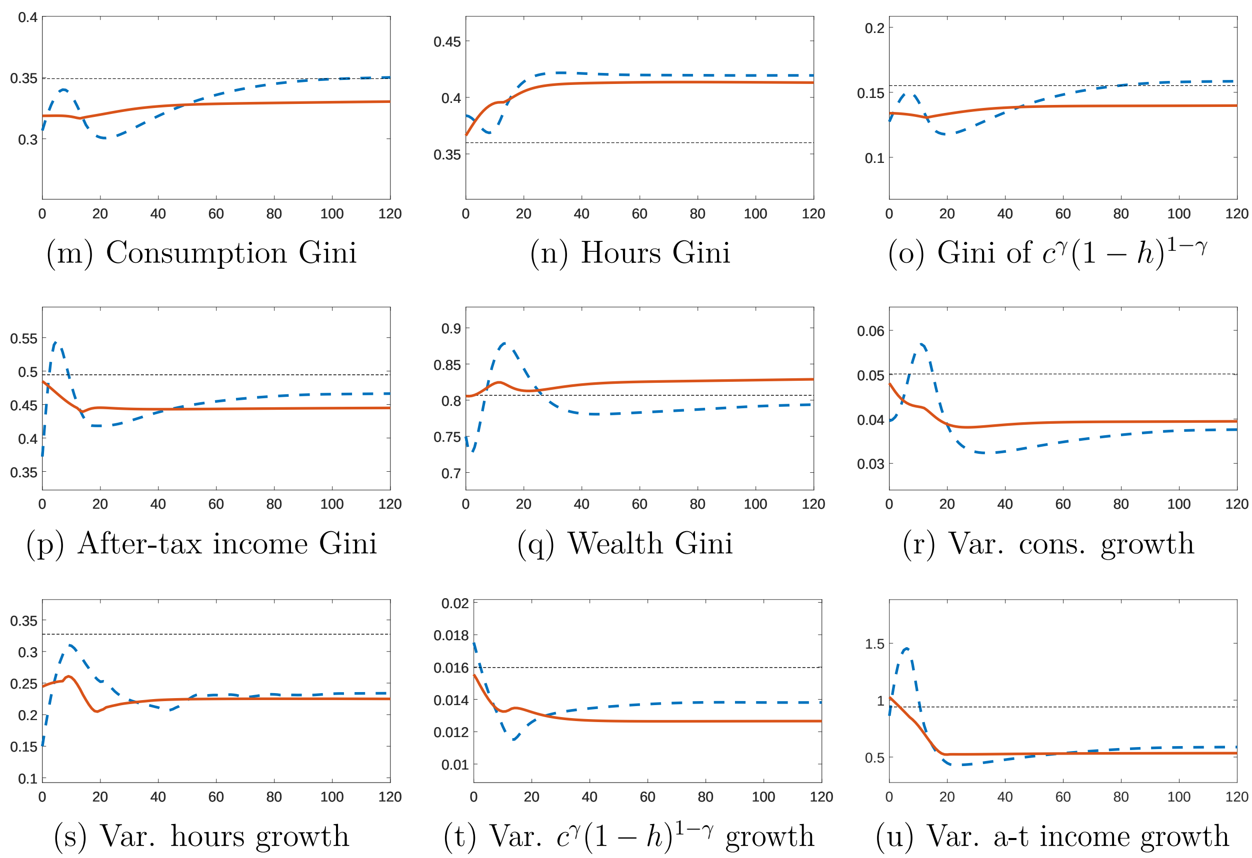
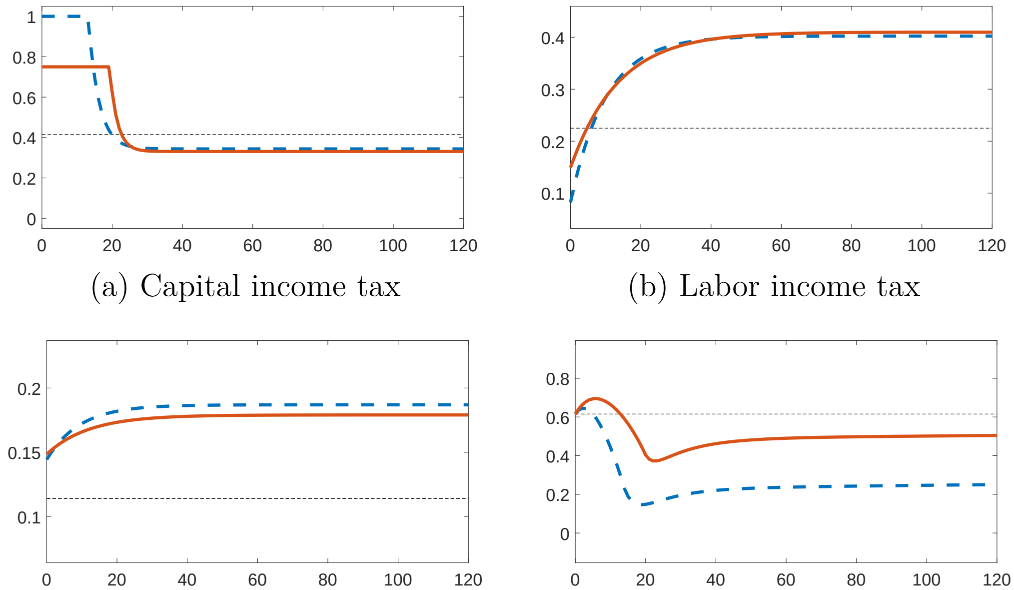


Figure 36: Inequality and Risk: Constant Lump-Sum Transfers

Notes: Thin dashed line: initial stationary equilibrium; Solid line: path that maximizes the utilitarian welfare function with the added restriction that lump-sum transfers are not allowed to vary over time after the initial change; Thick dashed line: benchmark results.

O.6 Different Upper Bounds on Capital Income Taxes

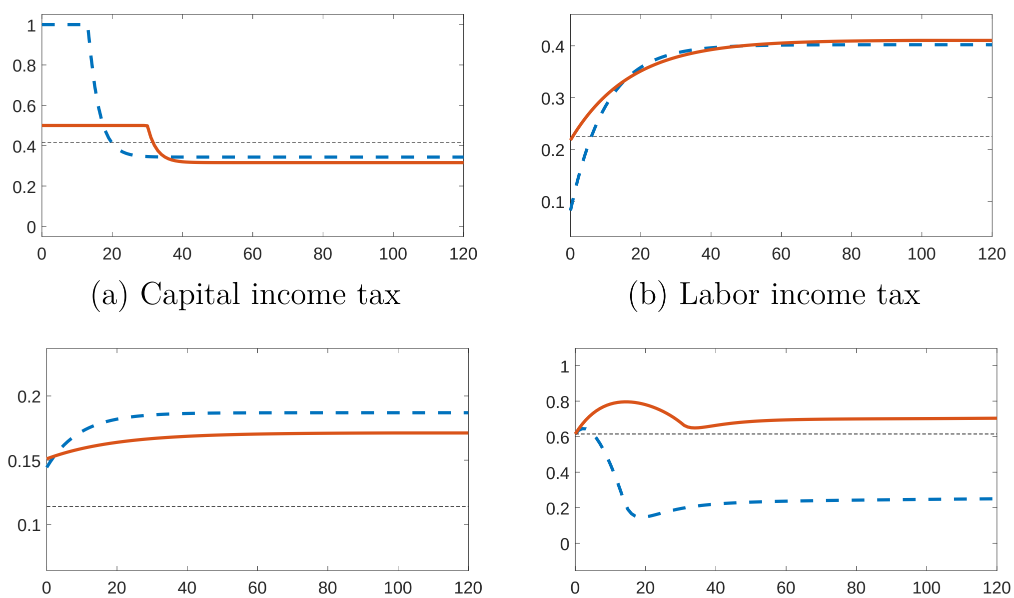


(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 37: Upper bound on capital income taxes at 0.75

Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: benchmark optimal policy with 8 parameters; Red solid curve: optimal policy with 8 parameters and upper bound on capital taxes at 0.75.



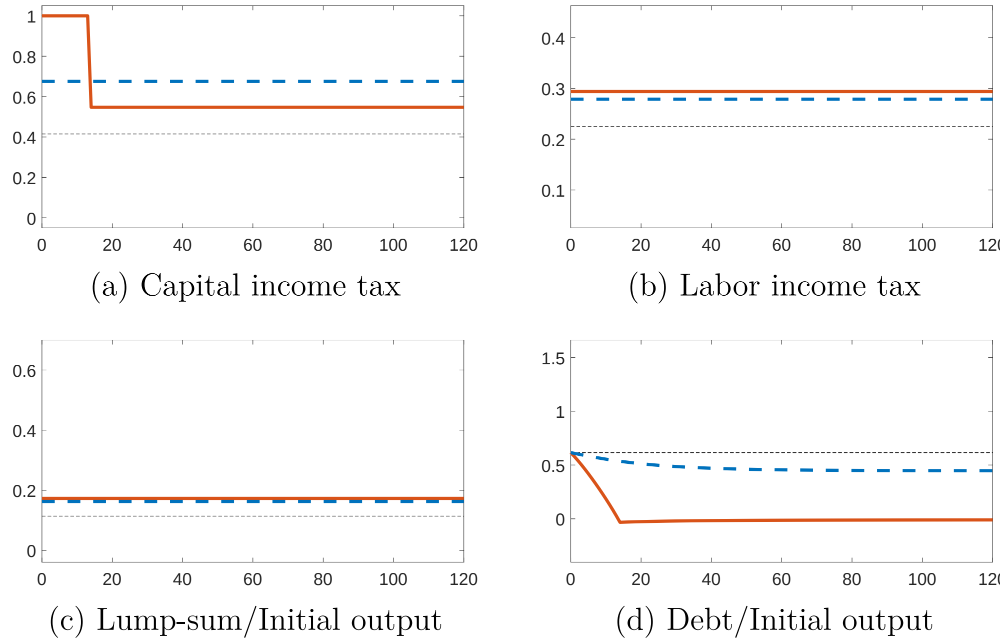
(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 38: Upper bound on capital income taxes at 0.50

Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: benchmark optimal policy with 8 parameters; Red solid curve: optimal policy with 8 parameters and upper bound on capital taxes at 0.50.

O.7 Adding Flexibility (see Appendix G.3)

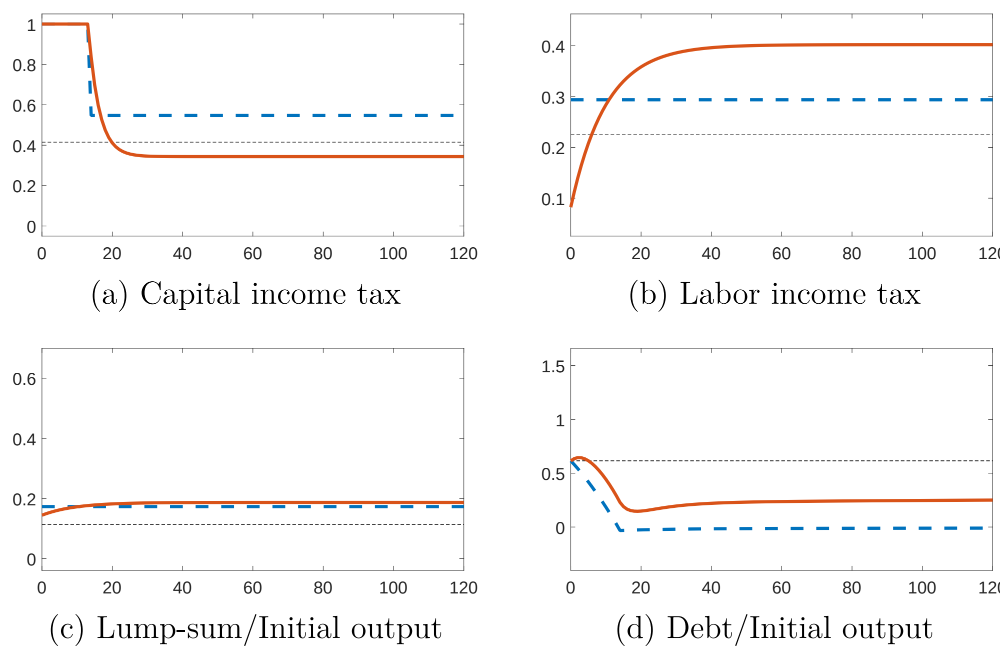


(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 39: Number of Parameters: 2 to 3

Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: optimal policy with 2 parameters (τ^k, τ^h); Red solid curve: optimal policy with 3 parameters (t^*, τ_F^k, τ^h).

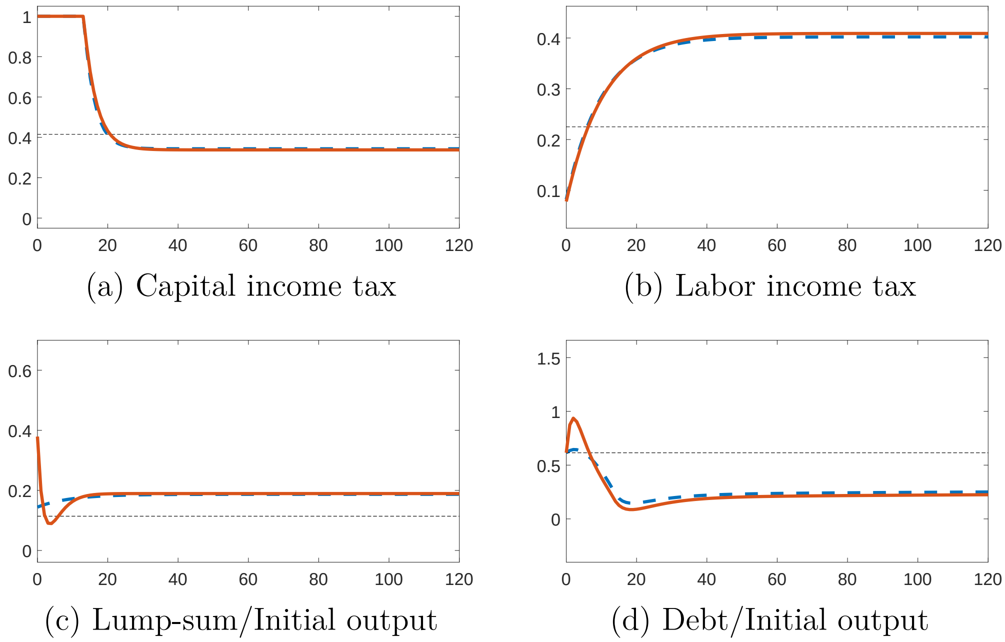


(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 40: Number of Parameters: 3 to 8

Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: optimal policy with 3 parameters; Red solid curve: optimal policy with 8 parameters $(\alpha_0^k, \beta_0^k, \lambda^k, \alpha_0^h, \beta_0^h, \lambda^h, \beta_0^T, \lambda^T)$.

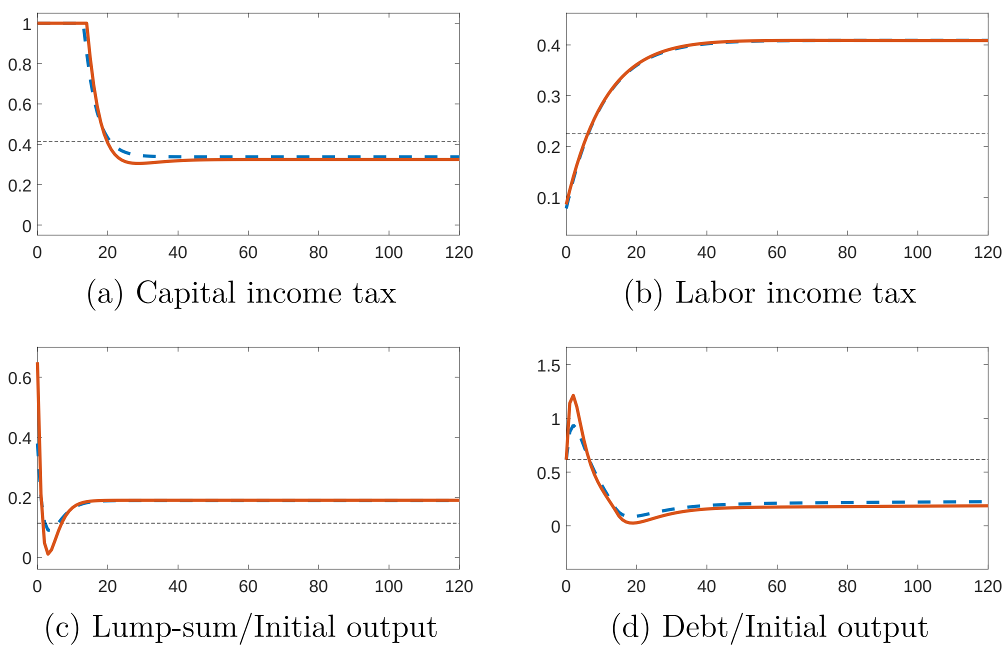


(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 41: Number of Parameters: 8 to 11

Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: optimal policy with 8 parameters; Red solid curve: optimal policy with 11 parameters $\left(\alpha_0^k, \alpha_1^k, \beta_0^k, \lambda^k, \alpha_0^h, \alpha_1^h, \beta_0^h, \lambda^h, \alpha_1^T, \beta_0^T, \lambda^T\right)$.

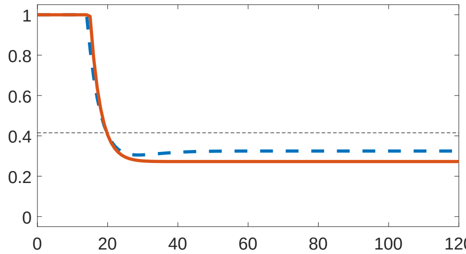


(a) Capital income tax (b) Labor income tax

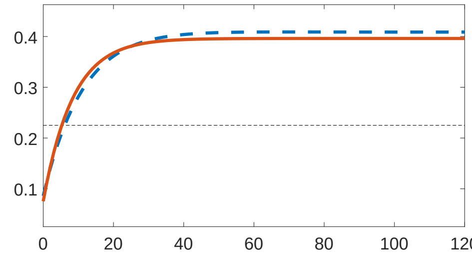
(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 42: Number of Parameters: 11 to 14

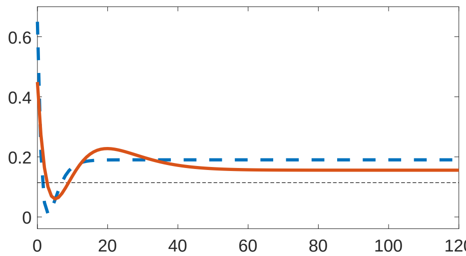
Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: optimal policy with 11 parameters; Red solid curve: optimal policy with 14 parameters $\left(\alpha_0^k, \alpha_1^k, \alpha_2^k, \beta_0^k, \lambda^k, \alpha_0^h, \alpha_1^h, \alpha_2^h, \beta_0^h, \lambda^h, \alpha_1^T, \alpha_2^T, \beta_0^T, \lambda^T\right)$.



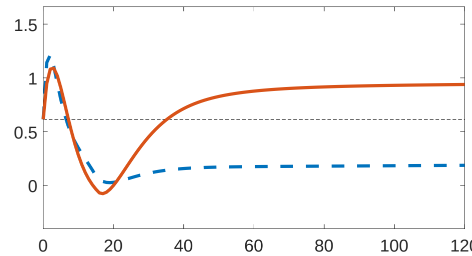
(a) Capital income tax



(b) Labor income tax



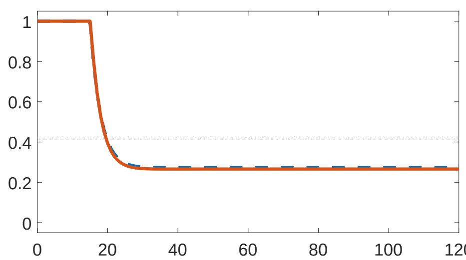
(c) Lump-sum/Initial output



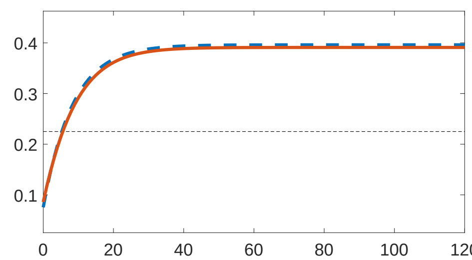
(d) Debt/Initial output

Figure 43: Number of Parameters: 14 to 16

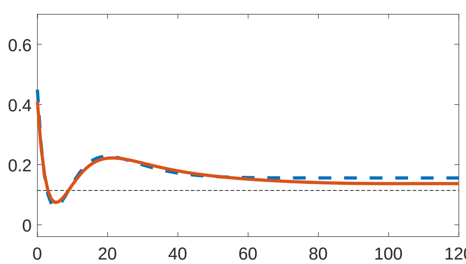
Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: optimal policy with 14 parameters; Red solid curve: optimal policy with 16 parameters $\left(\alpha_0^k, \alpha_1^k, \alpha_2^k, \beta_0^k, \lambda^k, \alpha_0^h, \alpha_1^h, \alpha_2^h, \beta_0^h, \lambda^h, \alpha_1^T, \alpha_2^T, \alpha_3^T, \alpha_4^T, \beta_0^T, \lambda^T\right)$.



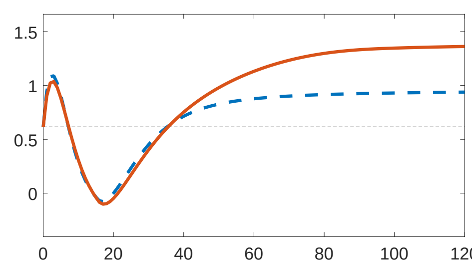
(a) Capital income tax



(b) Labor income tax



(c) Lump-sum/Initial output



(d) Debt/Initial output

Figure 44: Number of Parameters: 16 to 17

Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: optimal policy with 16 parameters; Red solid curve: optimal policy with 17 parameters $\left(\alpha_0^k, \alpha_1^k, \alpha_2^k, \beta_0^k, \lambda^k, \alpha_0^h, \alpha_1^h, \alpha_2^h, \beta_0^h, \lambda^h, \alpha_1^T, \alpha_2^T, \alpha_3^T, \alpha_4^T, \beta_0^T, \beta_1^T, \lambda^T\right)$, with β_2^T chosen such that the derivative of T_t at $t = 100$ is equal to zero.

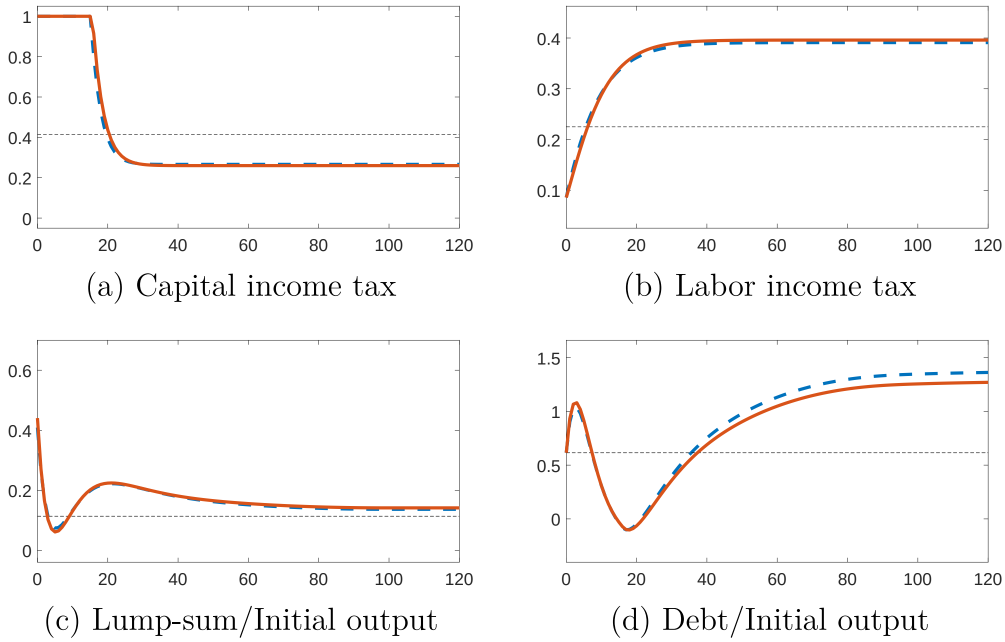
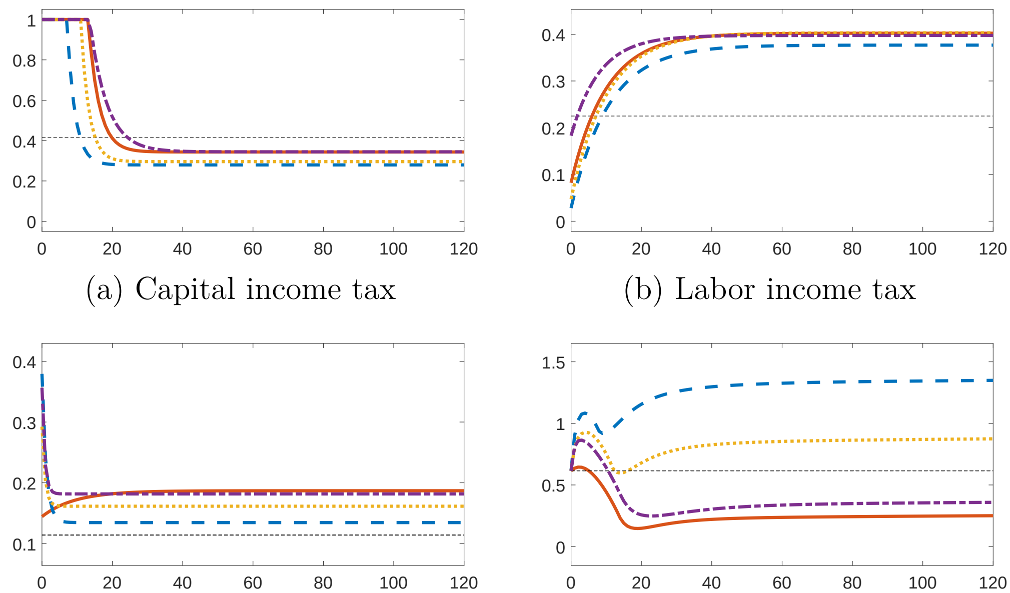


Figure 45: Number of Parameters: 17 to 20

Notes: Black dashed line: initial stationary equilibrium; Blue dashed curve: optimal policy with 17 parameters; Red solid curve: optimal policy (local search) with 20 parameters $\left(\alpha_0^k, \alpha_1^k, \alpha_2^k, \alpha_3^k, \beta_0^k, \lambda^k, \alpha_0^h, \alpha_1^h, \alpha_2^h, \alpha_3^h, \beta_0^h, \lambda^h, \alpha_1^T, \alpha_2^T, \alpha_3^T, \alpha_4^T, \alpha_5^T, \beta_0^T, \beta_1^T, \lambda^T\right)$, with β_2^T chosen such that the derivative of T_t at $t = 100$ is equal to zero.

O.8 Sensitivity with respect to Inequality Aversion (see Appendix G.1)

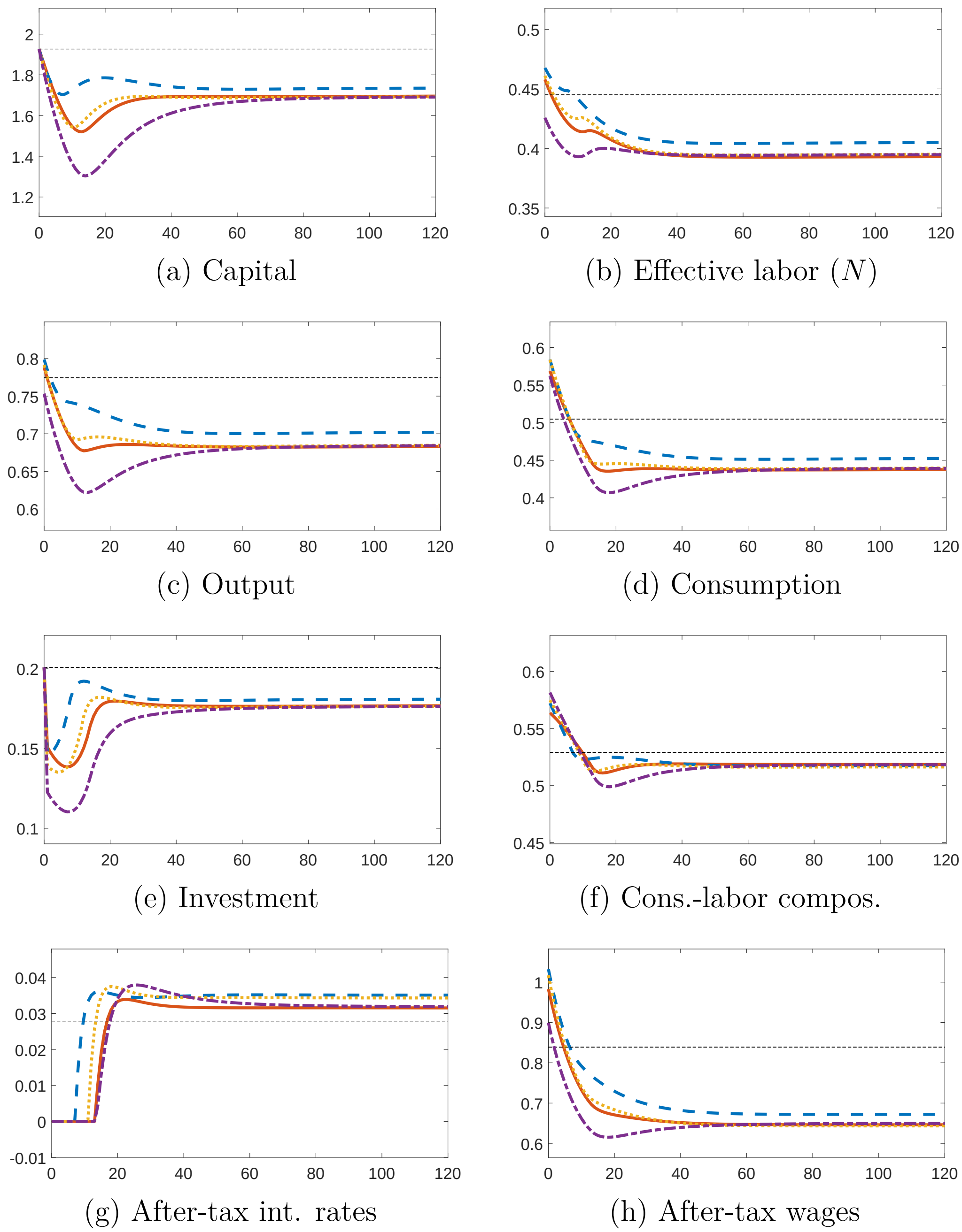


(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 46: Optimal Fiscal Policy: Sensitivity with respect to Inequality Aversion

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition with benchmark inequality aversion of 1.55; Blue dashed curve: optimal transition with inequality aversion equal to 0; Yellow dotted curve: optimal transition with inequality aversion equal to 1; Purple dash-dotted curve: optimal transition with inequality aversion equal to 10.

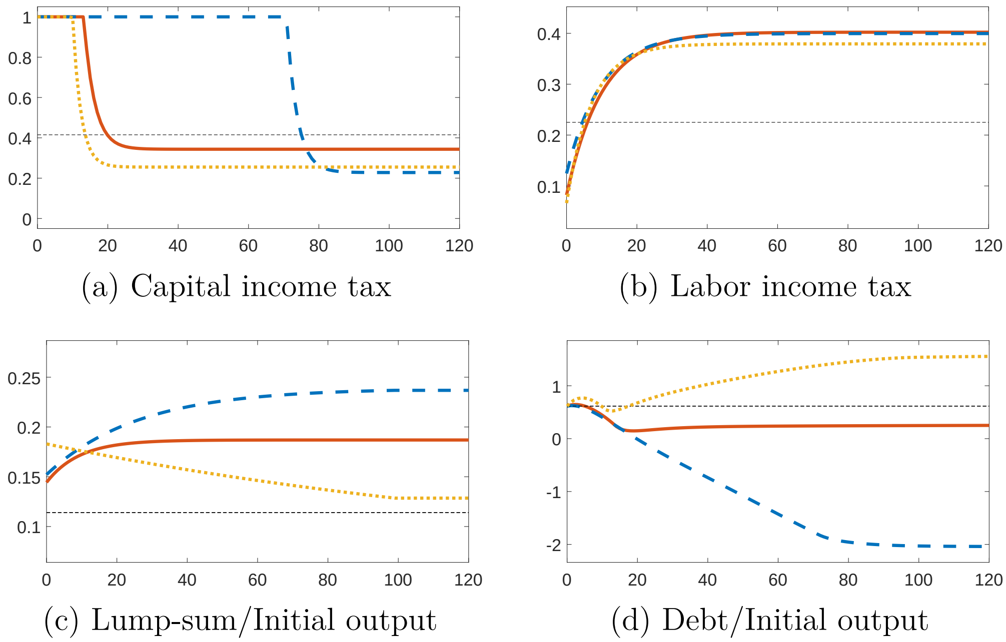


(a) Capital
(b) Effective labor (N)
(c) Output
(d) Consumption
(e) Investment
(f) Cons.-labor compos.
(g) After-tax int. rates
(h) After-tax wages

Figure 47: Aggregates: Sensitivity with respect to Inequality Aversion

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition with benchmark inequality aversion of 1.55; Blue dashed curve: optimal transition with inequality aversion equal to 0; Yellow dotted curve: optimal transition with inequality aversion equal to 1; Purple dash-dotted curve: optimal transition with inequality aversion equal to 10.

O.9 Sensitivity with respect to Intertemporal Elasticity of Substitution (see Appendix G.2)



(a) Capital income tax

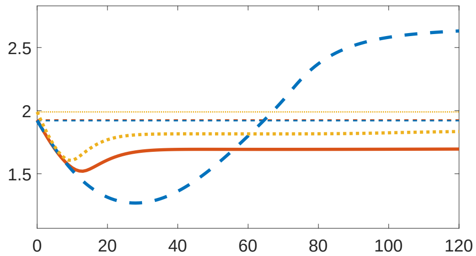
(b) Labor income tax

(c) Lump-sum/Initial output

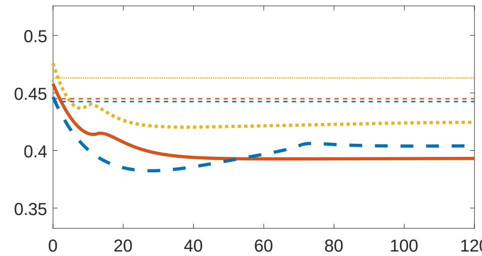
(d) Debt/Initial output

Figure 48: Optimal Fiscal Policy: Sensitivity with respect to IES

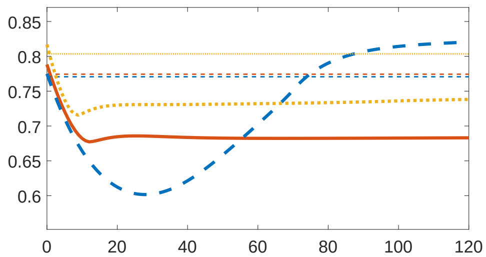
Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition with benchmark IES of 0.65; Blue dashed curve: optimal transition with IES equal to 0.5; Yellow dotted curve: optimal transition with IES equal to 0.8.



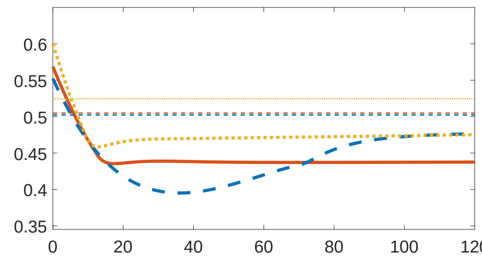
(a) Capital



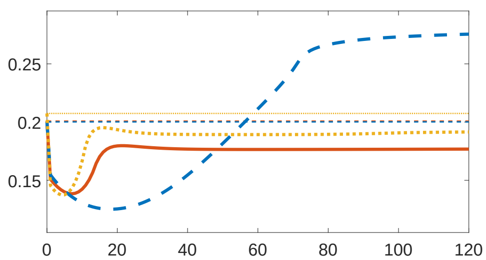
(b) Effective labor (N)



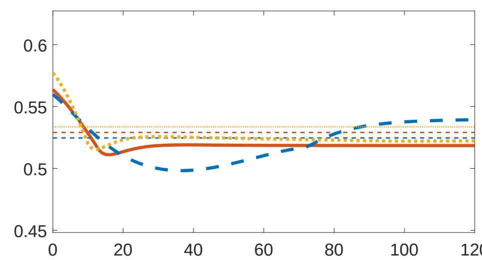
(c) Output



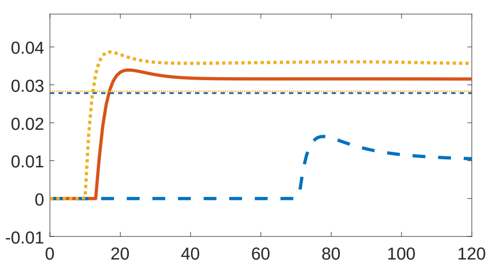
(d) Consumption



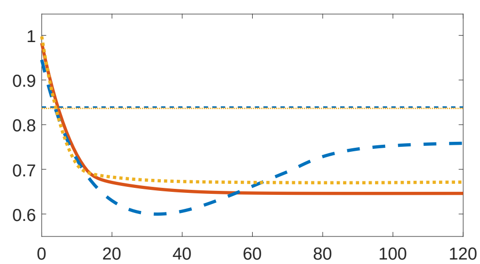
(e) Investment



(f) Cons.-labor compos.



(g) After-tax int. rates

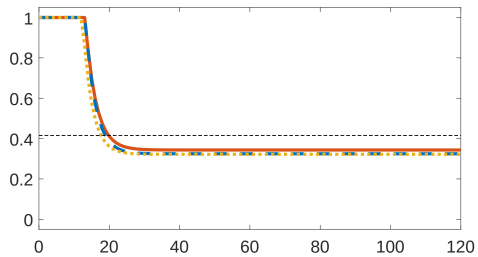


(h) After-tax wages

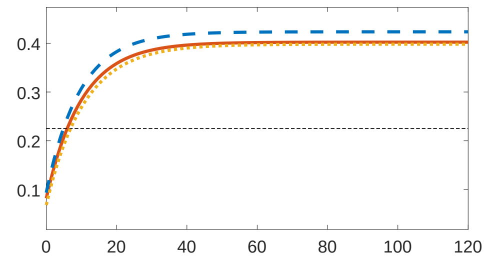
Figure 49: Aggregates: Sensitivity with respect to IES

Notes: Red solid curve: optimal transition with benchmark IES of 0.65; Blue dashed curve: optimal transition with IES equal to 0.5; Yellow dotted curve: optimal transition with IES equal to 0.8; Thin dashed lines: corresponding values in initial stationary equilibrium.

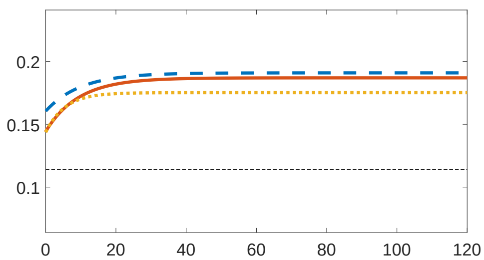
O.10 Sensitivity with respect to Frisch Elasticity (see Appendix G.2)



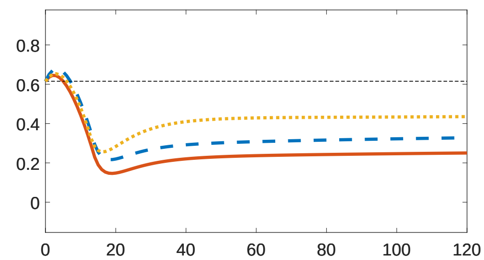
(a) Capital income tax



(b) Labor income tax



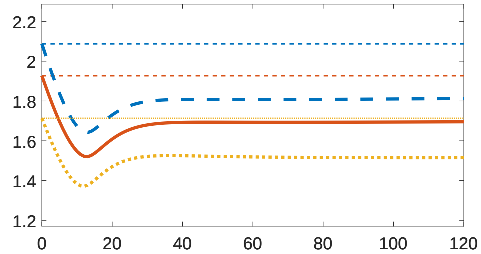
(c) Lump-sum/Initial output



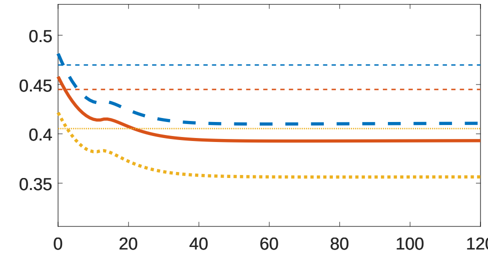
(d) Debt/Initial output

Figure 50: Optimal Fiscal Policy: Sensitivity with respect to Frisch

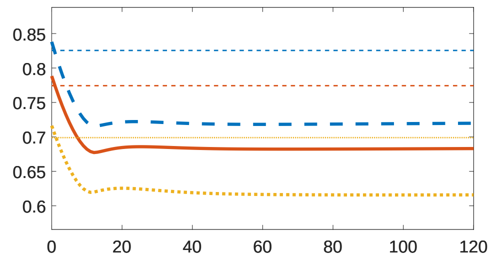
Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal transition with benchmark Frisch of 0.6; Blue dashed curve: optimal transition with Frisch equal to 0.45; Yellow dotted curve: optimal transition with Frisch equal to 0.75



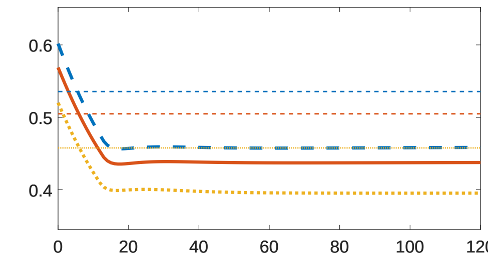
(a) Capital



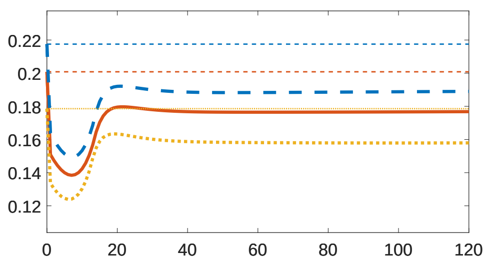
(b) Effective labor (N)



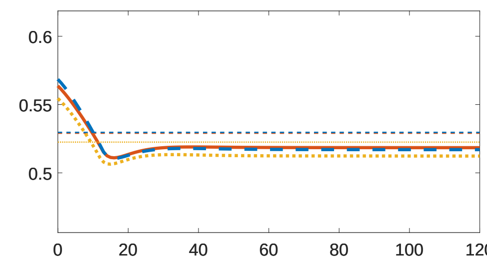
(c) Output



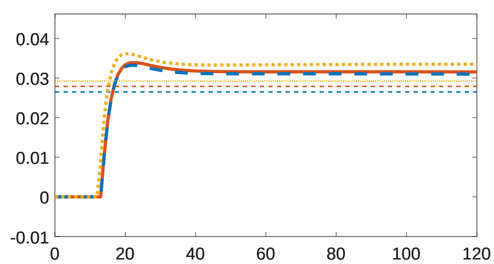
(d) Consumption



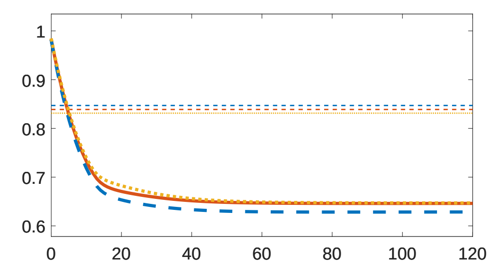
(e) Investment



(f) Cons.-labor compos.



(g) After-tax int. rates

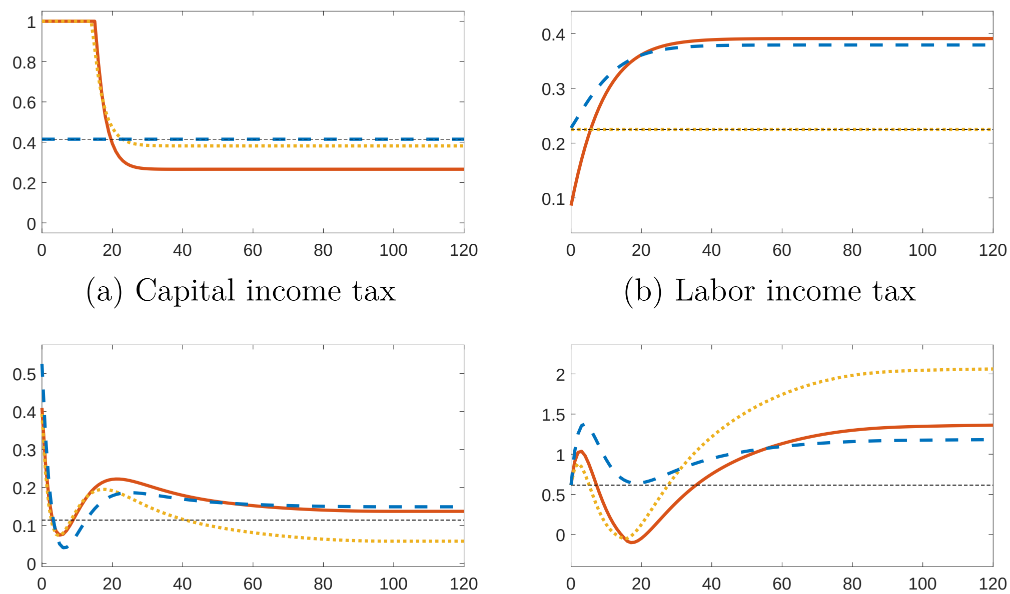


(h) After-tax wages

Figure 51: Aggregates: Sensitivity with respect to Frisch

Notes: Red solid curve: optimal transition with benchmark Frisch of 0.6; Blue dashed curve: optimal transition with Frisch equal to 0.45; Yellow dotted curve: optimal transition with Frisch equal to 0.75; Thin dashed lines: corresponding values in initial stationary equilibrium.

O.11 Fixed Capital and Labor Income Taxes Experiments (see Section 5.4)

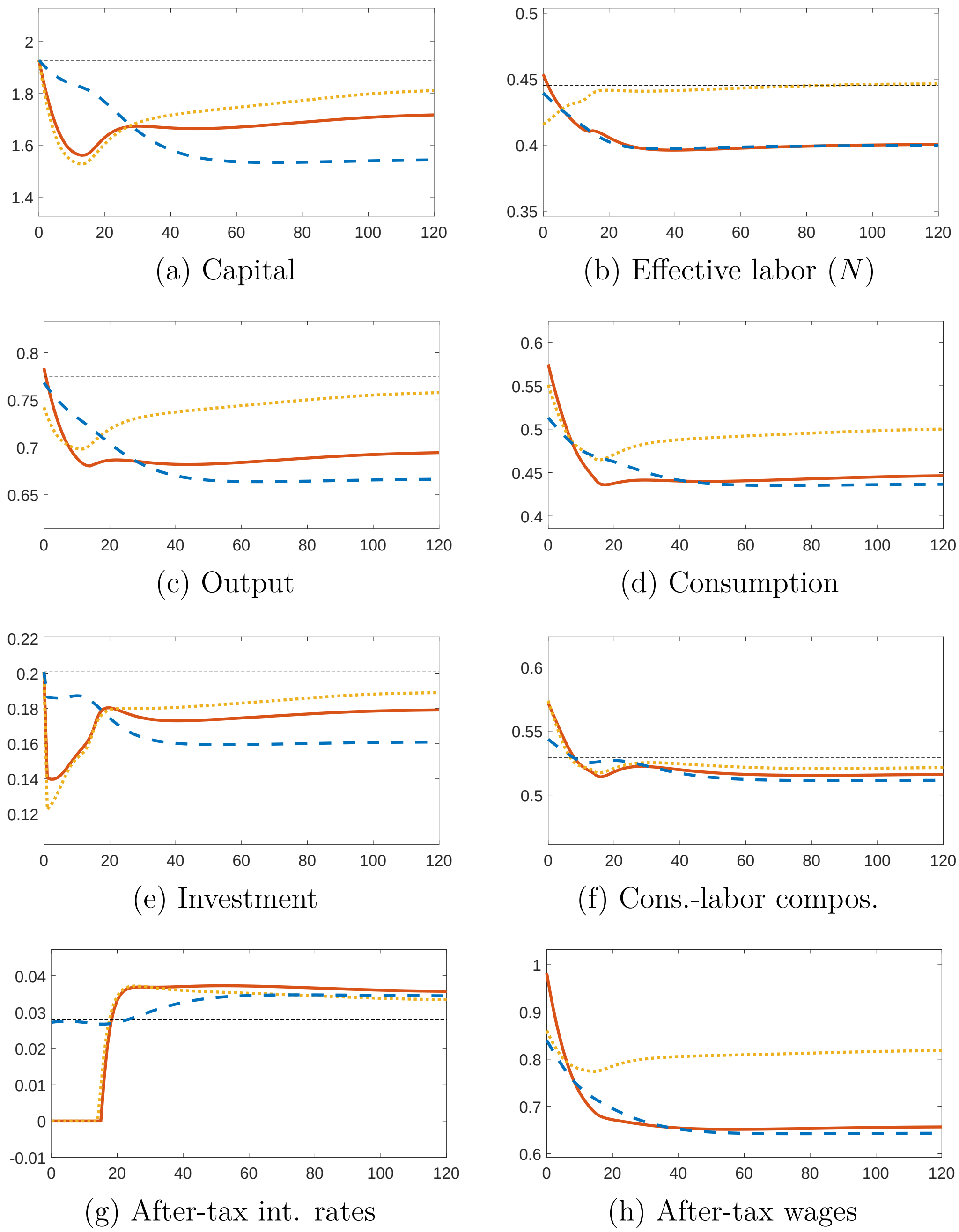


(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 52: Optimal Fiscal Policy: Fixed Capital and Labor Income Taxes Experiments

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal benchmark experiment; Blue dashed curve: reoptimized transition with capital income taxes fixed at their initial pre-reform level; Yellow dotted curve: reoptimized transition with labor income taxes fixed at their initial pre-reform level.



(a) Capital (b) Effective labor (N)

(c) Output (d) Consumption

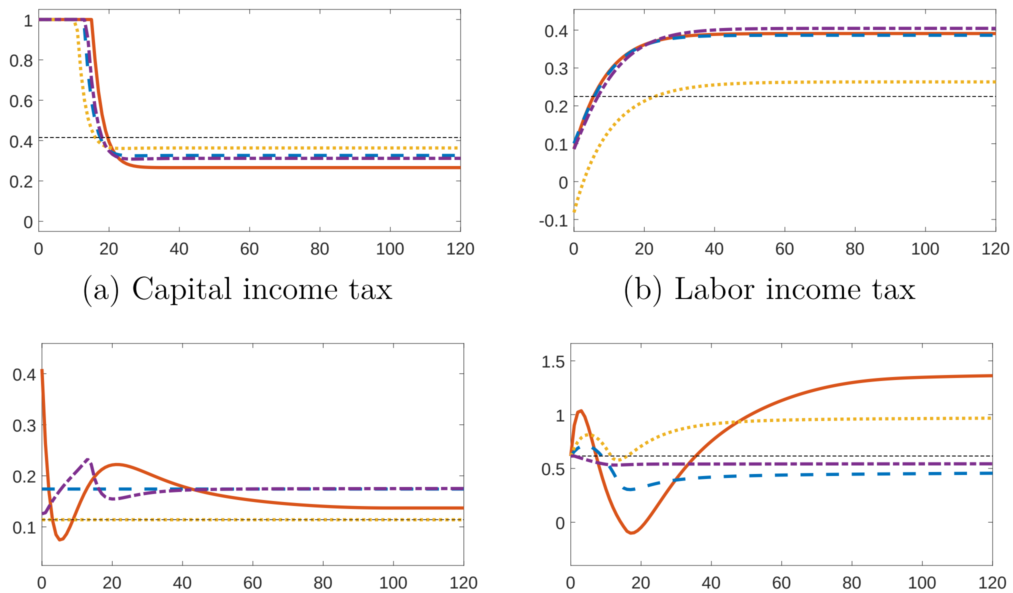
(e) Investment (f) Cons.-labor compos.

(g) After-tax int. rates (h) After-tax wages

Figure 53: Aggregates: Fixed Capital and Labor Income Taxes Experiments

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal benchmark experiment; Blue dashed curve: reoptimized transition with capital income taxes fixed at their initial pre-reform level; Yellow dotted curve: reoptimized transition with labor income taxes fixed at their initial pre-reform level.

O.12 Fixed Lump-Sum and Debt Experiments (see Section 5.4)

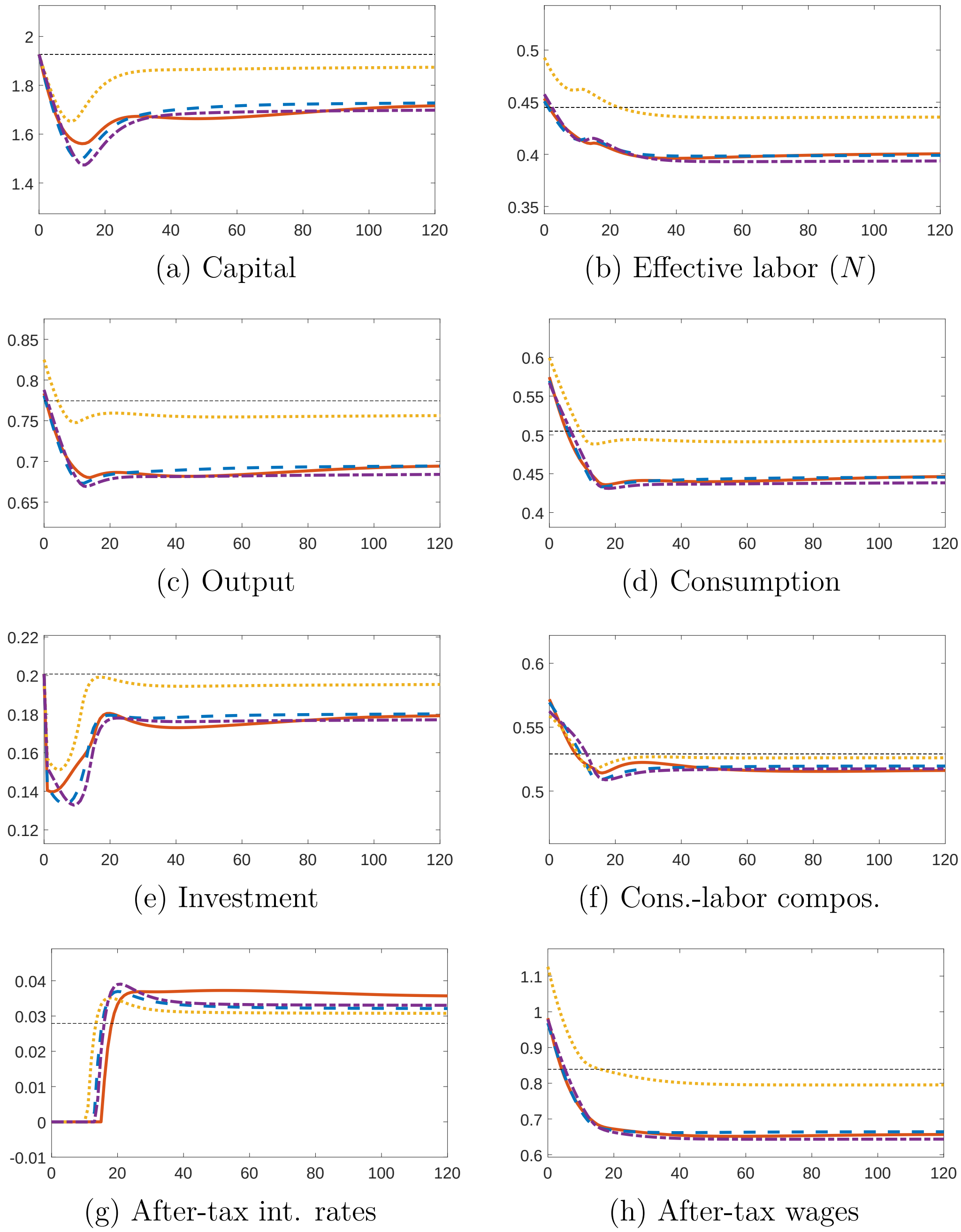


(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 54: Optimal Fiscal Policy: Fixed Lump-Sum and Debt Experiments

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal benchmark experiment; Blue dashed curve: reoptimized transition with lump-sum transfers constrained to be constant after an initial movement in period 0; Yellow dotted curve: reoptimized transition with lump-sum transfers fixed at their initial pre-reform level; Purple dash-dotted curve: reoptimized transition with debt-to-output fixed at their initial pre-reform level.



(a) Capital (b) Effective labor (N)

(c) Output (d) Consumption

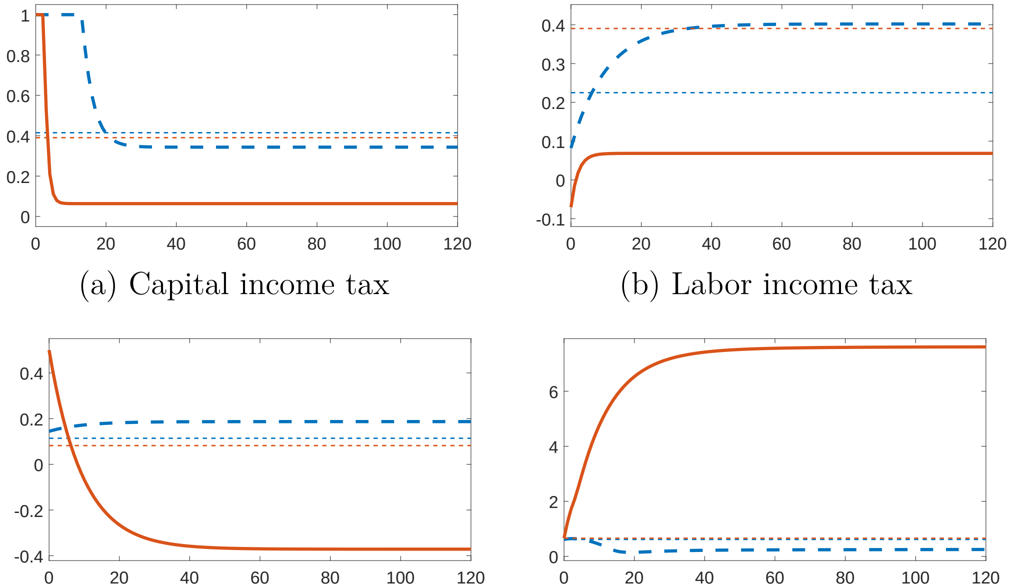
(e) Investment (f) Cons.-labor compos.

(g) After-tax int. rates (h) After-tax wages

Figure 55: Aggregates: Fixed Lump-Sum and Debt Experiments

Notes: Black dashed line: initial stationary equilibrium; Red solid curve: optimal benchmark experiment; Blue dashed curve: reoptimized transition with lump-sum transfers constrained to be constant after an initial movement in period 0; Yellow dotted curve: reoptimized transition with lump-sum transfers fixed at their initial pre-reform level; Purple dash-dotted curve: reoptimized transition with debt-to-output fixed at their initial pre-reform level.

O.13 Calibration from Aiyagari and McGrattan (1998) (see Appendix N.1)



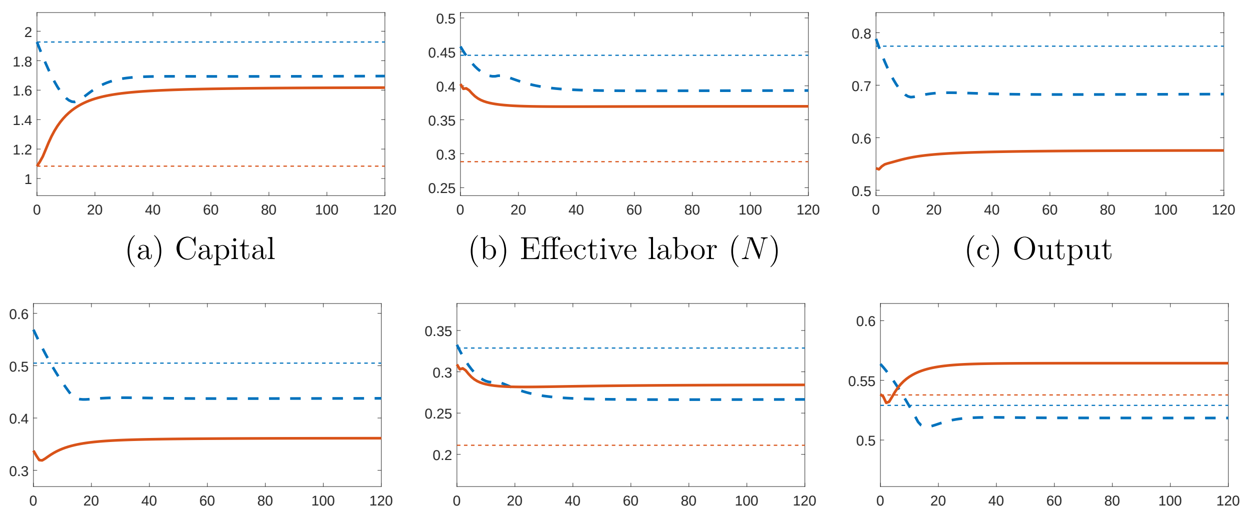
(a) Capital income tax

(b) Labor income tax

(c) Lump-sum/Initial output

(d) Debt/Initial output

Figure 56: Optimal Fiscal Policy: Calibration from Aiyagari and McGrattan (1998)



(a) Capital

(b) Effective labor (N)

(c) Output

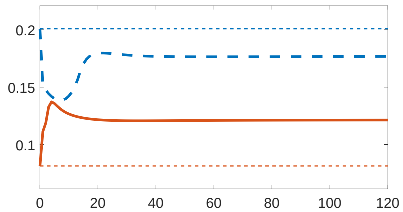
(d) Consumption

(e) Hours (H)

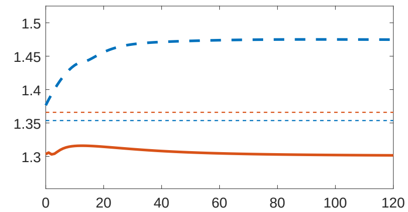
(f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 57: Aggregates: Calibration from Aiyagari and McGrattan (1998) (1)

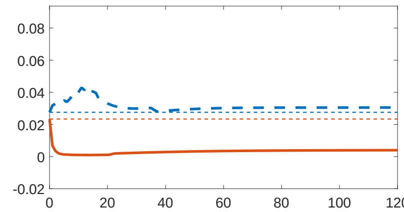
Notes: Red solid curve: optimal transition for calibration from Aiyagari and McGrattan (1998); Blue dashed curve: optimal transition with 8 variables for benchmark calibration; Thin dashed lines: corresponding values in initial stationary equilibrium.



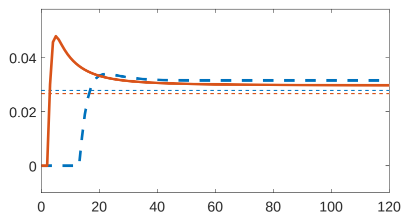
(g) Investment



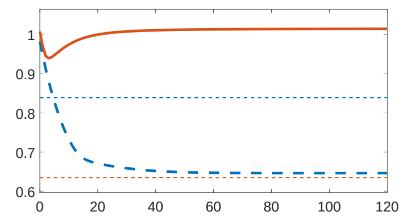
(h) Labor productivity



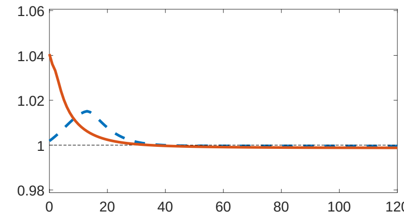
(i) Prop. borrow. constrained



(j) After-tax int. rates

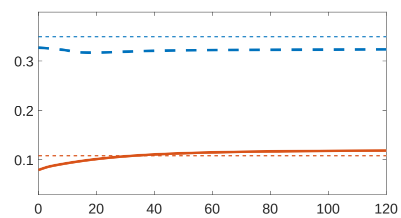


(k) After-Tax Wages

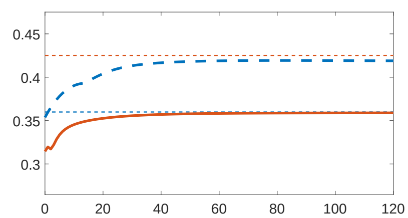


(l) Modified golden rule

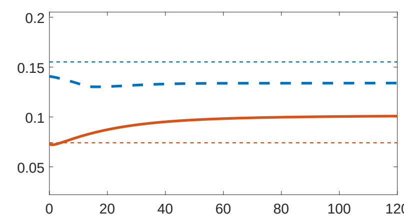
Figure 57: Aggregates: Calibration from Aiyagari and McGrattan (1998) (2)



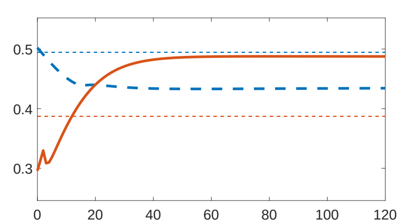
(m) Consumption Gini



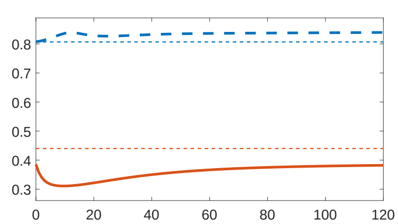
(n) Hours Gini



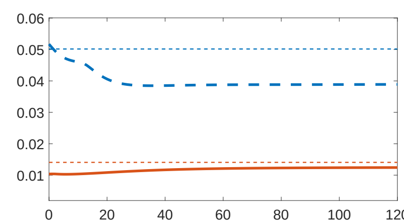
(o) Gini of $c^{\gamma}(1-h)^{1-\gamma}$



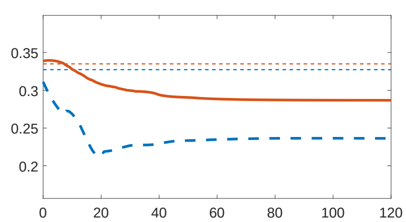
(p) After-tax income Gini



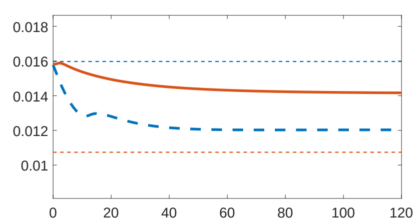
(q) Wealth Gini



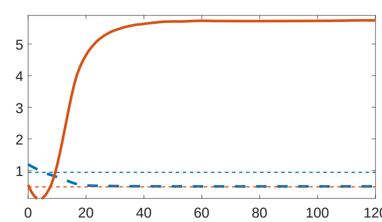
(r) Var. cons. growth



(s) Var. hours growth



(t) Var. $c^{\gamma}(1-h)^{1-\gamma}$ growth



(u) Var. a-t income growth

Figure 58: Inequality and Risk: Calibration from Aiyagari and McGrattan (1998)

Notes: Red solid curve: optimal transition for calibration from Aiyagari and McGrattan (1998); Blue dashed curve: optimal transition with 8 variables for benchmark calibration; Thin dashed lines: corresponding values in initial stationary equilibrium.

O.14 No-Inequality-Targets Calibration (see Appendix N.2)

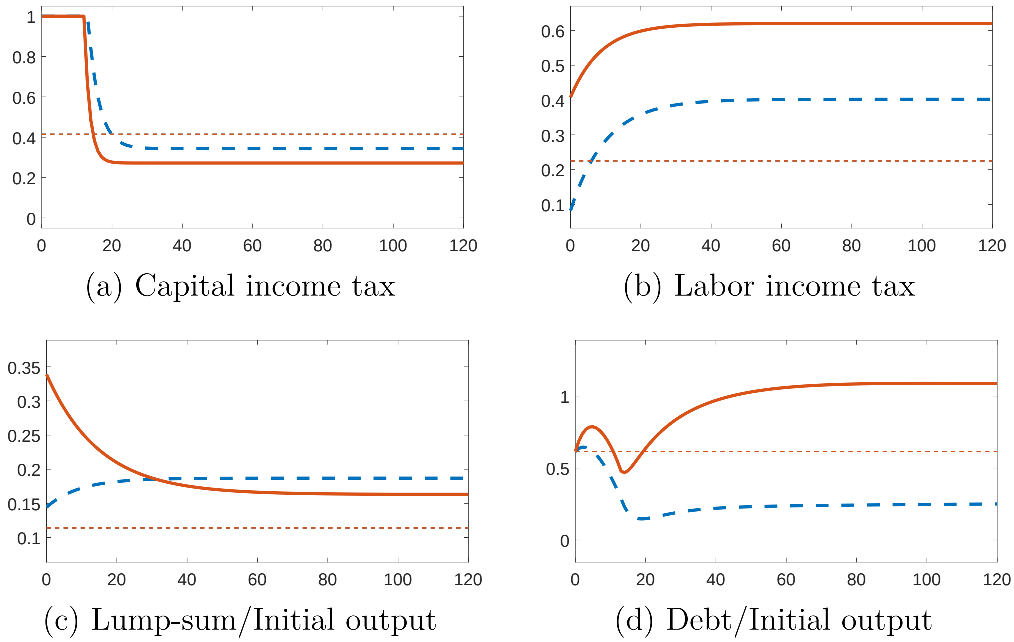


Figure 59: Optimal Fiscal Policy: No-Inequality-Targets Calibration

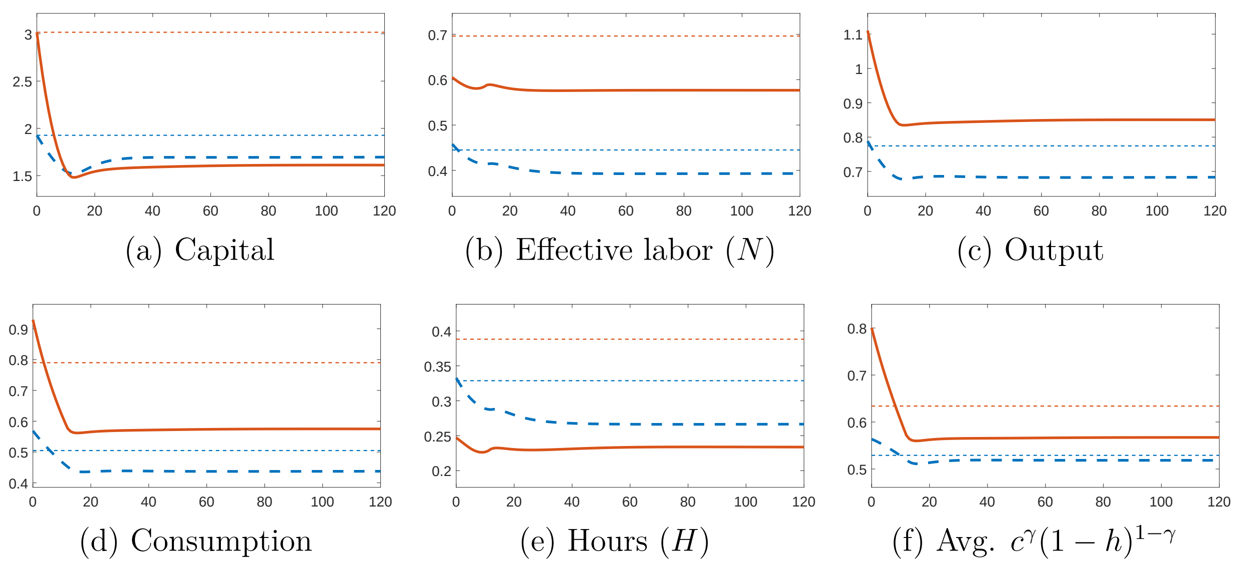
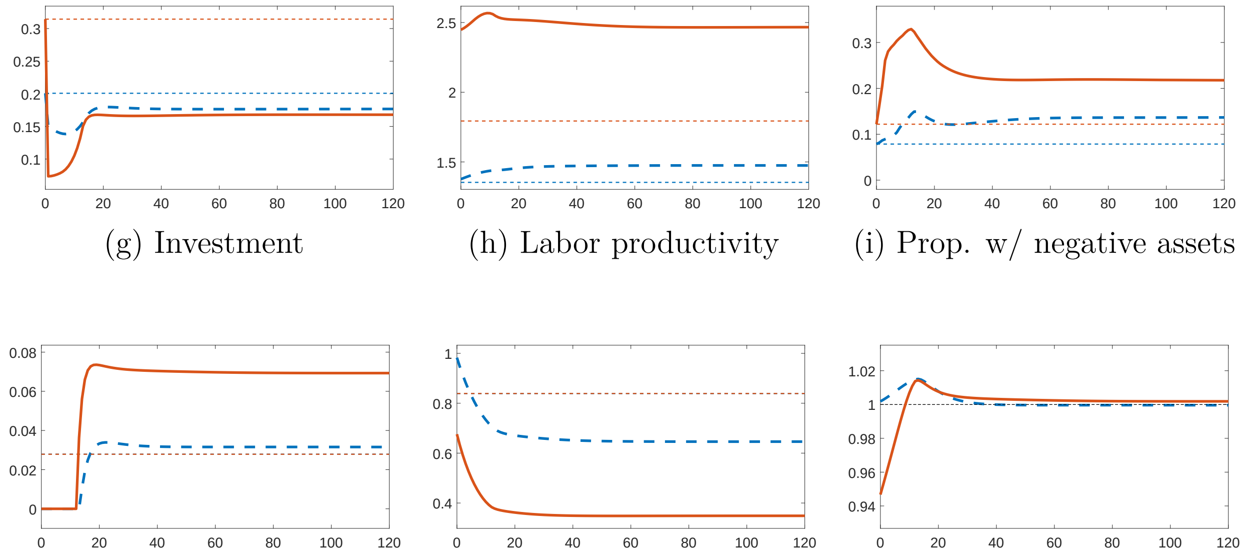


Figure 60: Aggregates: No-Inequality-Targets Calibration (1)

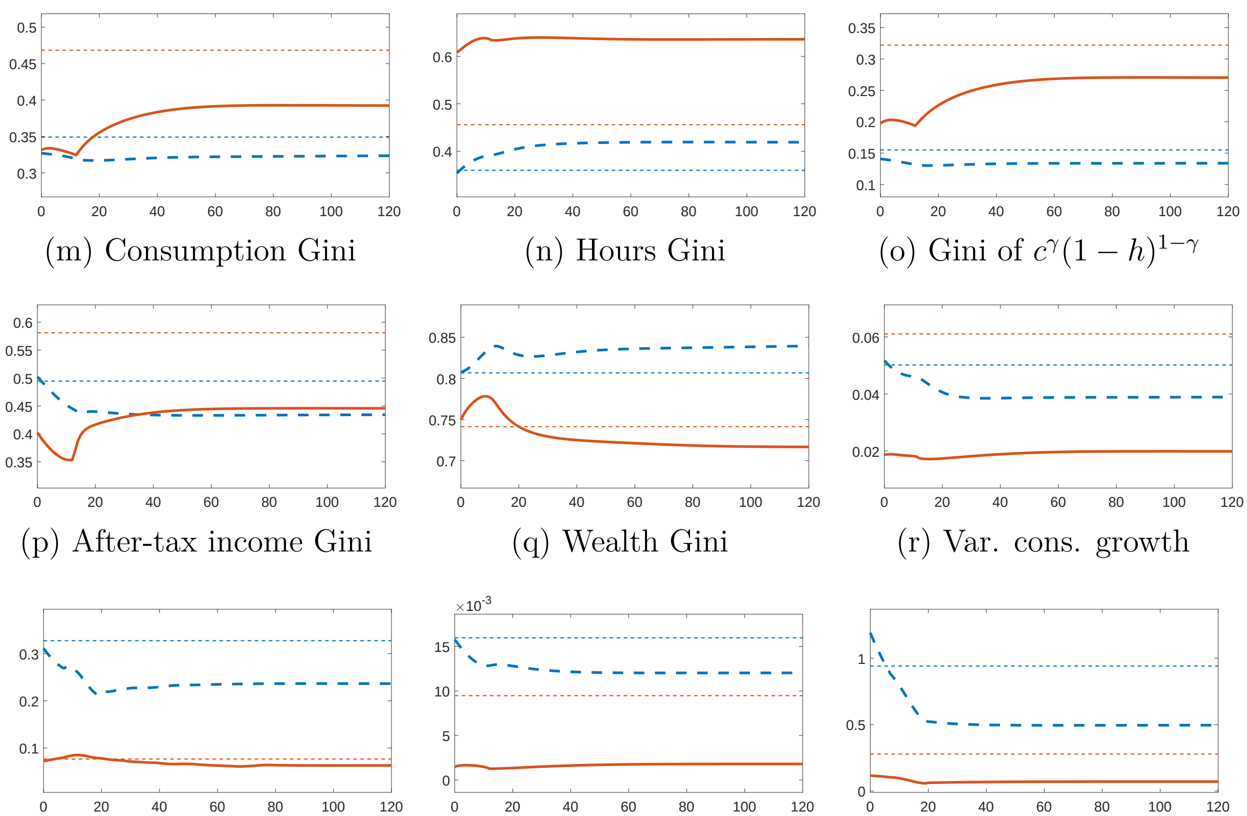
Notes: Red solid curve: optimal transition for calibration without inequality targets; Blue dashed curve: optimal transition with 8 variables for benchmark calibration; Thin dashed lines: corresponding values in initial stationary equilibrium.



(g) Investment (h) Labor productivity (i) Prop. w/ negative assets

(j) After-tax int. rates (k) After-Tax Wages (l) Modified golden rule

Figure 60: Aggregates: No-Inequality-Targets Calibration (2)



(m) Consumption Gini (n) Hours Gini (o) Gini of $c^\gamma(1-h)^{1-\gamma}$

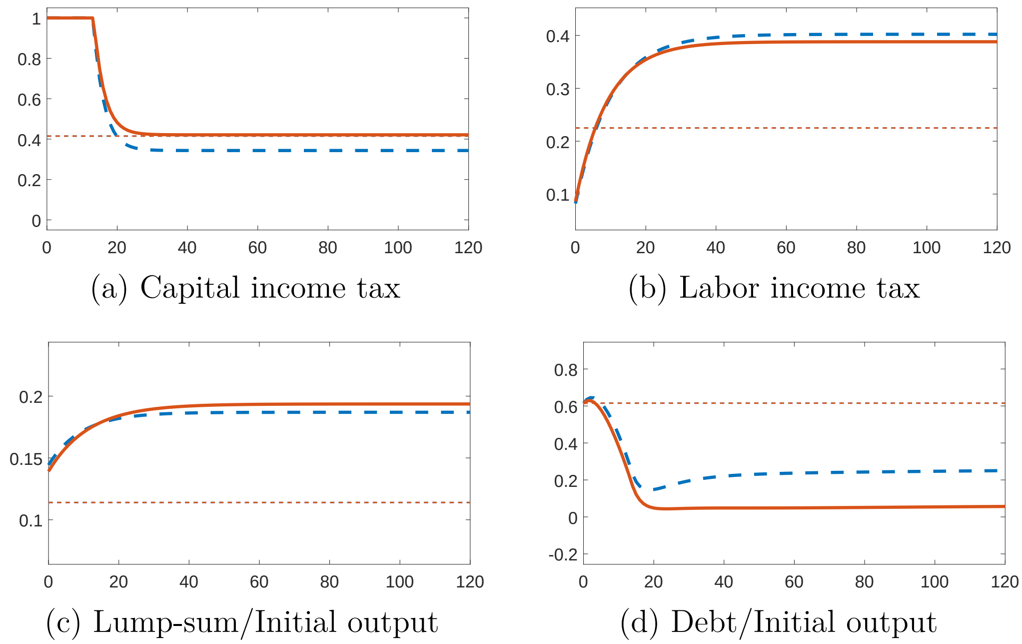
(p) After-tax income Gini (q) Wealth Gini (r) Var. cons. growth

(s) Var. hours growth (t) Var. $c^\gamma(1-h)^{1-\gamma}$ growth (u) Var. a-t income growth

Figure 61: Inequality and Risk: No-Inequality-Targets Calibration

Notes: Red solid curve: optimal transition for calibration without inequality targets; Blue dashed curve: optimal transition with 8 variables for benchmark calibration; Thin dashed lines: corresponding values in initial stationary equilibrium.

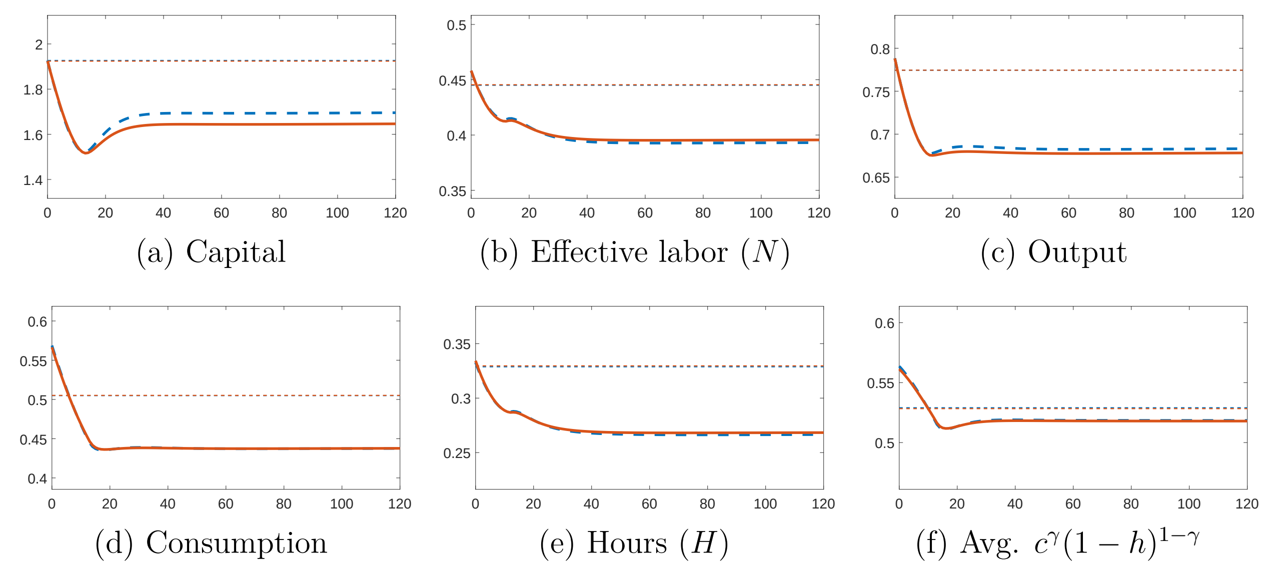
O.15 Return-Risk Calibration (see Appendix N.3)



(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 62: Optimal Fiscal Policy: Return-Risk Calibration

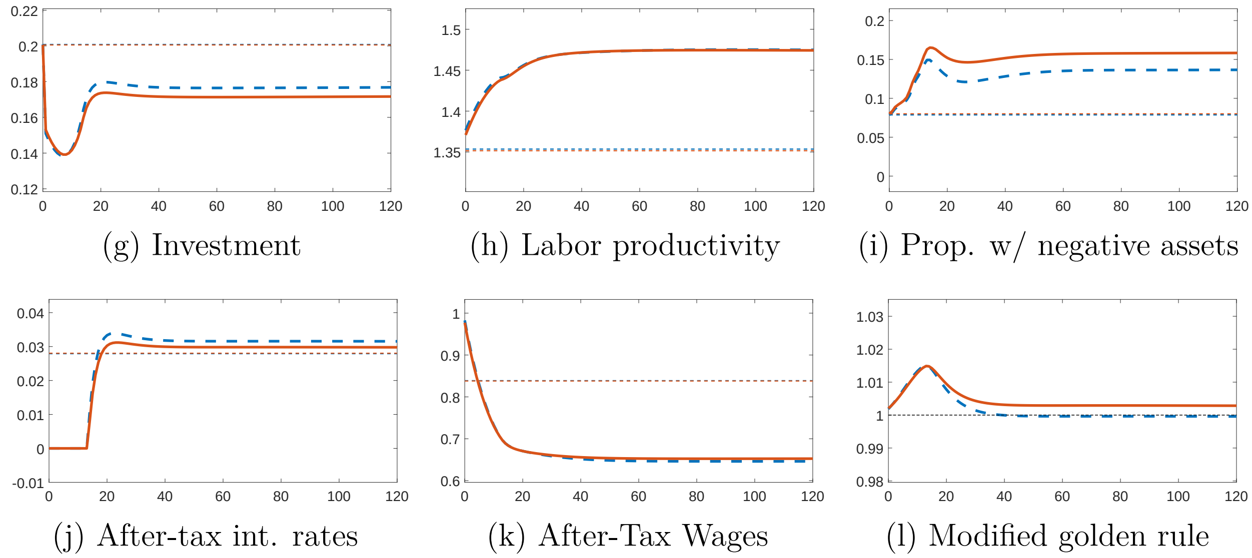


(a) Capital (b) Effective labor (N) (c) Output

(d) Consumption (e) Hours (H) (f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 63: Aggregates: Return-Risk Calibration (1)

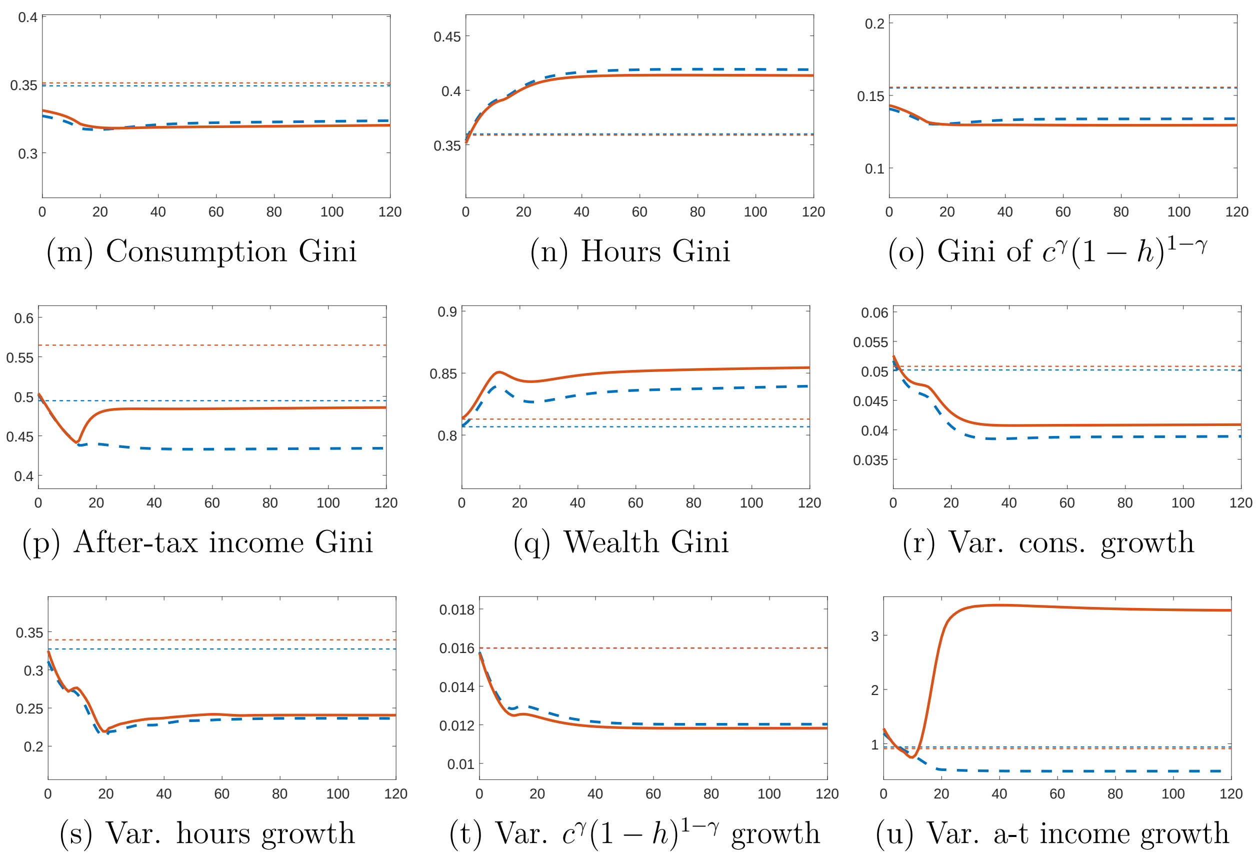
Notes: Red solid curve: optimal transition for calibration with return risk; Blue dashed curve: optimal transition with 8 variables for benchmark calibration; Thin dashed lines: corresponding values in initial stationary equilibrium.



(g) Investment (h) Labor productivity (i) Prop. w/ negative assets

(j) After-tax int. rates (k) After-Tax Wages (l) Modified golden rule

Figure 63: Aggregates: Return-Risk Calibration (2)



(m) Consumption Gini (n) Hours Gini (o) Gini of $c^{\gamma}(1-h)^{1-\gamma}$

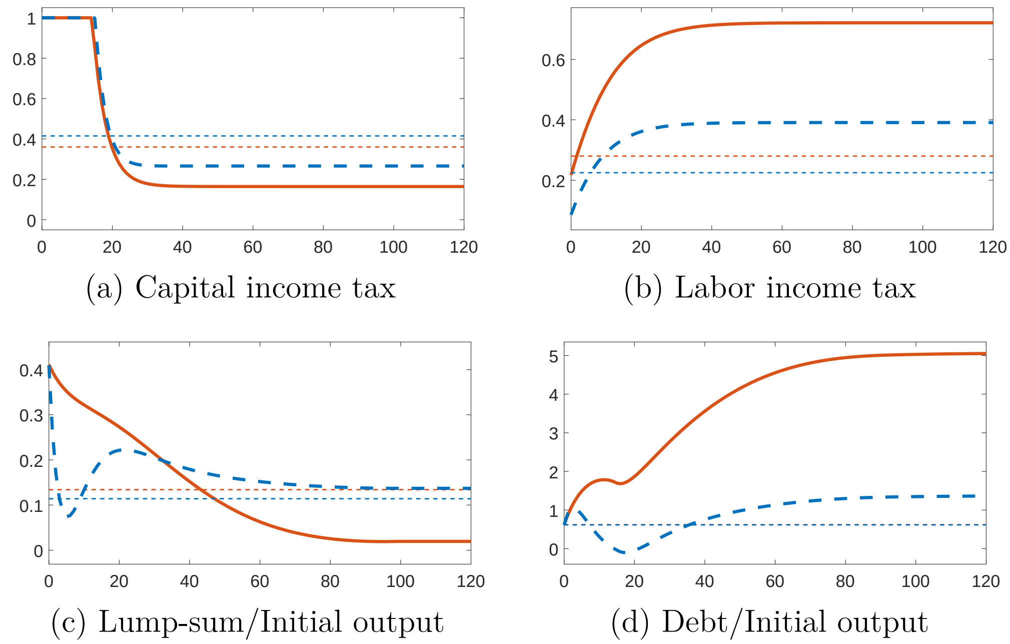
(p) After-tax income Gini (q) Wealth Gini (r) Var. cons. growth

(s) Var. hours growth (t) Var. $c^{\gamma}(1-h)^{1-\gamma}$ growth (u) Var. a-t income growth

Figure 64: Inequality and Risk: Return-Risk Calibration

Notes: Red solid curve: optimal transition for calibration with return risk; Blue dashed curve: optimal transition with 8 variables for benchmark calibration; Thin dashed lines: corresponding values in initial stationary equilibrium.

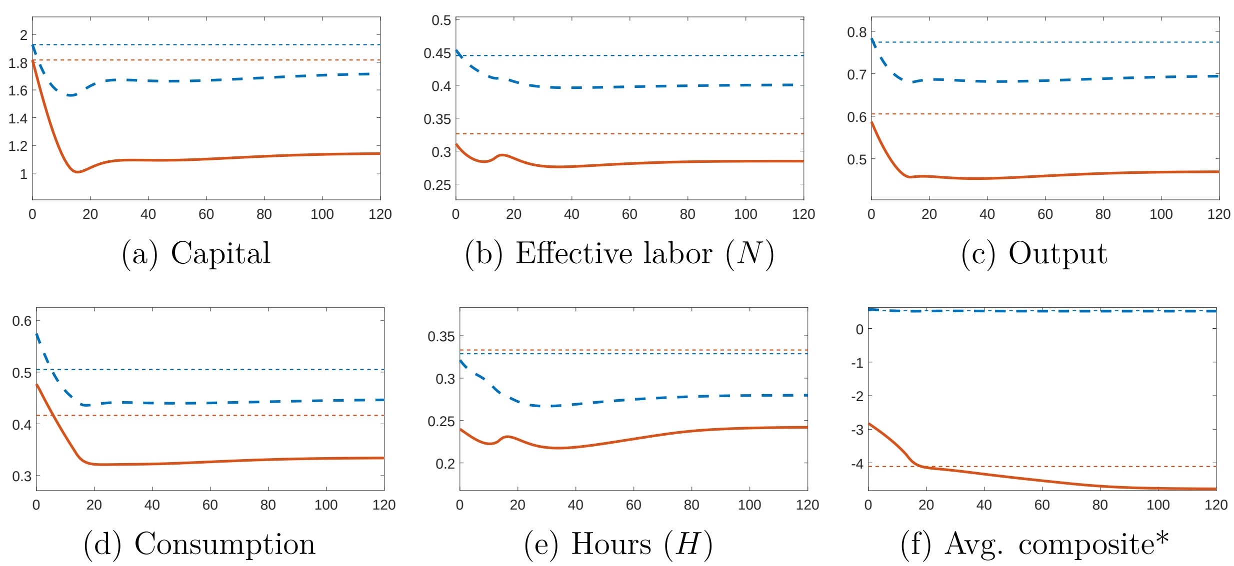
O.16 Calibration from Acikgoz, Hagedorn, Holter, and Wang (2018) (see Appendix M)



(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 65: Optimal Fiscal Policy: Calibration from Acikgoz et al. (2018)

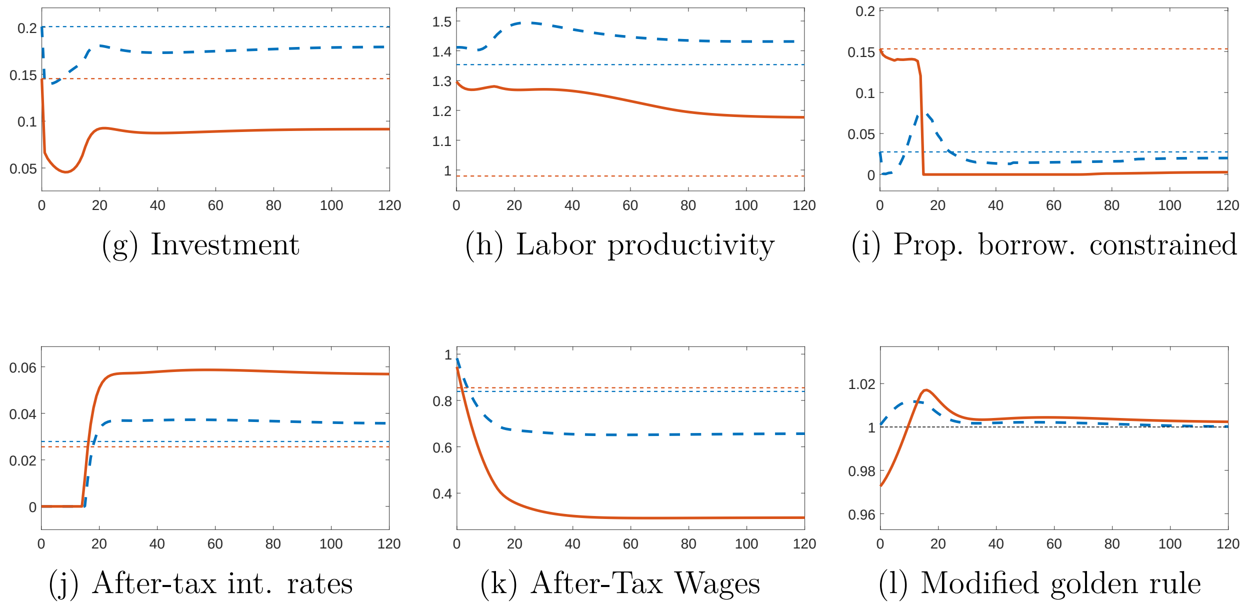


(a) Capital (b) Effective labor (N) (c) Output

(d) Consumption (e) Hours (H) (f) Avg. composite*

Figure 66: Aggregates: Calibration from Acikgoz et al. (2018) (1)

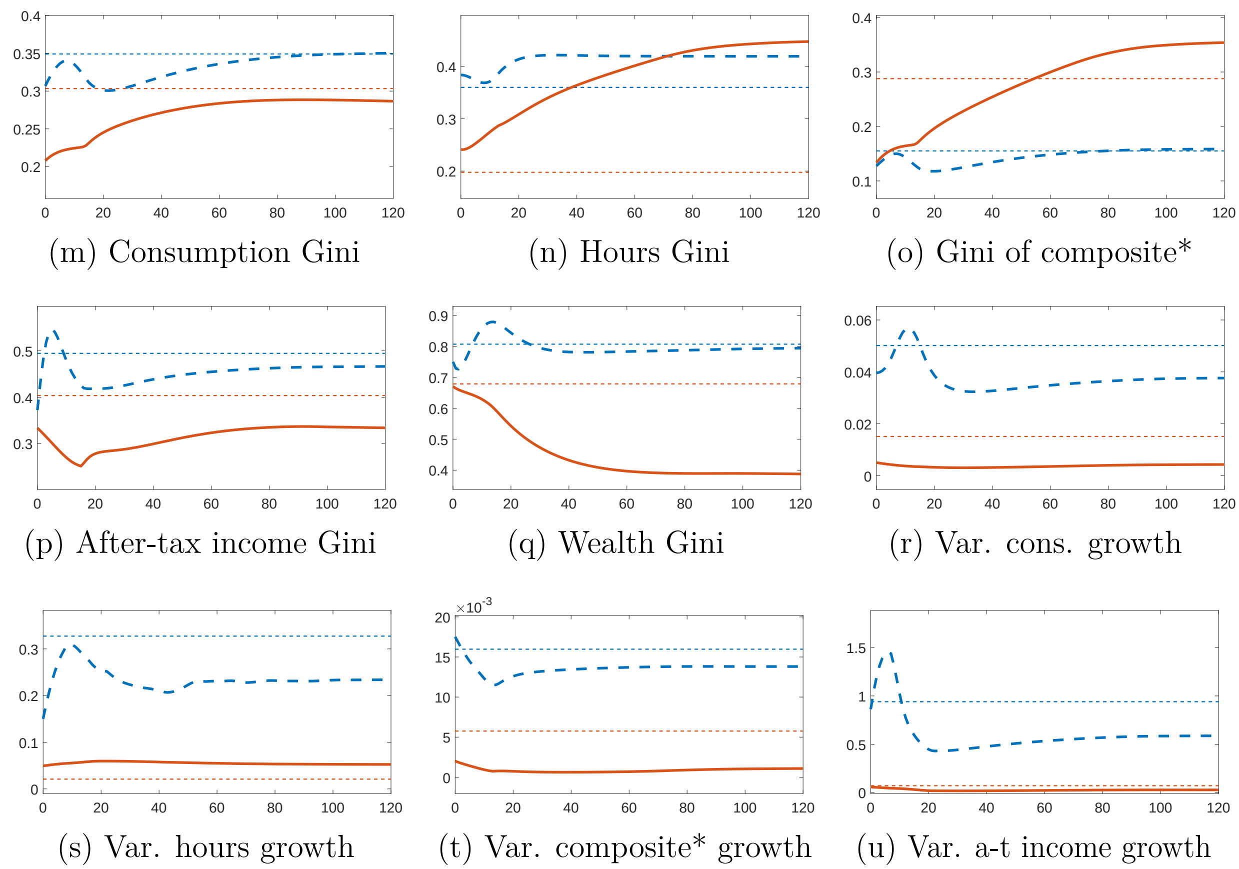
Notes: Red solid curve: optimal transition for calibration from Acikgoz et al. (2018); Blue dashed curve: optimal transition (benchmark); Thin dashed lines: corresponding values in initial stationary equilibrium. *Composite is $\frac{c^{1-\sigma}}{1-\sigma} - \chi \frac{h^{1+1/\phi}}{1+1/\phi}$ for Acikgoz et al. (2018) and $c^{\gamma}(1-h)^{1-\gamma}$ for the benchmark calibration.



(g) Investment (h) Labor productivity (i) Prop. borrow. constrained

(j) After-tax int. rates (k) After-Tax Wages (l) Modified golden rule

Figure 66: Aggregates: Calibration from Acikgoz et al. (2018) (2)



(m) Consumption Gini (n) Hours Gini (o) Gini of composite*

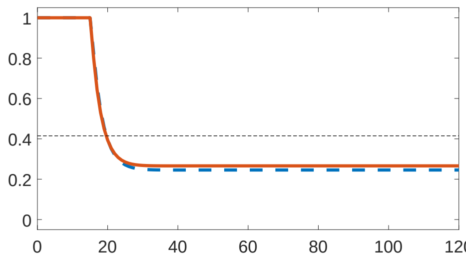
(p) After-tax income Gini (q) Wealth Gini (r) Var. cons. growth

(s) Var. hours growth (t) Var. composite* growth (u) Var. a-t income growth

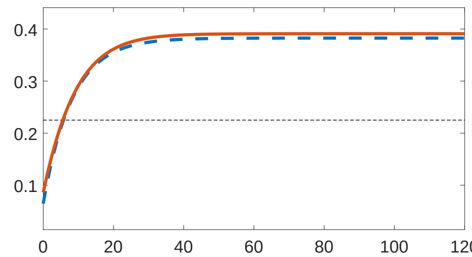
Figure 67: Inequality and Risk: Calibration from Acikgoz et al. (2018)

Notes: Red solid curve: optimal transition for calibration from Acikgoz et al. (2018); Blue dashed curve: optimal transition (benchmark); Thin dashed lines: corresponding values in initial stationary equilibrium. *Composite is $\frac{c^{1-\sigma}}{1-\sigma} - \chi\frac{h^{1+1/\phi}}{1+1/\phi}$ for Acikgoz et al. (2018) and $c^{\gamma}(1-h)^{1-\gamma}$ for the benchmark calibration.

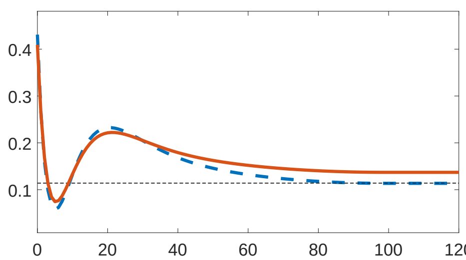
O.17 Benchmark Calibration: DP-AHHW Method Comparison (see Appendix M)



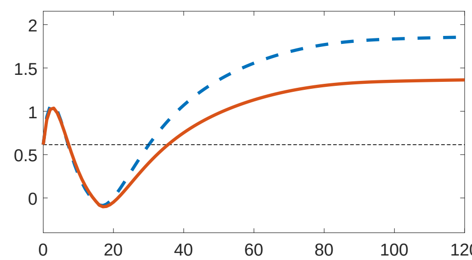
(a) Capital income tax



(b) Labor income tax

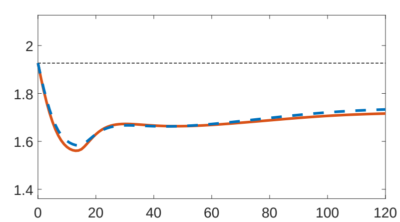


(c) Lump-sum/Initial output

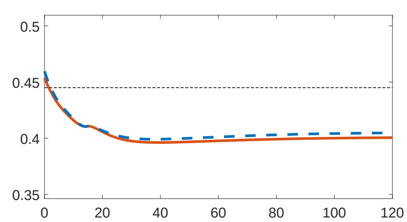


(d) Debt/Initial output

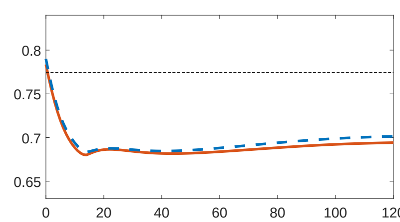
Figure 68: Optimal Fiscal Policy: DP-AHHW Method Comparison for Benchmark Calibration



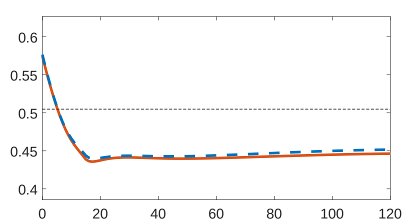
(a) Capital



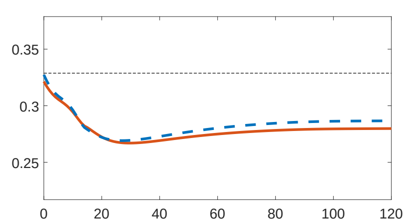
(b) Effective labor (N)



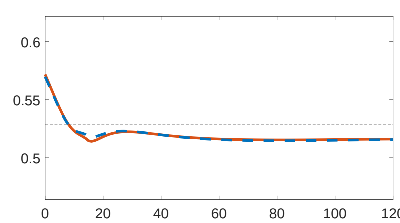
(c) Output



(d) Consumption



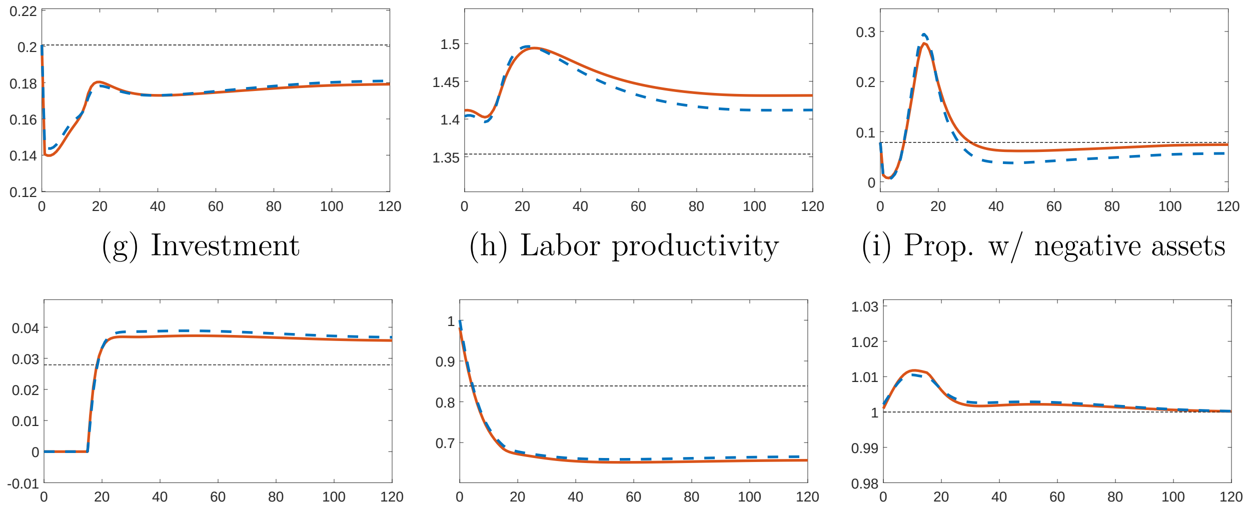
(e) Hours (H)



(f) Avg. $c^{\gamma}(1-h)^{1-\gamma}$

Figure 69: Aggregates: DP-AHHW Method Comparison for Benchmark Calibration (1)

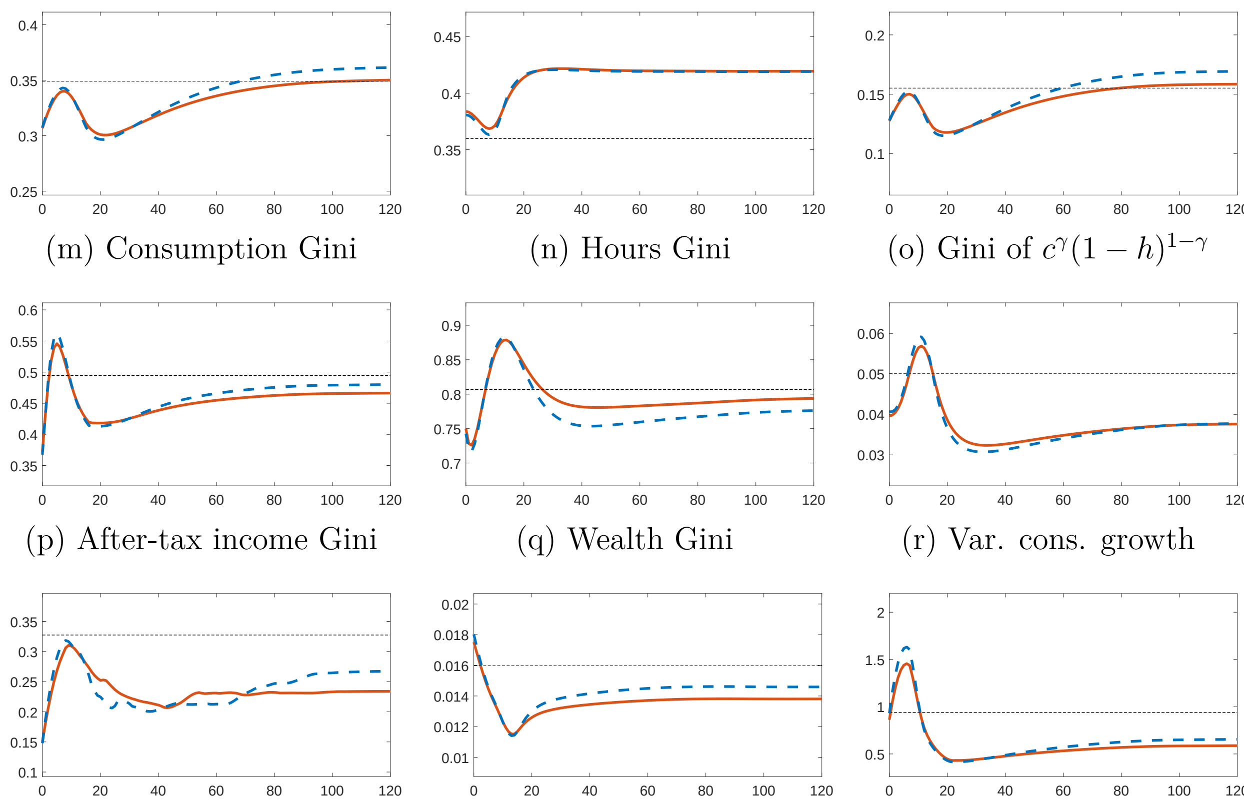
Notes: Red solid curve: optimal transition (benchmark); Blue dashed curve: optimal transition imposing the long-run policy obtained with the AHHW method; Thin dashed lines: corresponding values in initial stationary equilibrium.



(g) Investment (h) Labor productivity (i) Prop. w/ negative assets

(j) After-tax int. rates (k) After-Tax Wages (l) Modified golden rule

Figure 69: Aggregates: DP-AHHW Method Comparison for Benchmark Calibration (2)



(m) Consumption Gini (n) Hours Gini (o) Gini of $c^{\gamma}(1-h)^{1-\gamma}$

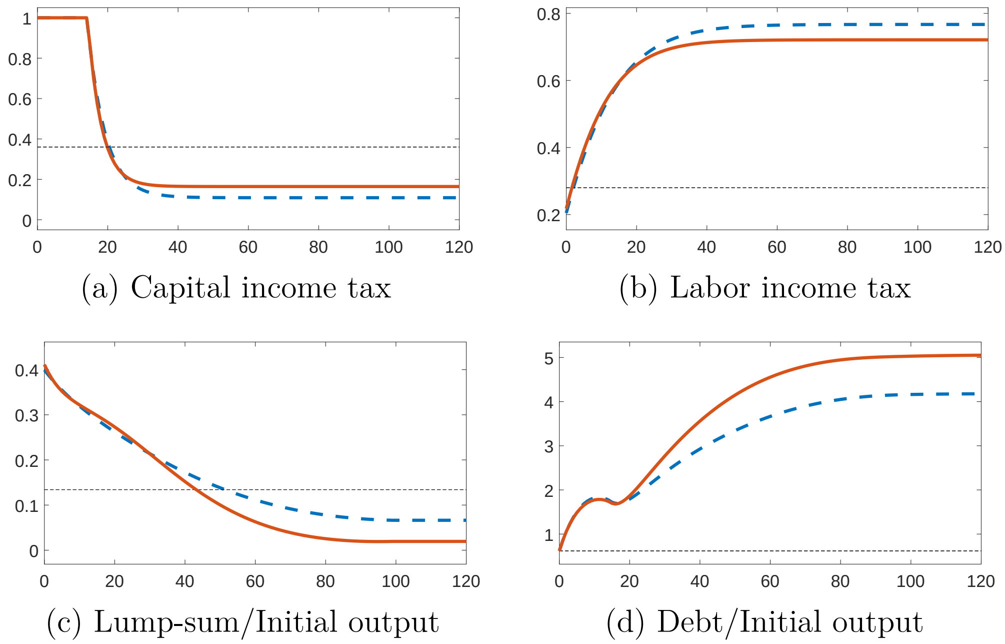
(p) After-tax income Gini (q) Wealth Gini (r) Var. cons. growth

(s) Var. hours growth (t) Var. $c^{\gamma}(1-h)^{1-\gamma}$ growth (u) Var. a-t income growth

Figure 70: Inequality and Risk: DP-AHHW Method Comparison for Benchmark Calibration

Notes: Red solid curve: optimal transition (benchmark); Blue dashed curve: optimal transition imposing the long-run policy obtained with the AHHW method; Thin dashed lines: corresponding values in initial stationary equilibrium.

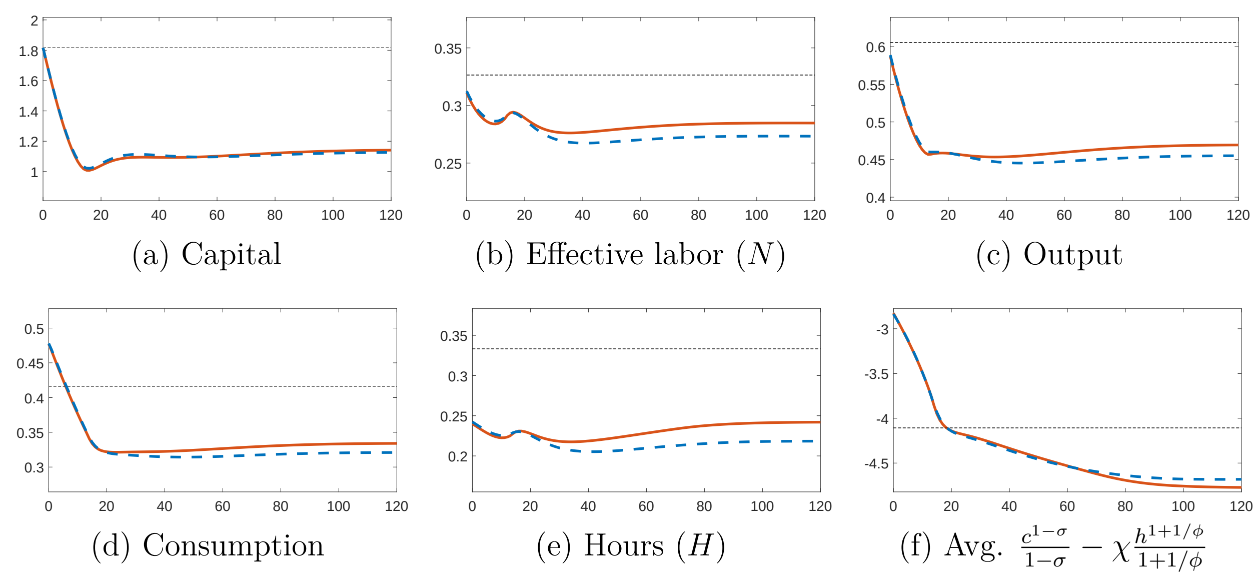
O.18 Calibration from Acikgoz et al. (2018): DP-AHHW Method Comparison (see Appendix M)



(a) Capital income tax (b) Labor income tax

(c) Lump-sum/Initial output (d) Debt/Initial output

Figure 71: Optimal Fiscal Policy: DP-AHHW Method Comparison for Calibration from Acikgoz et al. (2018)

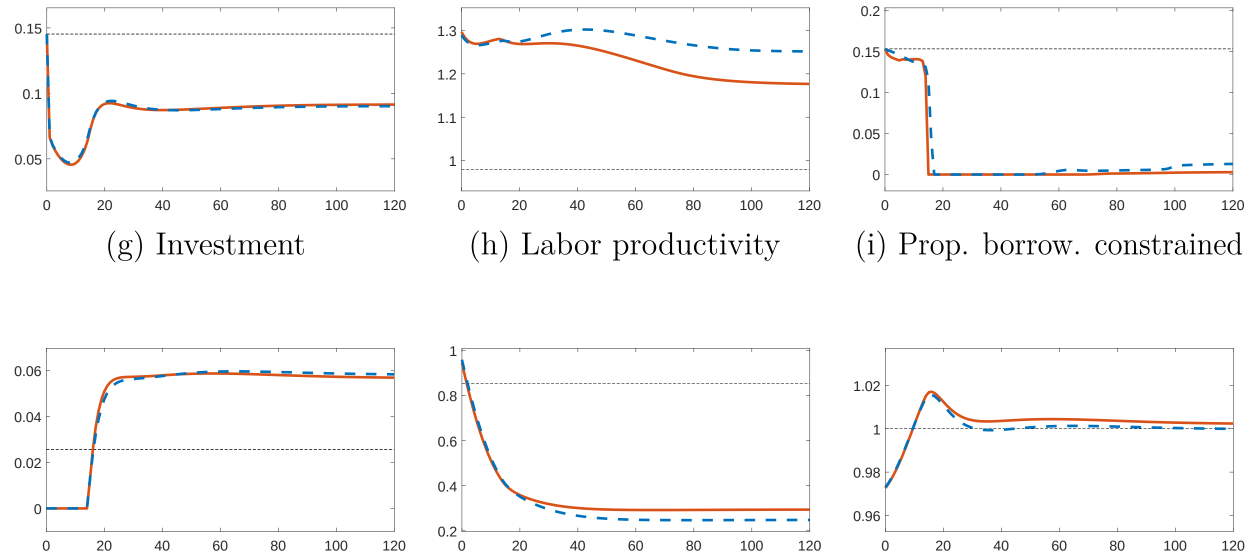


(a) Capital (b) Effective labor (N) (c) Output

(d) Consumption (e) Hours (H) (f) Avg. $\frac{c^{1-\sigma}}{1-\sigma} - \chi \frac{h^{1+1/\phi}}{1+1/\phi}$

Figure 72: Aggregates: DP-AHHW Method Comparison for Calibration from Acikgoz et al. (2018) (1)

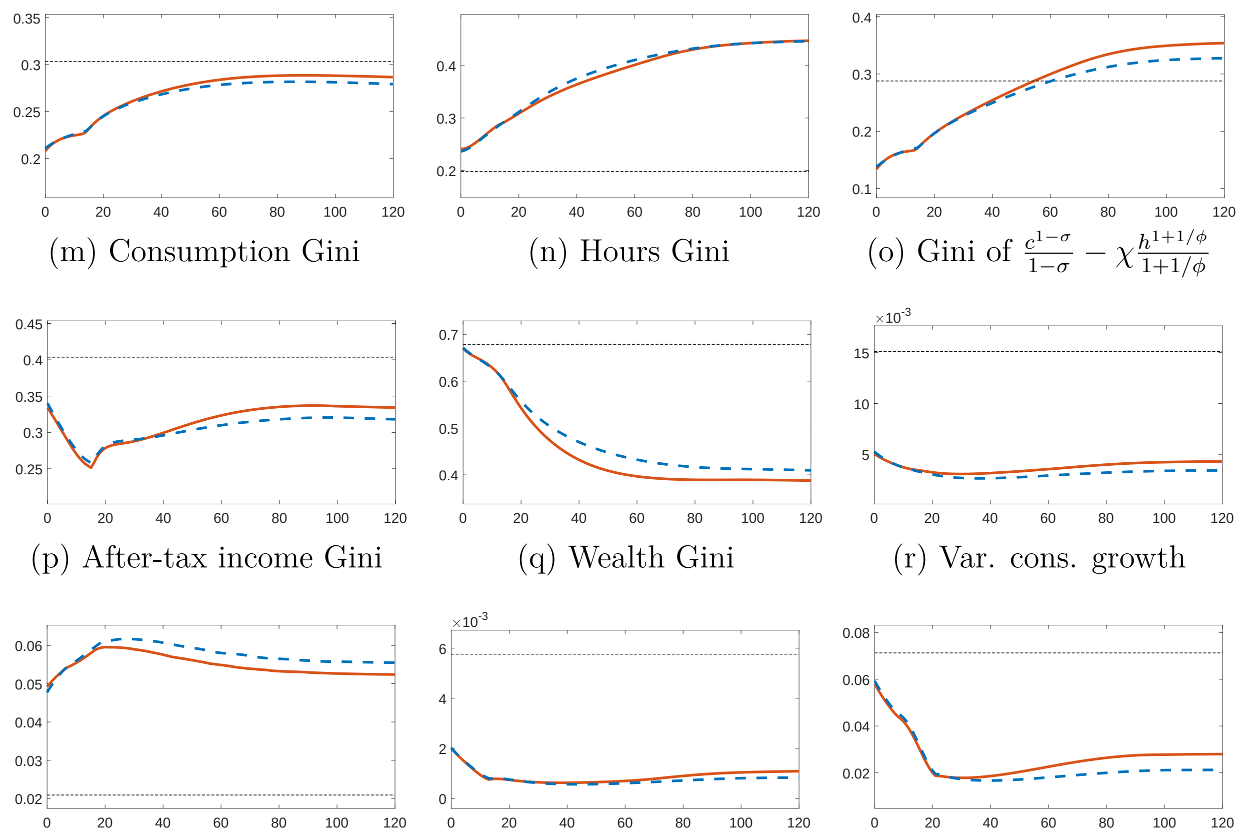
Notes: Red solid curve: optimal transition using our method; Blue dashed curve: optimal transition imposing the long-run policy obtained with the AHHW method; Thin dashed lines: corresponding values in initial stationary equilibrium.



(g) Investment (h) Labor productivity (i) Prop. borrow. constrained

(j) After-tax int. rates (k) After-Tax Wages (l) Modified golden rule

Figure 72: Aggregates: DP-AHHW Method Comparison for Calibration from Acikgoz et al. (2018) (2)



(m) Consumption Gini (n) Hours Gini (o) Gini of $\frac{c^{1-\sigma}}{1-\sigma} - \chi\frac{h^{1+1/\phi}}{1+1/\phi}$

(p) After-tax income Gini (q) Wealth Gini (r) Var. cons. growth

(s) Var. hours growth (t) Var. $\frac{c^{1-\sigma}}{1-\sigma} - \chi\frac{h^{1+1/\phi}}{1+1/\phi}$ growth (u) Var. a-t income growth

Figure 73: Inequality and Risk: DP-AHHW Method Comparison for Calibration from Acikgoz et al. (2018)

Notes: Red solid curve: optimal transition using our method; Blue dashed curve: optimal transition imposing the long-run policy obtained with the AHHW method; Thin dashed lines: corresponding values in initial stationary equilibrium.

References

Acikgoz, O., M. Hagedorn, H. Holter, and Y. Wang (2018): “The Optimum Quantity of Capital and Debt,” Tech. rep.

Aiyagari, S. R. and E. R. McGrattan (1998): “The optimum quantity of debt,” *Journal of Monetary Economics*, 42, 447–469.

Arnoud, A., F. Guvenen, and T. Kleineberg (2019): “Benchmarking Global Optimizers,” NBER Working Papers 26340, National Bureau of Economic Research, Inc.

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

Conesa, J. C., S. Kitao, and D. Krueger (2009): “Taxing Capital? Not a Bad Idea after All!” *American Economic Review*, 99, 25–48.

Dávila, J., J. H. Hong, P. Krusell, and J.-V. Ríos-Rull (2012): “Constrained Efficiency in the Neoclassical Growth Model With Uninsurable Idiosyncratic Shocks,” *Econometrica*, 80, 2431–2467.

Domeij, D. and J. Heathcote (2004): “On The Distributional Effects Of Reducing Capital Taxes,” *International Economic Review*, 45, 523–554.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020): “Heterogeneity and Persistence in Returns to Wealth,” *Econometrica*, 88, 115–170.

Fagereng, A., M. B. Holm, B. Moll, and G. Natvik (2019): “Saving Behavior Across the Wealth Distribution: The Importance of Capital Gains,” NBER Working Papers 26588, National Bureau of Economic Research, Inc.

Guvenen, F. (2011): “Macroeconomics with hetereogeneity : A practical guide,” *Economic Quarterly*, 97, 255–326.

Heathcote, J., F. Perri, and G. L. Violante (2010): “Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006,” *Review of Economic Dynamics*, 13, 15–51.

Hubmer, J., P. Krusell, and A. A. Smith (2020): “Sources of US Wealth Inequality: Past, Present, and Future,” in *NBER Macroeconomics Annual 2020, volume 35*, National Bureau of Economic Research, Inc, NBER Chapters.

Kan, A. R. and G. T. Timmer (1987a): “Stochastic Global Optimization Methods Part I: Clustering Methods,” *Mathematical Programming*, 39, 27–56.

——— (1987b): “Stochastic Global Optimization Methods Part II: Multilevel Methods,” *Mathematical Programming*, 39, 57–78.

KINDERMANN, F. AND D. KRUEGER (2021): "High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk," *American Economic Journal: Macroeconomics*, forthcoming.

KUCHERENKO, S. AND Y. SYTSKO (2005): "Application of Deterministic Low-Discrepancy Sequences in Global Optimization," *Computational Optimization and Applications*, 30, 297–318, 10.1007/s10589-005-4615-1.

KUHN, M. AND J.-V. RÍOS-RULL (2016): "2013 Update on the U.S. Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics," *Quarterly Review*, 1–75.

LOKEN, C., D. GRUNER, L. GROER, R. PELTIER, N. BUNN, M. CRAIG, T. HENRIQUES, J. DEMPSEY, C.-H. YU, J. CHEN, L. J. DURSI, J. CHONG, S. NORTHRUP, J. PINTO, N. KNECHT, AND R. V. ZON (2010): "SciNet: Lessons Learned from Building a Power-efficient Top-20 System and Data Centre," *Journal of Physics: Conference Series*, 256, 012026.

PONCE, M., R. VAN ZON, S. NORTHRUP, D. GRUNER, J. CHEN, F. ERTINAZ, A. FEDOSEEV, L. GROER, F. MAO, B. C. MUNDIM, M. NOLTA, J. PINTO, M. SALDARRIAGA, V. SLAVNIC, E. SPENCE, C.-H. YU, AND W. R. PELTIER (2019): "Deploying a Top-100 Supercomputer for Large Parallel Workloads: The Niagara Supercomputer," in *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)*, New York, NY, USA: Association for Computing Machinery, PEARC '19.

POWELL, M. (2009): "The BOBYQA algorithm for bound constrained optimization without derivatives," Tech. rep., Department of Applied Mathematics and Theoretical Physics, Cambridge University.

PRUITT, S. AND N. TURNER (2020): "Earnings Risk in the Household: Evidence from Millions of US Tax Returns," *American Economic Review: Insights*, 2, 237–54.

RÖHRS, S. AND C. WINTER (2017): "Reducing government debt in the presence of inequality," *Journal of Economic Dynamics and Control*, 82, 1–20.

RÍOS-RULL, J.-V. AND R. SANTAEULÀLIA-LLOPIS (2010): "Redistributive shocks and productivity shocks," *Journal of Monetary Economics*, 57, 931–948.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

STRAUB, L. AND I. WERNING (2020): "Positive Long-Run Capital Taxation: Chamley–Judd Revisited," *American Economic Review*, 110, 86–119.