A Heterogeneous-Agents RBC Model with Monopolistic Competition

Lilia Cavallari*

University of Rome III

and Ufficio Parlamentare Bilancio

Stefano D'Addona

University of Rome III

Paolo Porchia

CASMEF Institute, LUISS University

Abstract

We consider a flexible price continuous-time DSGE model, in a closed economy populated by heterogeneous households subject to uninsurable idiosyncratic income risk. Our main innovation is a rich representation of household consumption and saving behavior dictated by a utility function of the Increasing Elasticity of Substitution (IES) type, coupled with a new methodology to handle the endogenous dynamics of aggregate quantities implied by the presence of aggregate TFP shocks. We aim at studying how the interaction of aggregate shocks with income heterogeneity and IES preferences shapes the wealth distribution and creates interesting macroeconomic dynamics.

1 Introduction

The goal of this paper is to simultaneously unbundle two building blocks of traditional macroeconomic models: the representative agent and preferences. We introduce a general class of additive separable non-homothetic preferences over differentiated goods in an otherwise standard incomplete market (SIM) heterogeneous agents (HA) model, featuring both idiosyncratic (income) and aggregate (TFP) shocks, as in Krusell and Smith (1998). The combination of heterogeneity and non-homotheticity opens the way to a rich demand system, whereby optimal consumption (and saving) policies depend on novel mechanisms of intertemporal substitution varying across the population. We argue that demand plays an important role for wealth inequality and the propagation of shocks in the economy.

Homothetic preferences imply an income elasticity of demand equal to one, and a marginal propensity to consume (MPC) out of disposable income that is independent of the level of

*Email Addresses: lilia.cavallari@uniroma3.it, stefano.daddona@uniroma3.it, paolo.porchia@gmail.com

income. Over the years, many important insights have been facilitated by this assumption, including the idea - at the heart of the workhorse business cycle model - that the propagation of shocks is based on changes in relative prices that are independent of income. The assumption, however, stands in contrast to evidence based on consumers' behavior. In particular, we do not observe an income elasticity of demand equal to one in the data. In fact, the demand elasticity and the MPC are far from constant, and they both increase with the level of income (Attanasio and Browning (1995a)). Moreover, individuals experiencing income decreases have a higher propensity to consume compared to individuals experiencing income increases (Attanasio and Weber (2010)), while the rich have typically lower MPC compared to the poor. These facts suggest that different dimensions of heterogeneity are important for consumers' behavior. On the one hand, one needs to consider the specific conditions faced by each consumer, like his own income risk, in the attempt to explain why we observe different reactions to common shocks. On the other hand, individual demand itself varies in response to aggregate conditions, and these linkages may affect the general equilibrium of the economy. This paper sheds some light on both dimensions by introducing preferences with increasing elasticity of substitution (IES) over differentiated goods in a standard model à la Krusell and Smith (1998). To assess the role played by heterogeneity versus preferences, we compare the performance of our benchmark model with alternatives considering either homothetic preferences under idiosyncratic income risk or a representative agent with non-homothetic preferences.

The paper makes two contributions. First, it shows that a more general demand system is able to capture important features of the wealth distribution that are observed in the data, while providing an empirically plausible macroeconomic dynamics. Specifically, our benchmark model replicates the negligible share of wealth held at the bottom of the wealth distribution, its fat upper tail (though not as fat as in the data) and the counterciclicality of standard measures of wealth inequality, together with reasonable moments for aggregate output, consumption, investment and employment.

The benchmark model fares advantageously compared to the variant displaying homothetic preferences as regards both wealth and macroeconomic performance. The reason is a set of incentives brought about by endogenous variability in demand elasticity, which lead the consumption (and saving) profiles of less wealthier agents further apart from those of the rich. In order to see why, consider a cyclical downturn (a low TFP state, in our framework). In our imperfectly competitive economy, low demand elasticity and temporarily high markups imply a strong incentive to substitute current consumption with future consumption (and increase saving). The incentive, however, varies across the population. It is weaker for less wealthier agents and unemployed, who devote a larger share of their income to consumption (i.e., they have higher MPC) but have a lower elasticity of intertemporal substitution (EIS). The result is a more skewed distribution of wealth compared to the model with homothetic preferences. In addition, the skewness increases during cyclical downturns as is observed in the data.[1] The share of wealth held by the bottom 5 percent reduces while the share held by the top 5 percent increases, reflecting differences in the elasticity of asset demand between the

[1] Maestri and Roventini (2012) show that almost all inequality series in OECD countries are countercyclical at business cycle frequencies. The countercyclicality of income inequality is a stylized fact (e.g., Heathcote et al. (2010)).

rich and the poor. The variant with homothetic preferences, on the contrary, displays no heterogeneity in intertemporal substitution motives (the elasticity of intertemporal substitution is constant at unity for both the rich and the poor), and consumption smoothing is entirely driven by interest rate movements. Low interest rates in recessions curb the incentive to save in equal manner for the rich and the poor, so that their respective shares of wealth will reduce. The notion that less wealthier agents have a more inelastic asset demand compared to the wealthy resonates well with evidence showing that less wealthy nonstockholders have a lower elasticity of intertemporal substitution relative to wealthy stockholders (e.g., Guvenen (2006)).

Interestingly, the capacity to capture important aspects of the wealth distribution comes with further advantages in terms of macroeconomic dynamics. The benchmark outperforms both the alternative with homothetic preferences and the economy with a representative agent in replicating the volatility, ciclicality and persistence of consumption, output, employment and investments. The reason is that aggregate variables average the very distinct policies at the extremes of the wealth distribution, where most of the mass is concentrated. Consequently, consumption fluctuations reflect an average that puts more weight on the high MPC (though not as high as in the data) and the low EIS of the less wealthier agents, who contribute significantly to aggregate consumption. Capital and investments fluctuations, on the other hand, are mainly driven by wealthy agents (who own a large share of all the wealth in the economy), and reflect the high elasticity of intertemporal substitution of this group. In the model with homothetic preferences, on the contrary, the policies at the extremes of the distribution are less diverse, concern a smaller mass of individuals and the EIS and MPC are constant.

The second contribution is methodological. In any heterogeneous agent model featuring aggregate shocks, households should in principle keep track of the entire distribution of capital to plan their future consumption-savings. Practical considerations have then motivated researchers such as Krusell and Smith (1998) to propose notions of approximate rational expectations equilibria where agents form (and then verify) expectations about the behavior of selected moments of the distribution. A limitation of their approach is that agents do not factor the uncertainty of their forecasts in their decision process. We use an equilibrium a la Krusell and Smith (1998), where agents focus on (a restricted class of) dynamics of equilibrium prices, and explicitly consider forecast errors. We show that in our continuous-time framework a fixed point iteration to obtain equilibrium price dynamics is relatively easy to implement. Computing equilibria is thus easier than in discrete-time frameworks.

The paper speaks to different strands of a vast literature at the intersection of inequality, consumption theory and macroeconomics.[2] It provides a demand system characterized by a variable demand elasticity, in which both the intratemporal and the intertemporal substitution increase with the level of consumption. Several studies have considered non-homothetic preferences in the context of the SIM model for their ability to exhibit (heterogeneous) precautionary motives. The seminal contribution of De Nardi (2004), for instance, treats bequests as a luxury good (i.e., with lower elasticity compared to normal goods) to capture the notion that permanently richer agents save a larger fraction of their income.[3] Mankiw (1986)

[2]Recent surveys and books include, among others, Bertola et al. (2005); Heathcote et al. (2009); Guvenen (2011); Quadrini and Ríos-Rull (2015); Attanasio and Pistaferri (2016); Benhabib et al. (2019).

[3]Subsequent studies have considered non-homothetic bequests as well as non-homotheticity of consumption

exploits the variability of precautionary savings in the face of countercyclical earning risks to explain the equity premium puzzle (in this vein see, among others, Constantinides and Duffie (1996)). It is by now well-understood that non-homothetic preferences generate non-linear consumption policies and imply substantial deviations from the permanent income hypothesis. Our more modest contribution is to clarify the mechanisms by which these polices affect the wealth distribution and the business cycle in a context with imperfectly competitive markets. We show that markup variability has heterogeneous effects on agents' optimal policies depending on their wealth and employment status. The impact of this heterogeneity on the quantitative performance of the model is substantial. From a methodological standpoint Fernández-Villaverde et al. (2022) is related to our paper. They investigate the effect of financial frictions on the distribution of wealth in a continuous-time HA economy. In doing so, they implement a Krusell and Smith (1998) equilibrium where the endogenous state variables dynamics – as conjectured by agents – feature fully nonlinear and very general drifts. This hypothesis is more general than ours, yet more restrictive, in that it ignores the diffusion components of state variables dynamics.

The rest of the paper is organized as follows. Section 2 illustrates the model, while Sections 3 and 4 provide our solution approach, first in the steady-state, then in the full model. Section 5 reports on the empirical estimation exercise, and in Section 6 we discuss the propagation of shocks concerning individual policies and the wealth distribution, through the lens of comparative statics and impulse-response functions. Section 7 concludes.

2 Model

We consider a variant of the standard incomplete market model à la Krusell and Smith (1998) with imperfectly competitive goods markets and preferences exhibiting a time-varying elasticity of substitution across varieties (the elasticity of intratemporal substitution) and over time (the elasticity of intertemporal substitution). The demand structure draws on Cavallari (2022).

2.1 Intermediate Good Production

An intermediate good is produced by a perfectly competitive sector, which rents aggregate capital K_t at an instantaneous rental rate r_t, and employs aggregate labor L_t at an instantaneous wage w_t. The output rate (per unit of time) Y_t is modeled by a Cobb-Douglas production function with total factor productivity $\exp(a_t)$:

$$Y_t = \exp(a_t) K_t^{\alpha} L_t^{1-\alpha} \tag{1}$$

over an agent's life time Straub (2019).

where α is the constant capital share. The random evolution of the log-TFP a_t is described by a continuous-time Markov Chain with n_a states and $n_a \times n_a$ constant intensity matrix Λ^a:

$$\Lambda^a = \begin{bmatrix} -\sum_{i\neq 1}\lambda^a_{1i} & \lambda^a_{12} & \dots & \lambda^a_{1n_a} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda^a_{n_a 1} & \lambda^a_{n_a 2} & \dots & -\sum_{i\neq n_a}\lambda^a_{n_a i} \end{bmatrix} \tag{2}$$

with λ^a_{ij} denoting the instantaneous intensity of a transition from state i to state j. This choice, while computationally convenient, entails little loss of generality relative to the common assumption of a mean-reverting process such as the Ornstein-Uhlenbeck's.[4]

Markets for production factors are perfectly competitive, therefore factor rewards are pinned down by their marginal productivity:

$$w_t = (1-\alpha)e^{a_t}(K_t/L_t)^{\alpha} \tag{3}$$

$$r_t = \alpha e^{a_t}(K_t/L_t)^{\alpha-1} \tag{4}$$

As detailed below, the intermediate good can be consumed or invested by the households in a capital accumulation process, or used as production factor in a linear technology, to produce a variety of final consumption goods.

2.2 Households

The economy is populated by a continuum of households, whose measure is normalized to one.[5] Households face idiosyncratic, uninsurable labor productivity and time-preference shocks. The former is modeled by means of a binary continuous-time Markov chain ε_t, taking values $\varepsilon_t = 1$ (employed) and $\varepsilon_t = 0$ (unemployed). The constant 2×2 intensity matrix of this process is Λ^{ε}, with $\lambda^{\varepsilon}_{ij}$ denoting transition intensities. To model time-preference shocks, we allow the subjective discount rate β_t to follow a continuous-time Markov chain with three states, with constant 3×3 intensity matrix Λ^{β} and transition intensities λ^{β}_{ij}. Similarly to Krusell and Smith (1998), the purpose of time-preference shocks is to match the features of the empirical wealth distribution.

Households have homogeneous preferences over labor services l_t and a (measure one) continuum of final goods. In particular, $u(c_{j,t})$ is the subutility of consumption $c_{j,t}$ of the final good j, while we assume a logarithmic felicity function acting on a linear aggregator of subutilities.

Households can self-insure against idiosyncratic preference and labor productivity shocks by investing their savings in a capital accumulation technology. Each household is uniquely identified by capital (or wealth) holding k_t, its employment state ε_t and time-preference state β_t. Letting $p_{j,t}$ denote the price of a generic final good j, and $\Gamma_t = (r_t, w_t, p_{j,t})$ the vector of

[4]Indeed, in what follows we calibrate the states and the intensity matrix of our Markov Chain to match the properties of the steady state distribution of the Ornstein–Uhlenbeck process

$$da_t = -\eta_a a_t\, dt + \sigma_a dZ_t$$

fitted to US TFP data.

[5]Henceforth we drop the household index where no confusion may arise.

endogenous state-variables,[6] agents solve the following consumption-savings program:

$$V(k_t, \varepsilon_t, \beta_t, \Gamma_t) = \sup_{l_t \geq 0, c_{j,t}} \mathbb{E}\left[\int_t^\infty e^{-\int_t^s \beta_u\, du}\left(\log\left(\int_0^1 u(c_{j,s})\, dj\right) - v\,\frac{l_s^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}\right) ds\right] \tag{5}$$

subject to the budget constraint:

$$dk_t = \left[k_t(r_t - \delta) + w_t l_t \varepsilon_t - \int_0^1 p_{j,t} c_{j,t}\, dj + \Pi_t\right] dt \tag{6}$$

$$k_t \geq \underline{k} \tag{7}$$

In expression (5) φ is the Frisch elasticity of labor supply and v a scale parameter on the labor supply disutility. In (6), δ denotes the capital depreciation rate, and Π_t is the fraction of final goods sector's profits (to be detailed below) distributed to the individual households. The problem is state-constrained, in that we impose a lower capital bound $\underline{k}$. Households solve their program taking the equilibrium dynamics of Γ_t and Π_t as given, and face the net-worth constraint (7). The optimal decision rules for the consumption rate $c_j(k_t, \varepsilon_t, \beta_t, \Gamma_t)$ and labor supply $l(k_t, 1, \beta_t, \Gamma_t)$ imply an optimal drift for the capital process (6), which we denote by $s(k_t, \varepsilon_t, \beta_t, \Gamma_t)$:

$$s(k_t, \varepsilon_t, \beta_t, \Gamma_t) = k_t(r_t - \delta) + w_t l(k_t, \varepsilon_t, \beta_t, \Gamma_t)\varepsilon_t - \int_0^1 p_{j,t} c_j(k_t, \varepsilon_t, \beta_t, \Gamma_t)\, dj + \Pi_t \tag{8}$$

2.3 Final Good Production

Each final good variety is produced by a monopolistically competitive firm using a linear technology, with a unitary marginal cost. At each time instant, this firm aggregates the demand for good j across households and solves the intratemporal profit maximization problem:

$$\max_{p_{j,t}} \Pi_{j,t} := \left(p_{j,t} - 1\right) \int_0^1 c^i_{j,t}(p_{j,t})\, di \tag{9}$$

where the index i denotes the individual household. Notice that we have emphasized the dependence of the optimal demand on price. As we show in the next sections, households' optimization implies the following first order condition:

$$\frac{u'(c^i_{j,t})}{\int_0^1 u(c^i_{u,t})\, du} = \lambda^i_t p_{j,t} \tag{10}$$

with λ_t denoting a costate variable. This implies that

$$c^i(p_{j,t}) = h\left(p_{j,t}\lambda^i_t \int_0^1 u(c^i_{j,t})\, dj\right), \qquad h(\cdot) = (u')^{-1}(\cdot), \tag{11}$$

[6] Γ_t is in principle infinite dimensional, as it includes the continuum of final good prices. Since our equilibrium will be symmetric over final goods, the abuse of notation has little consequence.

where the function $h(\cdot)$ is the inverse marginal subutility. Using this expression, the first order condition of the price setting problem (9) reads:

$$\int c^i(p_{j,t})\,di + (p_{j,t}-1)\int \frac{\lambda_t^i \int_0^1 u(c_{u,t}^i)\,du}{u''(c_{s,t}^i)}\,di = 0$$

After using (10) to set $\lambda_t^i \int_0^1 u(c_{u,t}^i)\,du = u'(c_{j,t}^i)/p_{j,t}$, we obtain the optimal price:

$$p_{j,t}^* = \frac{1}{1-\vartheta_t}, \qquad \vartheta_t = -\frac{\int c_{j,t}^i\,di}{\displaystyle\int \frac{u'(c_{j,t}^i)}{u''(c_{j,t}^i)}\,di} \tag{12}$$

The optimal mark-up $m_t = \vartheta_t/(1-\vartheta_t)$ is reminiscent of the expressions typically derived in a representative agent context, however with household heterogeneity the expression $1/\vartheta_t$ is interpreted as an aggregate elasticity of intratemporal substitution, as we clarify in the next section.

2.4 IES preferences

The traditional choice in the macroeconomics literature for the subutility index $u(\cdot)$ is a power form such as $u(c) = \frac{c^{1-1/\theta}}{1-1/\theta}$, with $\theta > 1$. This delivers "log-CES" homothetic preferences where θ can be interpreted as the intratemporal elasticity of substitution between goods while the intertemporal elasticity is unitary (due to the logarithmic transformation of the consumption index). With this choice, optimal prices and markups in expression (12) become constant, since $\vartheta_t = 1/\theta$.

Instead, we opt for the following specification of the subutility:

$$u(c) = \gamma c + \frac{c^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}}, \qquad \theta > 1, \quad \gamma > 0 \tag{13}$$

This is the *IES* (or Increasing Elasticity of Substitution) utility specification introduced in Bertoletti and Etro (2016). As its name suggests, it implies an *intratemporal* elasticity of substitution between varieties which is increasing in consumption, at the single agent level:

$$-\frac{\partial \ln\left(c_i/c_j\right)}{\partial \ln p_i} = -\left(\frac{\partial \ln\left(p_i/p_j\right)}{\partial \ln c_i}\right)^{-1} = -\frac{u'(c)}{u''(c)c} = \theta\left(1+\gamma c^{\frac{1}{\theta}}\right), \tag{14}$$

where we have evaluated the expression under the symmetry hypothesis ($c_i = c_j = c$), which will be the case of interest. The *intertemporal* elasticity of substitution for a single agent can be defined as:

$$-\frac{\partial \ln(c)}{\partial \ln p} = \frac{1}{-\frac{u''(c)c}{u'(c)} + \frac{u'(c)c}{u(c)}} \tag{15}$$

Substituting this utility specification in the optimal pricing rule of expression (12), we obtain:

$$\vartheta_t = \frac{\int c_t^i \, di}{\int \theta c_t^i \left(1 + \gamma (c_t^i)^{\frac{1}{\theta}}\right) di} \tag{16}$$

The assumptions $\theta > 1$ and $\gamma > 0$ guarantee that the markup m_t is uniformly positive (i.e. $p_t^* > 1$). Importantly, with $\gamma > 0$, it is also easy to show that the markup is countercyclical, in that a reduction in consumption across the population increases it. By analogy with the single-agent case discussed in Cavallari and Etro (2020), we define an 'aggregate' elasticity of *intratemporal* substitution as $\epsilon_t = 1/\vartheta_t$, and an 'aggregate' elasticity of *intertemporal* substitution as

$$\chi_t = \frac{1}{\epsilon_t + \iota_t}, \qquad \iota_t = \frac{\int c_t^i \, di}{\int \frac{u(c_t^i)}{u'(c_t^i)} \, di} \tag{17}$$

While an increase in consumption across the population always augments ϵ_t, provided $\theta > 1$ and $\gamma > 0$, the effect on χ_t is ambiguous from a theoretical standpoint. In our empirical application, though, χ_t will turn out to be procyclical, with important ramifications for the properties of the wealth distribution in the economy.

2.5 Equilibrium

Let $G_t(k, \varepsilon, \beta; \Gamma_t)$ denote the joint distribution of the household population over capital, labor productivity and time-preference states, and let $g_t(k, \varepsilon, \beta; \Gamma_t)$ denote its density in the capital dimension. A symmetric (relative to consumption varieties) equilibrium in this economy is defined as a path for individual household decisions $\{k_t, c_t, l_t\}_{t \geq 0}$, input and consumption-good prices $\{r_t, w_t, p_t^*\}_{t \geq 0}$ and densities $\{g_t\}_{t \geq 0}$ such that, at every t:

1. Households decisions are optimal policies of the consumption-saving problem (5), taking as given equilibrium prices $\Gamma_t = (r_t, w_t, p_t^*)$

2. The sequence of densities satisfies aggregate consistency conditions

3. Equilibrium prices $\Gamma_t = (r_t, w_t, p_t^*)$ satisfy (4), (3), and

$$p_t^* = \frac{1}{1 - \vartheta_t}, \qquad \vartheta_t = -\frac{\sum_{\varepsilon=0,1} \sum_{\beta=1,2,3} \int c(k, \varepsilon, \beta, \Gamma_t) g_t(k, \varepsilon, \beta; \Gamma_t) \, dk}{\sum_{\varepsilon=0,1} \sum_{\beta=1,2,3} \int \frac{u'(c(k, \varepsilon, \beta, \Gamma_t))}{u''(c(k, \varepsilon, \beta, \Gamma_t))} g_t(k, \varepsilon, \beta; \Gamma_t) \, dk}, \tag{18}$$

   respectively, where $u(\cdot)$ is given by (13).

4. Markets for capital and labor clear:

$$K_t = \sum_{\varepsilon=0,1} \sum_{\beta=1,2,3} \int k \, g_t(k, \varepsilon, \beta; \Gamma_t) \, dk \tag{19}$$

$$L_t = \sum_{\beta=1,2,3} \int l(k, 1, \beta, \Gamma_t) \, g_t(k, 1, \beta; \Gamma_t) \, dk \tag{20}$$

Notice that the last condition implies clearing of the goods market, since aggregating the budget constraint (6) across households we obtain:

$$\int \frac{dk_t}{dt} \, dG_t = (r_t - \delta)K_t + w_t L_t - p_t^* \int c \, dG_t + \int \Pi_t \, dG_t \tag{21}$$

where we have dropped functional arguments for ease of notation. Defining aggregate investment (including depreciation) $I_t = \int \frac{dk_t}{dt} \, dG_t + \delta K_t$, aggregate consumption $C_t = \int c \, dG_t$, using factor prices (4) and (3), and noting that aggregate profits of the final goods sector are $\int \Pi_t \, dG_t = (p_t^* - 1)C_t$, we obtain the goods market-clearing condition:

$$Y_t = C_t + I_t \tag{22}$$

3 Steady State Solution Approach

We first illustrate the solution of our model in the steady-state, where idiosyncratic labor productivity and time-preference shock are still present at the household level, but aggregate TFP shocks are not. Therefore, the output rate becomes:

$$\overline{Y} = \overline{K}^{\alpha} \, \overline{L}^{1-\alpha}$$

where an overbar denotes constant steady-state values. $\overline{G}(k, \varepsilon, \beta)$ is the joint steady state distribution of capital, labor productivity and time-preference states, with a density $\overline{g}(k, \varepsilon, \beta)$ in the capital dimension. The following equilibrium conditions then hold:

$$\overline{K} = \sum_{i=1,2} \sum_{j=1,2,3} \int k \, \overline{g}(k, \varepsilon_i, \beta_j) \, dk \tag{23}$$

$$\overline{L} = \sum_{j=1,2,3} \int l(k, 1, \beta_j) \, \overline{g}(k, 1, \beta_j) \, dk \tag{24}$$

$$\overline{C} = \sum_{i=1,2} \sum_{j=1,2,3} \int c(k, \varepsilon_i, \beta_j) \, \overline{g}(k, \varepsilon_i, \beta_j) \, dk \tag{25}$$

$$\overline{w} = (1-\alpha)(\overline{K}/\overline{L})^{\alpha}, \qquad \overline{r} = \alpha(\overline{K}/\overline{L})^{\alpha-1} \tag{26}$$

$$\overline{p}^* = \frac{1}{1-\overline{\vartheta}}, \qquad \overline{\vartheta} = -\frac{\overline{C}}{\displaystyle\sum_{i=1,2} \sum_{j=1,2,3} \int \frac{u'(c(k, \varepsilon_i, \beta_j))}{u''(c(k, \varepsilon_i, \beta_j))} \overline{g}(k, \varepsilon_i, \beta_j) \, dk}, \tag{27}$$

where $c(\cdot)$ and $l(\cdot)$ denote the optimal individual consumption and labor policies derived below.

3.1 Households' Problem and Steady-State Distribution

Taking as given production factor and consumption good steady-state prices $(\overline{r}, \overline{w}, \overline{p}^*)$, and assuming a vanishing individual share of redistributed profits,[7] $\Pi \approx 0$, households solve

[7] This is due to our assumption of a diffuse measure of households.

their individual consumption-investment problem (5) subject to the continuous-time Markov Chain dynamics of the labor productivity and time-preference states.

The problem could in principle be solved with the martingale method developed in He and Pagès (1993), whose dual formulation leads to a linear variational equality to be solved. We sketch this approach in the Appendix C.1, for completeness, as it would lead to a very explicit characterization if the subutility function was of the CES type. In our IES case, though, the inability to solve explicitly for the optimal consumption rule (as a function of the state-price density and the dynamic Lagrange multiplier) undermines the advantages of this method. Hence, we directly tackle the dynamic programming formulation of the problem.

Let $V(k,\varepsilon_i,\beta_j)$ denote the value function of the dynamic program (5), in labor productivity state $i \in \{1,2\}$ and time-preference state $j \in \{1,2,3\}$. It satisfies the following Hamilton-Bellman-Jacobi equation

$$\begin{aligned}\beta_j V(k,\varepsilon_i,\beta_j) = \sup_{c_t,l_t\geq 0}&\left[\log u(c_t) - v\frac{l_t^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}} + V_k(k,\varepsilon_i,\beta_j)\left((\bar{r}-\delta)k + \overline{w}l_t\varepsilon_i - \overline{p}^* c_t\right)\right] \\ &+ \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2}\left(V(k,\varepsilon_{i_2},\beta_j) - V(k,\varepsilon_i,\beta_j)\right) + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2}\left(V(k,\varepsilon_i,\beta_{j_2}) - V(k,\varepsilon_i,\beta_j)\right) \quad (28)\end{aligned}$$

where $u(\cdot)$ has the IES form (13). The First Order Conditions read:

$$\frac{u'(c_t)}{u(c_t)} = V_k(k,\varepsilon_i,\beta_j)\overline{p}^*, \qquad l_t = \begin{cases} 0 & \text{if } \varepsilon_i = 0 \\ \left(\frac{\overline{w}V_k(k,\varepsilon_i,\beta_j)}{v}\right)^{\varphi} & \text{if } \varepsilon_i = 1\end{cases} \quad (29)$$

Let $c_t^* = f(V_k(k,\varepsilon_i,\beta_j))$ and $l_t^* = h(V_k(k,\varepsilon_i,\beta_j))$ denote the optimal policies as derived (numerically, in the consumption's case) from these FOCs. The value functions solve the following nonlinear system of ODEs:

$$\begin{aligned}\left(\beta_j + \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2} + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2}\right)V(k,\varepsilon_i,\beta_j) &= \log u(c_t^*) - v\frac{(l_t^*)^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}} + V_k(k,\varepsilon_i,\beta_j)s(k,\varepsilon_i,\beta_j) \\ &+ \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2}V(k,\varepsilon_{i_2},\beta_j) + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2}V(k,\varepsilon_i,\beta_{j_2}) \\ &\qquad i=1,2 \quad j=1,2,3. \quad (30)\end{aligned}$$

where we have used the optimal saving policy:

$$s(k,\varepsilon_i,\beta_j) = (\bar{r}-\delta)k + \overline{w}l_t^*\varepsilon_i - \overline{p}^* c_t^* \quad (31)$$

The state constraint $k \geq \underline{k}$ translates into $s(\underline{k},\varepsilon_i,\beta_j) \geq 0$, which amounts to a nonlinear constraint on $V_k(\underline{k},\varepsilon_i,\beta_j)$. The numerical solution of this system of ODEs, hence of the optimal policies $c(k,\varepsilon_i,\beta_j)$, $l(k,1,\beta_j)$ and $s(k,\varepsilon_i,\beta_j)$, is discussed in Appendix A.1.

The joint steady-state density of households over capital, labor productivity and time-preference satisfies the following Forward Kolmogorov equation (see Achdou, Han, Lasry, Lions, and Moll (2017)):

$$-\frac{d}{dk}\left[g(k,\varepsilon_i,\beta_j)s(k,\varepsilon_i,\beta_j)\right]+\sum_{i_2\neq i}\lambda^{\varepsilon}_{i_2 i}g(k,\varepsilon_{i_2},\beta_j)+\sum_{j_2\neq j}\lambda^{\beta}_{j_2 j}g(k,\varepsilon_i,\beta_{j_2})$$
$$-\left(\sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2}+\sum_{j_2\neq j}\lambda^{\beta}_{jj_2}\right)g(k,\varepsilon_i,\beta_j)=0 \quad (32)$$

where the equation depends on the optimal saving policy (31) derived previously. Again, we obtain a numerical solution using a finite-difference method, which reduces the ODE (32) to an eigenvalue problem.

3.2 Equilibrium Algorithm

In order to obtain an equilibrium, the ex-ante values of factor and consumption good-prices taken as given (i.e. conjectured) by the households upon solving their consumption-saving problem must coincide with their ex-post counterparts stemming from the market-clearing conditions (26)-(27). To this end, define a mapping $(r_2, w_2, p_2^*) = F(r_1, w_1, p_1^*)$ as follows:

(i) Solve the HBJ equation (30) using candidate steady-state prices (r_1, w_1, p_1^*), and obtain optimal policies $c(k,\varepsilon_i,\beta_j)$, $l(k,1,\beta_j)$ and $s(k,\varepsilon_i,\beta_j)$.

(ii) Using the latter, find the candidate steady-state distribution $g(k,\varepsilon_i,\beta_j)$ as solution of the Forward Kolmogorov equation (32).

(iii) Compute aggregate capital, labor and consumption using conditions (23)-(25), and obtain ex-post steady-state values of prices, (r_2, w_2, p_2^*), with conditions (26)-(27).

A steady-state equilibrium consists of a fixed point of the mapping $F(\cdot)$, that is:[8]

$$(\bar{r},\bar{w},\bar{p}^*) = F(\bar{r},\bar{w},\bar{p}^*) \quad (33)$$

4 Full Model Solution Approach

Let us now consider the problem of solving for equilibrium when aggregate TFP shocks are present, so that the output rate is given by expression (1).

In equilibrium, factor and consumption-good prices (r_t, w_t, p_t^*) are now stochastic process, the dynamics of which depend on the evolution of the whole distribution $G_t(k,\varepsilon,\beta;\Gamma_t)$, through the market-clearing conditions (4), (3), and (18). From the household standpoint, the implication is that the distribution becomes a state variable for the individual consumption-saving problem. The challenge of tracking the evolution of an infinite-dimensional object has

[8]In practice, a nonlinear equation solver can be employed to solve this problem, though the Jacobian of the mapping needs to be evaluated numerically by finite difference. Alternatively, the dumped iteration

$$(r_{t+1}, w_{t+1}, p_{t+1}^*) = (1-\omega)\,(r_t, w_t, p_t^*) + \omega\, F(r_t, w_t, p_t^*)$$

converges robustly to a fixed point, provided the parameter ω is small enough.

motivated Krusell and Smith (1998) to propose a notion of equilibrium with 'bounded rationality', whereby agents consider only a few selected moments (typically the first) of the distribution as state variables, and conjecture a dynamic equation for their evolution in time, which is verified ex-post through an iterative simulation scheme.

We follow the same idea of selecting a particular, boundedly rational equilibrium, by assuming that agents conjecture dynamics of few specific variables which depend on collective behavior. Prices – $\Gamma_t = (r_t, w_t, p_t^*)$ – seem the natural candidates. Because of market clearing, this is equivalent to conjecturing the behavior of (functions of) the aggregate distribution of capital. Following Kaplan et al. (2020) we restrict agents' conjectures about equilibrium prices to deterministic functions of the log-TFP state a, which is then the only additional state variable households need to track in their individual plans. Being these functions independent of time, we are implicitly assuming that the economy has reached its stochastic steady state. The finite dimension of the unknown price functions renders an adaption of Krusell and Smith (1998)'s methodology quite straightforward, as we outline next. For candidate price processes $(r_1(a_t), w_1(a_t), p_1^*(a_t))$, define a mapping $(r_2(a_t), w_2(a_t), p_2^*(a_t)) = F(r_1(a_t), w_1(a_t), p_1^*(a_t))$ as follows:

(i) Solve the household's consumption-saving problem (5) subject to the candidate price processes $(r_1(a_t), w_1(a_t), p_1^*(a_t))$. Note that these are just finite-dimensional vectors, the entries of which are values taken in correspondence of each log-TFP state. Extending the argument of the previous section, the household's value function solves the HBJ equation:

$$\left(\beta_j + \sum_{i_2 \neq i} \lambda^{\varepsilon}_{ii_2} + \sum_{j_2 \neq j} \lambda^{\beta}_{jj_2} + \sum_{z_2 \neq z} \lambda^{a}_{zz_2}\right) V(k, \varepsilon_i, \beta_j, a_z) = \log u(c_t^*) - v\frac{(l_t^*)^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}$$
$$+ V_k(k, \varepsilon_i, \beta_j, a_z) s(k, \varepsilon_i, \beta_j, a_z) + \sum_{i_2 \neq i} \lambda^{\varepsilon}_{ii_2} V(k, \varepsilon_{i_2}, \beta_j, a_z) + \sum_{j_2 \neq j} \lambda^{\beta}_{jj_2} V(k, \varepsilon_i, \beta_{j_2}, a_z)$$
$$+ \sum_{z_2 \neq z} \lambda^{a}_{zz_2} V(k, \varepsilon_i, \beta_j, a_{z_2}) \quad (34)$$

where we have used the optimal saving policy:

$$s(k, \varepsilon_i, \beta_j, a_z) = (r_1(a_z) - \delta)k + w_1(a_z) l_t^* \varepsilon_i - p_1^*(a_z) c_t^* \quad (35)$$

Moreover, $c_t^* = f(V_k(k, \varepsilon_i, \beta_j, a_z))$ and $l_t^* = h(V_k(k, \varepsilon_i, \beta_j, a_z))$ denote the optimal policies derived (numerically, in the consumption's case) from the first order conditions

$$\frac{u'(c_t)}{u(c_t)} = V_k(k, \varepsilon_i, \beta_j, a_z) p_1^*(a_z), \qquad l_t = \begin{cases} 0 & \text{if } \varepsilon_i = 0 \\ \left(\frac{w_1(a_z) V_k(k, \varepsilon_i, \beta_j, a_z)}{v}\right)^{\varphi} & \text{if } \varepsilon_i = 1 \end{cases} \quad (36)$$

with $u(\cdot)$ as in (13). The state constraint $k \geq \underline{k}$ translates into $s(\underline{k}, \varepsilon_i, \beta_j, a_z) \geq 0$, which amounts to a nonlinear constraint on $V_k(\underline{k}, \varepsilon_i, \beta_j, a_z)$. The numerical method used to obtain a solution of this system of ODEs, hence of the optimal policies $c(k, \varepsilon_i, \beta_j, a_z)$, $l(k, 1, \beta_j, a_z)$ and $s(k, \varepsilon_i, \beta_j, a_z)$, is described in Appendix A.1.

ii) Let $g_t(k,\varepsilon,\beta)$ denote the joint density of capital, labor productivity and time-preference at time t, for a given initial condition $g_0(k,\varepsilon,\beta)$. Its evolution in time reads:

$$\begin{aligned} d g_t(k,\varepsilon_i,\beta_j) = \Bigg[-\frac{d}{dk}\left[g_t(k,\varepsilon_i,\beta_j) s(k,\varepsilon_i,\beta_j,a_t)\right] + \sum_{i_2\neq i} \lambda^{\varepsilon}_{i_2 i} g_t(k,\varepsilon_{i_2},\beta_j) \\ + \sum_{j_2\neq j} \lambda^{\beta}_{j_2 j} g_t(k,\varepsilon_i,\beta_{j_2}) - \left(\sum_{i_2\neq i} \lambda^{\varepsilon}_{i i_2} + \sum_{j_2\neq j} \lambda^{\beta}_{j j_2}\right) g_t(k,\varepsilon_i,\beta_j) \Bigg] dt \\ i=1,2 \quad j=1,2,3 \quad z=1,\dots,n_a \end{aligned} \tag{37}$$

A trajectory of the density function implicitly depends on the specific history of log-TFP until time t, through the optimal savings function $s(k,\varepsilon_i,\beta_j,a_t)$. However in the long-run, stationarity implies that an invariant value of the conditional density $g(k,\varepsilon_i,\beta_j;a_z)$ prevails, for $z=1,\dots,n_z$. To compute the latter, we simulate a very long trajectory $\tilde{a}_t$ over a fine discretization of the time interval, and obtain the corresponding trajectory of density functions $\tilde{g}_t(k,\varepsilon_i,\beta_j)$ by iterating on (37) with an Euler scheme.[9] After discarding an initial burn-in period, we obtain an estimate of the conditional density $g(k,\varepsilon_i,\beta_j;a_z)$ by averaging simulations $\tilde{g}_t(k,\varepsilon_i,\beta_j)$ for all dates where a realization a_z took place.

(iii) Using the distribution just obtained, we compute state-wise aggregate capital, labor, consumption, and new candidate prices $(r_2(a_t), w_2(a_t), p_2^*(a_t))$:

$$K(a_z) = \sum_{i=1,2}\sum_{j=1,2,3} \int k\, g(k,\varepsilon_i,\beta_j;a_z)\, dk \tag{38}$$

$$L(a_z) = \sum_{j=1,2,3} \int l(k,1,\beta_j,a_z)\, g(k,1,\beta_j;a_z)\, dk \tag{39}$$

$$C(a_z) = \sum_{i=1,2}\sum_{j=1,2,3} \int c(k,\varepsilon_i,\beta_j,a_z)\, g(k,\varepsilon_i,\beta_j;a_z)\, dk \tag{40}$$

$$w_2(a_z) = (1-\alpha)(K(a_z)/L(a_z))^{\alpha}, \qquad r_2(a_z) = \alpha (K(a_z)/L(a_z))^{\alpha-1} \tag{41}$$

$$p_2^*(a_z) = \frac{1}{1-\vartheta(a_z)}, \qquad \vartheta(a_z) = -\frac{C(a_z)}{\sum_{i=1,2}\sum_{j=1,2,3}\int \frac{u'(c(k,\varepsilon_i,\beta_j,a_z))}{u''(c(k,\varepsilon_i,\beta_j,a_z))} g(k,\varepsilon_i,\beta_j;a_z)\, dk},$$

$$z=1,\dots,n_a$$

An equilibrium consists of a fixed point of the mapping $F(\cdot)$, that is:[10]

$$(r(a_z), w(a_z), p^*(a_z)) = F(r(a_z), w(a_z), p^*(a_z)) \tag{42}$$

[9] Using an implicit scheme does not change results substantially.
[10] In practice, we use the dumped iteration

$$(r_{t+1}(a_z), w_{t+1}(a_z), p^*_{t+1}(a_z)) = (1-\omega)\,(r_t(a_z), w_t(a_z), p^*_t(a_z)) + \omega\, F(r_t(a_z), w_t(a_z), p^*_t(a_z))$$

which converges robustly to a fixed point, provided the parameter ω is small enough, rather than a nonlinear equation solver, since the Jacobian of $F(\cdot)$ needs to be computed by finite difference, which can be a daunting task, depending on the number of log-TFP states n_a.

5 Estimation

In order to obtain reasonable parameter values for preferences, technology and the dynamics of state variables, we rely on a combination of calibration and econometric estimation. We refer to Appendix A.2 for a more detailed description of the procedure.

Borrowing from the mainstream literature, we set the capital share to $\alpha = 0.33$ and the capital depreciation rate to $\delta = 0.025$, to match the 10% rate of capital depletion per year found in US data. We normalize ν, the scale parameter for the disutility of labor, so that the (deterministic) steady-state value of employment is 1.[11] We calibrate the intensity matrix Λ^{ε} of the idiosyncratic employment state in order to match the discrete-time annual equivalent used in Krusell and Smith (1998), while we dogmatically set the 3 possible values of the idiosyncratic preference rate to account for reasonable degrees of patience: $\beta_1 = -\log 0.975$, $\beta_2 = -\log 0.988$ and $\beta_3 = -\log 0.999$. We also restrict the intensity matrix of the latter variable, so that no transitions are possible between extreme states: $\lambda^{\beta}_{13} = \lambda^{\beta}_{31} = 0$.

We work with $n_a = 9$ possible states for log-TFP a_t. We adopt the following strategy to identify the transition matrix Λ^a: we assume that the embedded discrete-time Markov chain of a_t approximates at yearly frequency the AR-1 process

$$a_t = \rho_a \, a_{t-1} + \sigma_a \, u_t \tag{43}$$

where u_t is a standard Gaussian innovation. Given parameters ρ_a and σ_a,[12] the Rouwenhorst (1995) discretization procedure identifies the states a_i, $i = 1, \ldots, n_a$, and the yearly transition probability matrix of the approximating Markov chain. The continuous-time intensities in Λ^a are chosen to match the latter.

We estimate the set of free parameters $\Omega = (\lambda^{\beta}_{12}, \lambda^{\beta}_{21}, \lambda^{\beta}_{23}, \lambda^{\beta}_{32}, \rho_a, \sigma_a, \varphi, \gamma, \theta)$ with the Simulated Method of Moments of Duffie and Singleton (1993) and Lee and Ingram (1991). We consider a set of 11 moments of aggregate consumption, output, investment, and labor supply: in particular, we match quarterly variances and one-quarter autocovariances of each variable, and contemporaneous covariance of output with the other three. Real moments are computed on quarterly time series from 1947 to 2017 on real GDP per capita, real non-durable and services consumption expenditure per capita, real total private fixed investment (including durable consumption) per capita, and total hours in the non-farm business sector. All series are from the US Department of Commerce, Bureau of Economic Analysis (BEA) or the US Bureau of Labor Statistics (BLS). We augment this collection with two additional moments of the wealth distribution extracted from the US *Survey of Consumer Finances*: the fraction of total wealth owned by the wealthiest 5% and the poorest 5% of the population.

We now turn to estimation results, focusing on the baseline IES specification given in (13). Table 1 reports the SMM parameter estimates, and the corresponding t-statistics. Point estimates of preference parameters are $\gamma = 0.9$ and $\theta = 1.14$. A positive value for γ implies an elasticity of *intratemporal* substitution that is increasing in the level of consumption across the population while markups are decreasing in the same quantity. Figure 1 confirms this

[11] See equation (A.16) in Appendix A.2.

[12] 'Though ρ_a and σ_a are part of the SMM procedure below, we check that their point estimates are not far off the values obtained by OLS (auto)regression on the time series of Total Factor Productivity at Constant National Prices for United States' (part of FRED database maintained by St. Louis FED).

coutercyclicality by plotting the equilibrium mark-up (42) conditional on the log-TFP state a_t. Mark-up values range from approximately 31% in good times to 60% in bad times, consistent with estimates based on macro data, though larger than those reported by the micro literature.[13] The unconditional value is 41%, larger than the 16% obtained by Cavallari and Etro (2020) in the Bayesian estimation of their representative agent model with *IES* preferences.[14] The 'aggregate' elasticity of *intratemporal* substitution ϵ_t – defined in Section 2.4 – implied by these estimates ranges from 4.2 in good times to 2.6 in bad times, and the 'aggregate' elasticity of *intertemporal* substitution χ_t displays a similar procyclical pattern, increasing from 1.31 in bad times, to 1.54 in good times. Both ranges are consistent with those typically entertained in macroeconomics.[15] In particular, the procyclical behavior of the intertemporal elasticity is consistent with evidence from Attanasio and Browning (1995b), who argue that it is increasing in the level of aggregate consumption.

Table 2 reports actual and simulated moments, with t-statistics for the hypothesis that they are statistically distinct, in addition to the J-statistics for the test of over-identifying restrictions. The J-test does not reject the specification at even the 10% level, implying that the model provides a good overall match to the set of moments viewed collectively. Most simulated moments in Table 2 match the corresponding data moments with sufficient accuracy. For all moment conditions we can reject the hypothesis that simulated and actual moments are statistically different at the 5% confidence level. Only in two cases we can't reject it at the 10% level: variance of aggregate consumption and, (marginally) correlation between aggregate consumption and output, whereby the model overshoots in both cases. This is a consequence of the well-known tension that frictionless RBC models – with aggregate uncertainty driven by log-TFP alone – face when asked to match both volatile investment and smooth consumption. The estimation tries to cope with this trade-off by increasing the volatility of log-TFP (σ_a) to levels higher than usual, while decreasing its persistence (ρ_a) to the bottom of the plausible range.

In Table 3 we report a set of statistics implied by the fitted moments which is easier to interpret. The goodness-of-fit of the model with regard to standard deviations and correlations of aggregate consumption, investment, output and labor is quite apparent. In particular, the correlation between consumption and output closely resembles the data, overcoming a well-known difficulty of standard RBC models in this respect. Investments and labor policies, though, appear excessively procyclical. First order auto-correlations, on the other hand, show that empirical persistence levels are closely mimicked. Importantly, the model is close to replicating a few characteristics of the unconditional distribution of wealth. The 2% wealth share of the bottom 5% of the population is matched exactly, whereas the model-implied wealth share of the top 5% is 38%, smaller than the 48% observed empirically. 2 confirms that the latter moment condition is significantly different from zero at standard confidence level. Indeed, a Gini coefficient of 0.61, as opposed to an empirical value of 0.82, suggests that the model falls short in generating a degree of wealth inequality that is fully consistent with the data. This reflects the well-known difficulty to match the high concentration observed in the extreme upper tail of the wealth distribution in the data. Moreover, the requirement to fit both a battery of

[13]See, for instance, Rotemberg and Woodford (1999) and Basu and Fernald (1997)

[14]Cavallari and Etro (2020) obtain a larger point estimate for the linear component, $\gamma = 1.99$.

[15]See for instance the meta-analysis in Havranek et al. (2015).

macroeconomic moments and key characteristics of the wealth distribution has proven quite a challenging task, leading to a slight worse outcome in the distribution compartment, relative to studies targeting only the latter but ignoring macro-moments.[16] Interestingly, Figure 3 shows that wealth inequality is countercyclical in our model, because both the Gini coefficient and the top 5% wealth share in the *conditional* wealth distribution are inversely related to the log-TFP state. We will come back to this point in the sequel.

It is interesting at this point to gauge the extent to which the time-varying elasticity of substitution contributes to the goodness-of-fit, in our baseline specification. To this end, we have applied the same estimation strategy to a version of the model featuring constant elasticity of substitution, namely a *CES* aggregator, which amounts to imposing the constraint $\gamma = 0$. Results of this exercise are reported in columns 'CES-HA' of Tables 1, 2 and 3. The J-statistics clearly shows that the CES specification provides a worse overall match to the set of moments viewed collectively. In particular, it struggles to match the low standard deviation of aggregate consumption and the large variability of investment. Moreover, the properties of the theoretical distribution of wealth appear inconsistent with their empirical counterparts, due to an inadequate amount of inequality, highlighted by the small top 5% wealth share and Gini coefficients. We can conclude that the data appear to favor our specification with increasing elasticity of substitution.

6 Wealth and shock propagation

We now turn to the transmission mechanism at work in the model. We start with an intuitive illustration of the aggregate dynamics by means of impulse response functions. Then, we consider optimal policies and the wealth distribution conditional on the state of the economy. Figure 7 displays the responses of selected variables to a one standard deviation increase in productivity (positive supply hock) together with confidence bands at the 5% level of significance (shaded areas). Black lines refer to the benchmark model with IES elasticity, while red lines correspond to the variant with constant elasticity (CES). Figure 8 does the same for the benchmark model (black lines) against the representative agent model (red lines).

The aggregate dynamics is qualitatively similar across specifications: output and its components boost in the aftermath of the productivity rise. Markups, on the contrary, are countercyclical under IES preferences. Quantitatively, the responses are larger under IES preferences, though the differences are not always statistically significant (confidence bands overlap for output, employment and investment). The response of consumption, for instance, is more than twice as large, reflecting a strong incentive to consume in periods in which markups are low. Interestingly, the wealth effects are very large indeed in the benchmark economy, much larger (in absolute value) than under CES preferences. The share of wealth held by the bottom 5% increases by almost 6 percent against less than 1 percent with CES preferences. In addition, there is a significant drop in the wealth share at the top of the distribution (the magnitude is around 10 percent) while the response is not statistically significant at standard confidence levels under CES preferences. The reason is the lack of amplification effects in the

[16] Carroll et al. (2017), for instance, propose an heterogeneous-agents RBC model with life-cycle labor income. As in our case, they exploit the idea of time-preference heterogeneity to match the empirical wealth distribution, originally proposed by Krusell and Smith (1998).

CES case. Not only is the coutercyclical markup channel muted – as the intratemporal elasticity is constant – but also the intertemporal elasticity channel is absent, since this quantity is identically equal to one in the log-CES case. In particular, there is no increase in the substitution motives of the less wealthy to propel the responses, nor a substantial redistribution of wealth in favor of the lower range of the population, to the extent that the response of wealth shares is negligible. Our insight is that macroeconomic dynamics and wealth inequality are linked, and the reasons why increasing elasticity generates more aggregate volatility are the same reasons why it also generates a more skewed wealth distribution compared to the CES case. In fact, under increasing elasticity the optimal policies at the extremes of the distribution are further apart from each other. The more heterogeneous are the optimal policies in the tails of the distribution and the fatter these tails, the larger is the effect of aggregate shocks on wealth inequality. We will soon come back to this.

Not surprisingly, the responses in the benchmark economy are not significantly different from those generated by a representative agent economy with the same preferences (Figure 8).[17] In fact, a well-known implication of approximate aggregation is that aggregate variables can be almost perfectly described as functions of the mean of the wealth distribution and the aggregate shock. The approximation, however, is less likely to hold for consumption, especially in the very early aftermath of the shock. As it will be apparent in a while, this reflects the fact that consumption policies are non-linear with respect to wealth and they are heterogeneous across agents, so that aggregate shocks have a strong impact on wealth inequality. Large wealth effects are known to weaken the approximation (for instance, Huggett (1997), and Krueger and Kubler (2004)). At this point, it is worth exploring in detail the mechanics of shock propagation under increasing elasticity. In both the HA and the RA models, we observe the amplification mechanism described in Cavallari and Etro (2020). A positive TFP shock reduces markups (because of increasing *intratemporal* elasticity), while increasing wages and interest rates. The effect of lower prices today compared to the future is amplified by an increase in the *intertemporal* elasticity of substitution. Hence, the additional labor income, prompted by the wage and the demand increase, finances both consumption (substitution due to lower markup) and investment (substitution due to higher interest rate), thereby boosting output. The dynamics of consumption (and saving) is however different. In the RA model, the response of consumption is initially muted, it picks up in later periods and then dies out slowly compared to the heterogeneous agent case. This reflects the fact that it takes time for markups to adjust to a higher level of consumption: the initial drop is almost exclusively driven by the TFP shock alone, while markups further reduce over time as long as consumption increases. Approximate aggregation in the HA model implies that agents fully anticipate the entire dynamics of markups, and are therefore induced to react more strongly in the immediate aftermath of the shock. In addition, the response of aggregate consumption reflects the large contribution of the less wealthy group of the population, who have higher MPC.

The benchmark and its representative agent counterpart appear to achieve things differently also in the long run. In order to see why, compare the steady-state density of wealth in the RA model and the distribution of wealth in the benchmark economy (Figure 2 plots).

[17]The moment fit is also similar. In Table 2, the J-test does not reject the model and moment conditions are not statistically distinguishable from zero at standard confidence levels.

The density in the RA economy considers a rather limited range of wealth, concentrated in the upper end of the range visited by the population in the heterogeneous-agent model. Therefore, while the two models may generate similar aggregate outcomes, the representative agent model does so by visiting a rather concentrated portion of the optimal policies, whereas the heterogeneous agent model averages the very distinct policies at the extremes of the wealth distribution, where most of the mass is concentrated (Figure 2).

To gauge the extent to which the demand structure is important for wealth consider the optimal policies. Figure 4 displays optimal consumption, investment and employment policies as functions of wealth, conditional on the aggregate state (good versus bad state) for the benchmark economy and for the variant with CES preferences. Remarkably, the level of desired (optimal) consumption is higher for any level of wealth under IES preferences (while investment and employment are correspondingly lower), reflecting a stronger incentive to consume out of disposable income (higher MPC). In addition, consumption increases non-linearly with wealth and does so for a much larger range of wealth before turning linear (recall that non-linearities are critical for generating wealth effects). Last but not least, the benchmark economy implies far more diverse policies in response to the employment status, so that unemployed have much lower consumption levels than employed for a large span of the wealth distribution (around the bottom three deciles). In the CES model, on the contrary, heterogeneity is not only less substantial but it is also concentrated at the very bottom of the wealth distribution.

These differences reflect the incentives brought about by endogenous variability in demand elasticity. An increasing elasticity of *intratemporal* substitution generates countercyclical fluctuations in markups that are absent under CES preferences. These, in turn, affect the price of current relative to future consumption in a direction that reduces consumption smoothing. In order to see why, consider a cyclical downturn (a low productivity state in our framework) leading to a temporary drop in incomes (and therefore in consumption) throughout the economy. High markups induce agents to cut current consumption even further, in order to take advantage of the lower price of future consumption. The effect is particularly strong for unemployed and less wealthy agents, who spend a larger share of their income for consumption purposes.

Moreover, the elasticity of *intertemporal* substitution is positively related with the level of consumption, while it is constant under CES preferences. This implies that the population will respond heterogeneously to aggregate fluctuations. As Figure 4 shows, households with low levels of wealth display weaker *intertemporal* substitution motives (especially if unemployed) compared to wealthier households. Consequently, their precautionary savings will be smaller for any level of wealth and for any state of the economy. The endogenous variability of the elasticity of intertemporal substitution in our setup is in line with evidence stressing that the asset demand of wealthier agents is far more elastic than the demand of the less wealthier group (e.g., Guvenen (2006)).

The ample heterogeneity of intertemporal substitution in our benchmark generates not only a more skewed profile of wealth distribution compared to the CES case (see Figure 2), but also countercyclical movements in wealth inequality. In Figure 2, the Lorenz curve in the bad state lays below the curve in the good state, pointing to an increase of inequality in recessions, in line with the behavior of the Gini coefficient as a function of the aggregate state

(Figure 3). The variability of inequality over the business cycle comes mainly from the behavior of households in the lower range of the distribution, who grab a higher (lower) fraction of wealth in good (bad) times. For poorer households, a positive TFP shock increases the elasticity of intertemporal substitution much more than for wealthier households (Figure 4). Therefore, the additional labor income (for the employed, who are the majority in the steady state, and benefit from higher wages) finances not only increased consumption (reacting to lower markups), but also higher savings (reacting to an increase in the interest rate). On top of these effects, there is the usual desire to accumulate wealth in good times to smooth consumption in bad times (remember that the productivity process is mean-reverting).

It is useful at this point to consider the properties of the marginal propensity to consume both across the population and in aggregate terms. The MPC is the (expected) fraction consumed out of a windfall wealth gain over a discrete time interval, a year in our case. It is important to understand the behavior of this indicator over different fringes of the population, in order to quantify the aggregate effect. Figure 6 reports MPCs conditional on the state of the economy for the baseline model (IES-HA), for the variant with heterogeneous agents and CES preferences (CES-HA) and for the representative-agent model with IES preferences (IES-RA). Extensive microeconomic evidence reports estimates in the range between 0.2 and 0.6 (for a comprehensive survey see, among others, Carroll et al. (2017)). In addition, it supports the notion that the MPC is higher in recessions (Gross et al. (2020), for instance, argue that the MPC was 20 to 30% higher during the Great Recession relative to normal times).

The baseline model delivers an unconditional value for the aggregate MPC equal to 0.21, which is in the lower range of empirical estimates. Interestingly, the MPC varies in a contercyclical way as is observed in the data, ranging between a value as low as 0.1 in very good times and a value of 0.3 in very bad times. At a microeconomic level, the MPC decreases with wealth, and is maximal for poor and unemployed households, once again in line with evidence. Therefore, the aggregate MPC is mainly driven by the large cluster of population in the low-wealth range. Qualitatively, the behavior of the MPC is similar also in the variant with CES preferences and in the representative agent model under IES preferences.[18] In both these models, however, the unconditional value falls largely short of the empirical range. Moreover, the RA model generates far less variability of the MPC in terms of wealth. In fact, the economy here spends most of the time in a wealth region where the MPC peaks at about 0.11 (see the steady state distribution reported in Figure 2). We deduce that IES preferences, by delivering a state-dependent elasticity of intertemporal substitution – smallest for poor households – are important for the performance of the model. Clearly, the absence of idiosyncratic risk in the RA version takes its toll.

Given the importance of the demand structure in our setup, we finally explore its sensitivity with respect to the parameter γ, which governs the linear (non-homothetic) component of consumption utility (13). A non-zero value for γ represents a deviation from the CES paradigm, while $\gamma > 0$ ($\gamma < 0$) captures an increasing (decreasing) elasticity of substitution. In Figure 9, we consider four different values, two of which larger and two smaller than the point estimate obtained in the SMM exercise. All other parameters are unchanged. We then

[18] Carroll et al. (2017) have shown similar results in the context of a CES-HA model with idiosyncratic income shocks and heterogeneous time-preference rates. In quantitative terms, our variant of the CES-HA model performs disadvantageously compared to previous studies. On the bright side, our variant targets a broader range of moment conditions.

solve the model for each case, and report optimal policies and other equilibrium quantities.

Larger values of γ increase the *intratemporal* elasticity of substitution, and therefore generate smaller equilibrium markups for all TFP states. The *intertemporal* elasticity of substitution is also increasing in γ for all wealth levels. The combination of these two effects increases the optimal consumption policy for both the employed and the unemployed, while the smaller marginal utility of wealth depresses labor demand and investment. In Figure 10 we report the properties of the wealth distribution and the marginal propensity to consume obtained from the same exercise. Consistently with the notion that increasing elasticity favors the clustering of the wealth distribution towards the extremes, the wealth share of the richest (poorest) 5% of the population increases (decreases) with γ. The enhanced bimodality of the distribution, in turn, leads to more inequality, as is confirmed by the positive relation between the Gini coefficient and γ. The exercise confirms the ability of our setup with increasing elasticity to predict wealth distributions aligned with empirical evidence. The same conclusion holds for the marginal propensity to consume. The MPC is increasing in γ for less wealthy households, and especially in bad states. Richer households, on the contrary, experience almost no effect.

7 Conclusions

In this paper, we have introduced a general class of additive separable non-homothetic preferences over differentiated goods in an otherwise standard incomplete market (SIM) heterogeneous agents (HA) model, featuring both idiosyncratic (income) and aggregate (TFP) shocks, as in Krusell and Smith (1998). The combination of heterogeneity and non-homotheticity opens the way to a rich demand system, whereby optimal consumption (and saving) policies depend on novel mechanisms of intertemporal substitution varying across the population. The model's numerical predictions – made possible by a tractable yet global solution approach – have allowed us to argue that demand plays an important role for wealth inequality and the propagation of shocks in the economy.

References

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2017): "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," .

Attanasio, O. and M. Browning (1995a): "Consumption over the Life Cycle and over the Business Cycle," *American Economic Review*, 85, 1118–37.

——— (1995b): "Consumption over the Life Cycle and over the Business Cycle," *American Economic Review*, 85, 1118–37.

Attanasio, O. and G. Weber (2010): "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy," NBER Working Papers 15756, National Bureau of Economic Research, Inc.

Attanasio, O. P. and L. Pistaferri (2016): "Consumption Inequality," *Journal of Economic Perspectives*, 30, 3–28.

Basu, S. and J. Fernald (1997): "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 105, 249–83.

Benhabib, J., A. Bisin, and M. Luo (2019): "Wealth Distribution and Social Mobility in the US: A Quantitative Approach," *American Economic Review*, 109, 1623–47.

Bertola, G., R. Foellmi, and J. Zweimüller (2005): *Income Distribution in Macroeconomic Models*, no. 8058 in Economics Books, Princeton University Press.

Bertoletti, P. and F. Etro (2016): "Preferences, entry, and market structure," *RAND Journal of Economics*, 47, 792–821.

Carroll, C., J. Slacalek, K. Tokuoka, and M. N. White (2017): "The distribution of wealth and the marginal propensity to consume," *Quantitative Economics*, 8, 977–1020.

Cavallari, L. (2022): "The international real business cycle when demand matters," *Journal of Macroeconomics*, 73, 103445.

Cavallari, L. and F. Etro (2020): "Demand, markups and the business cycle," *European Economic Review*, 127, 103471.

Constantinides, G. M. and D. Duffie (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104, 219–240.

Cox, J. C. and C.-f. Huang (1989): "Optimal consumption and portfolio policies when asset prices follow a diffusion process," *Journal of Economic Theory*, 49, 33–83.

De Nardi, M. (2004): "Wealth Inequality and Intergenerational Links," *The Review of Economic Studies*, 71, 743–768.

Duffie, D. and K. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–52.

Fernández-Villaverde, J., S. Hurtado, and G. Nuño (2022): "Financial Frictions and the Wealth Distribution," CESifo Working Paper Series 8482, CESifo.

Gross, T., M. J. Notowidigdo, and J. Wang (2020): "The Marginal Propensity to Consume over the Business Cycle," *American Economic Journal: Macroeconomics*, 12, 351–84.

Guvenen, F. (2006): "Reconciling conflicting evidence on the elasticity of intertemporal substitution: A macroeconomic perspective," *Journal of Monetary Economics*, 53, 1451–1472.

——— (2011): "Macroeconomics With Heterogeneity: A Practical Guide," Working Paper 17622, National Bureau of Economic Research.

Havranek, T., R. Horvath, Z. Irsova, and M. Rusnák (2015): "Cross-country heterogeneity in intertemporal substitution," *Journal of International Economics*, 96, 100–118.

He, H. and H. F. Pagès (1993): "Labor Income, Borrowing Constraints, and Equilibrium Asset Prices," *Economic Theory*, 3, 663–696.

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, special issue: Cross-Sectional Facts for Macroeconomists.

Heathcote, J., K. Storesletten, and G. L. Violante (2009): "Quantitative Macroeconomics with Heterogeneous Households," Working Paper 14768, National Bureau of Economic Research.

Huggett, M. (1997): "The one-sector growth model with idiosyncratic shocks: Steady states and dynamics," *Journal of Monetary Economics*, 39, 385–403.

Kaplan, G., K. Mitman, and G. L. Violante (2020): "The Housing Boom and Bust: Model Meets Evidence," *Journal of Political Economy*, 128, 3285–3345.

Krueger, D. and F. Kubler (2004): "Computing equilibrium in OLG models with stochastic production," *Journal of Economic Dynamics and Control*, 28, 1411–1436.

Krusell, P. and A. A. J. Smith (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, pp. 867–896.

Lee, B.-S. and B. Ingram (1991): "Simulation estimation of time-series models," *Journal of Econometrics*, 47, 197–205.

Maestri, V. and A. Roventini (2012): "Inequality and Macroeconomic Factors: A Time-Series Analysis for a Set of OECD Countries," .

Mankiw, N. G. (1986): "The equity premium and the concentration of aggregate shocks," *Journal of Financial Economics*, 17, 211–219.

Newey, W. and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–08.

Quadrini, V. and J.-V. Ríos-Rull (2015): "Chapter 14 - Inequality in Macroeconomics," in *Handbook of Income Distribution*, ed. by A. B. Atkinson and F. Bourguignon, Elsevier, vol. 2 of *Handbook of Income Distribution*, 1229–1302.

Rotemberg, J. and M. Woodford (1999): "The cyclical behavior of prices and costs," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, Elsevier, vol. 1, Part B, chap. 16, 1051–1135, 1 ed.

Rouwenhorst, K. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," *In: Cooley, T.F. (Ed.), Frontiers of Business Cycle Research. Princeton University Press, Princeton, NJ*, 294–330.

Straub, L. (2019): "Consumption, Savings, and the Distribution of Permanent Income," Revise and resubmit at Econometrica.

Tables

			Model: IES-HA				
λ^{β}_{12}	λ^{β}_{21}	λ^{β}_{23}	λ^{β}_{32}	ρ_a	σ_a	γ	φ
0.747	0.687	0.942	1.0	0.750	0.021	0.900	0.974
(0.075)	(0.093)	(0.174)	(0.168)	(0.19)	(0.009)	(0.306)	(0.261)
θ							
1.144							
(0.08)							
			Model: CES-HA				
λ^{β}_{12}	λ^{β}_{21}	λ^{β}_{23}	λ^{β}_{32}	ρ_a	σ_a	φ	θ
0.891	0.698	0.209	0.393	0.750	0.0096	0.874	1.434
(0.190)	(0.326)	(0.119)	(0.132)	(0.092)	(0.003)	(0.161)	(0.182)
			Model: IES-RA				
ρ_a	σ_a	φ	γ	θ			
0.927	0.0098	0.973	1.276	2.275			
(0.108)	(0.0051)	(0.319)	(0.432)	(0.292)			

Tab. 1: SMM Point Estimates. *Point estimates for the parameters of the heterogeneous agents model with IES (IES-HA) or CES (CES-HA) preferences, and the representative-agent model with IES preferences (IES-RA). Estimates are obtained with the Simulated Method of Moments, as described in the Appendix. Asymptotic standard errors of parameter estimates are reported in parenthesis.*

	IES-HA	CES-HA	IES-RA	Data
$\mathbb{E}[(\int_0^{0.25} Y_s ds)^2] - \overline{Y}^2$	0.0002999 (1.132)	0.0003409 (1.740)	0.0003519 (1.832)	0.000256
$\mathbb{E}[(\int_0^{0.25} L_s ds)^2] - \overline{L}^2$	0.000085 (0.361)	0.000015 (-1.069)	0.000115 (1.269)	0.000079
$\mathbb{E}[(\int_0^{0.25} C_s ds)^2] - \overline{C}^2$	0.000189 (1.564)	0.000321 (2.949)	0.00029 (1.864)	0.000151
$\mathbb{E}[(\int_0^{0.25} I_s ds)^2] - \overline{I}^2$	0.006525 (1.042)	0.004056 (-2.057)	0.004925 (-1.042)	0.005576
$\mathbb{E}[(\int_0^{0.25} Y_s ds)(\int_0^{0.25} L_s ds)] - \overline{YL}$	0.00007927 (-0.735)	0.0000543 (-2.474)	0.00012927 (0.435)	0.000095
$\mathbb{E}[(\int_0^{0.25} Y_s ds)(\int_0^{0.25} C_s ds)] - \overline{YC}$	0.000199 (1.7699)	0.000279 (1.471)	0.0000799 (-2.3699)	0.000152
$\mathbb{E}[(\int_0^{0.25} Y_s ds)(\int_0^{0.25} I_s ds)] - \overline{YI}$	0.0011459 (1.036)	0.000575 (-0.458)	0.001279 (1.236)	0.000973
$\mathbb{E}[(\int_0^{0.25} Y_s ds)(\int_{0.25}^{0.5} Y_s ds)] - \overline{Y}^2$	0.000240 (0.698)	0.000317 (2.006)	0.000290 (1.618)	0.000217
$\mathbb{E}[(\int_0^{0.25} L_s ds)(\int_{0.25}^{0.5} L_s ds)] - \overline{L}^2$	0.000059 (-0.207)	0.000013 (-1.308)	0.000089 (0.907)	0.000062
$\mathbb{E}[(\int_0^{0.25} C_s ds)(\int_{0.25}^{0.5} C_s ds)] - \overline{C}^2$	0.000154 (1.466)	0.000294 (15.741)	0.000104 (-1.066)	0.000124
$\mathbb{E}[(\int_0^{0.25} I_s ds)(\int_{0.25}^{0.5} I_s ds)] - \overline{I}^2$	0.005456 (1.299)	0.00600 (1.594)	0.004435 (0.084)	0.004432
$\mathbb{E}[WL_t]$	0.021 (0.116)	0.014 (-1.166)	-	0.02
$\mathbb{E}[WH_t]$	0.38 (-1.916)	0.162 (-4.166)	-	0.48
J-test	3.21 (p : 0.201)	5.31 (p : 0.150)	8.11 (p : 0.231)	

Tab. 2: Optimal Moment Conditions in the SMM estimation. *Theoretical (i.e. fitted) and empirical moment restrictions in the Simulated Method of Moments estimation procedure, for the heterogeneous agents model with IES (IES-HA) or CES (CES-HA) preferences, and the representative-agent model with IES preferences (IES-RA). Theoretical (simulated) moments are evaluated at the point estimates, minimizing the SMM quadratic form. Parenthesis report the T-statistics (computed with asymptotic standard errors) for the hypothesis that moment conditions (i.e. theoretical moment - empirical moment) differ from zero.* WL_t *(*WH_t*) denotes the fraction of total wealth owned by the poorest (wealthiest) 5% of the population. The last row reports Hansen's J-test statistics, with its* $p-$*value in parenthesis. The notation is as follows: Aggregate Output (*Y*), Consumption (*C*), Investment (*I*) and Labor Supply (*L*).* $\overline{Y} = \mathbb{E}[\int_0^{0.25} Y_s ds]$*, and similarly for the other policies*

	IES-HA	CES-HA	IES-RA	Data
$\sigma(Y_t)$	0.0173	0.0184	0.0187	0.0160
$\sigma(L_t)$	0.0092	0.0038	0.0107	0.0089
$\sigma(C_t)$	0.0137	0.0179	0.0298	0.0123
$\sigma(I_t)$	0.0807	0.0636	0.0701	0.0746
$\rho(Y_t, L_t)$	0.495	0.759	0.642	0.668
$\rho(Y_t, C_t)$	0.836	0.844	0.515	0.773
$\rho(Y_t, I_t)$	0.819	0.489	0.971	0.815
$\rho(Y_t, Y_{t-1})$	0.803	0.931	0.824	0.847
$\rho(L_t, L_{t-1})$	0.689	0.912	0.774	0.785
$\rho(C_t, C_{t-1})$	0.818	0.914	0.570	0.821
$\rho(I_t, I_{t-1})$	0.836	0.616	0.980	0.794
% wealth bottom 5%	0.021	0.014	-	0.02
% wealth top 5%	0.38	0.162	-	0.48
Gini coeff. %	0.61	0.26	-	0.82

Tab. 3: Model-Implied Moments vs Empirical Moments. *Unconditional standard deviation (σ) and correlation (ρ) of Aggregate Output (Y), Consumption (C), Investment (I) and Labor Supply (L), implied by the heterogeneous agents models with IES (IES-HA) or CES (CES-HA) preferences. Model-implied moments are computed by simulation using the point estimates of the SMM, with the policy series simulated in the estimation itself. '% wealth bottom (top) 5%' is the percentage of wealth held by the population in the bottom (top) 5% of the unconditional wealth distribution. 'Gini coeff.' is the Gini coefficient of the unconditional wealth distribution.*

Figures

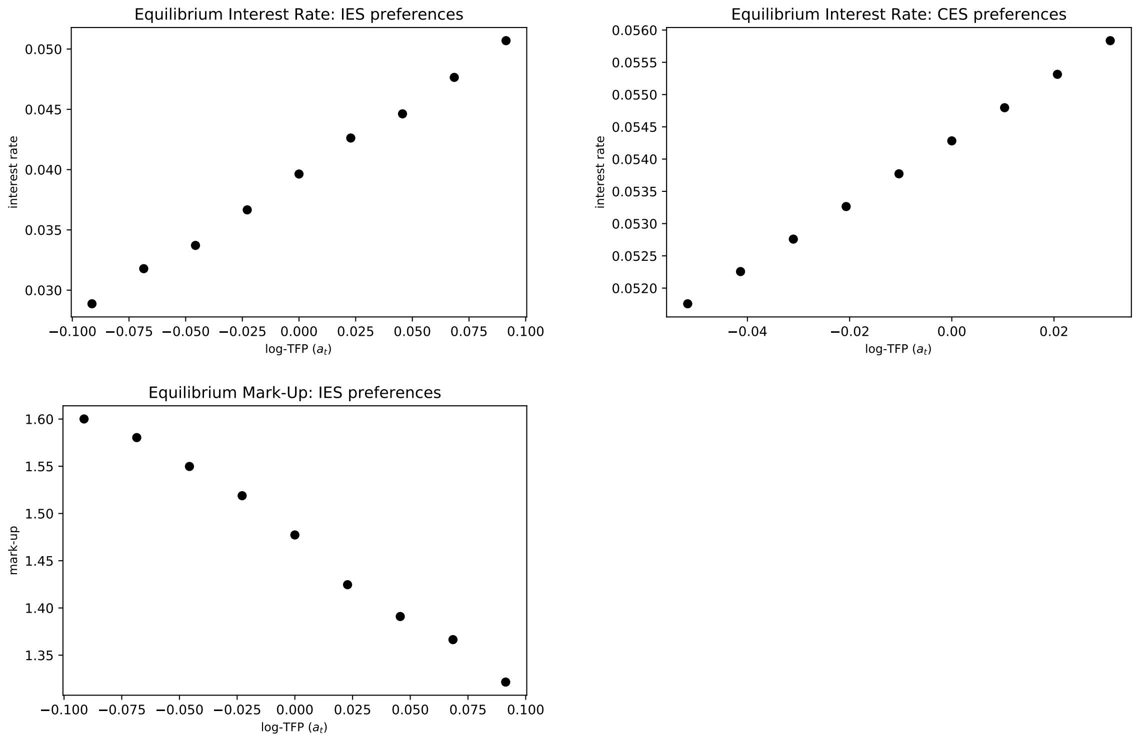

Fig. 1: Equilibrium Interest Rate and Mark-Up*. Upper Panels: equilibrium interest rate in the heterogeneous-agents model with IES preferences (left panel) and CES preferences (right panel). Bottom Panel: equilibrium mark-up in the heterogeneous-agents model with IES preferences. All quantities are obtained used SMM parameter estimates.*

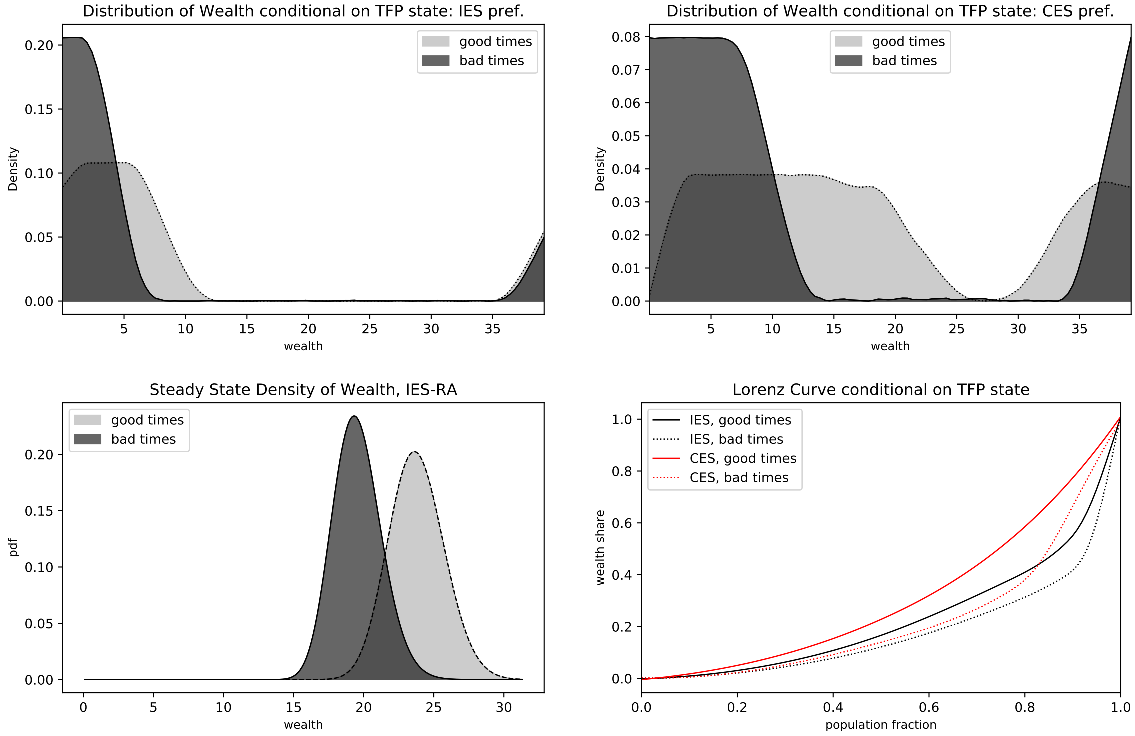

*FIG. 2: **Equilibrium Distributions of Wealth and Lorenz Curves**. Equilibrium distribution of wealth conditional on the lowest (dim area) and highest (light area) log-TFP state, for the heterogeneous-agents model with IES (left-upper panel) and CES (right-upper panel) preferences. The left-bottom panel shows the ergodic density of wealth for the representative-agent model with IES preferences. The right-bottom panel reports the equilibrium wealth share held by each fraction of the population (Lorenz curve), conditional on the lowest (dotted line) and highest (solid line) log-TFP state, for the heterogeneous-agents model with IES (black lines) and CES (red lines) preferences. Equilibrium distributions are obtained with SMM parameter estimates.*

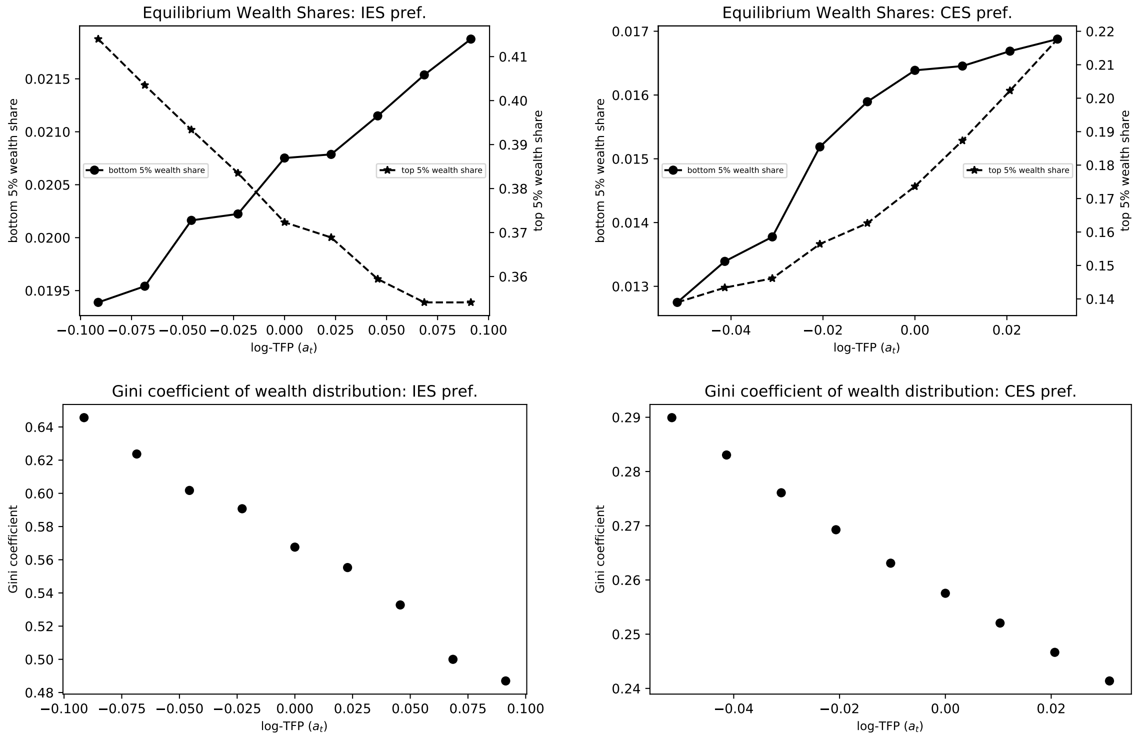

Fig. 3: Wealth Shares and Gini Coefficients of Equilibrium Distribution of Wealth*. Upper Panel: Percentage of wealth held by bottom 5% (circles, left $y-$axis) and top 5% (stars, right $y-$axis) of the equilibrium distribution of wealth, conditional on the log-TFP state, for the heterogeneous-agents model with IES (left panel) and CES (right panel) preferences. Bottom Panel: Gini coefficients of the equilibrium wealth distribution conditional on the log-TFP state, for the heterogeneous-agents model with IES (left panel) and CES (right panel) preferences. Equilibrium distributions are obtained with SMM parameter estimates.*

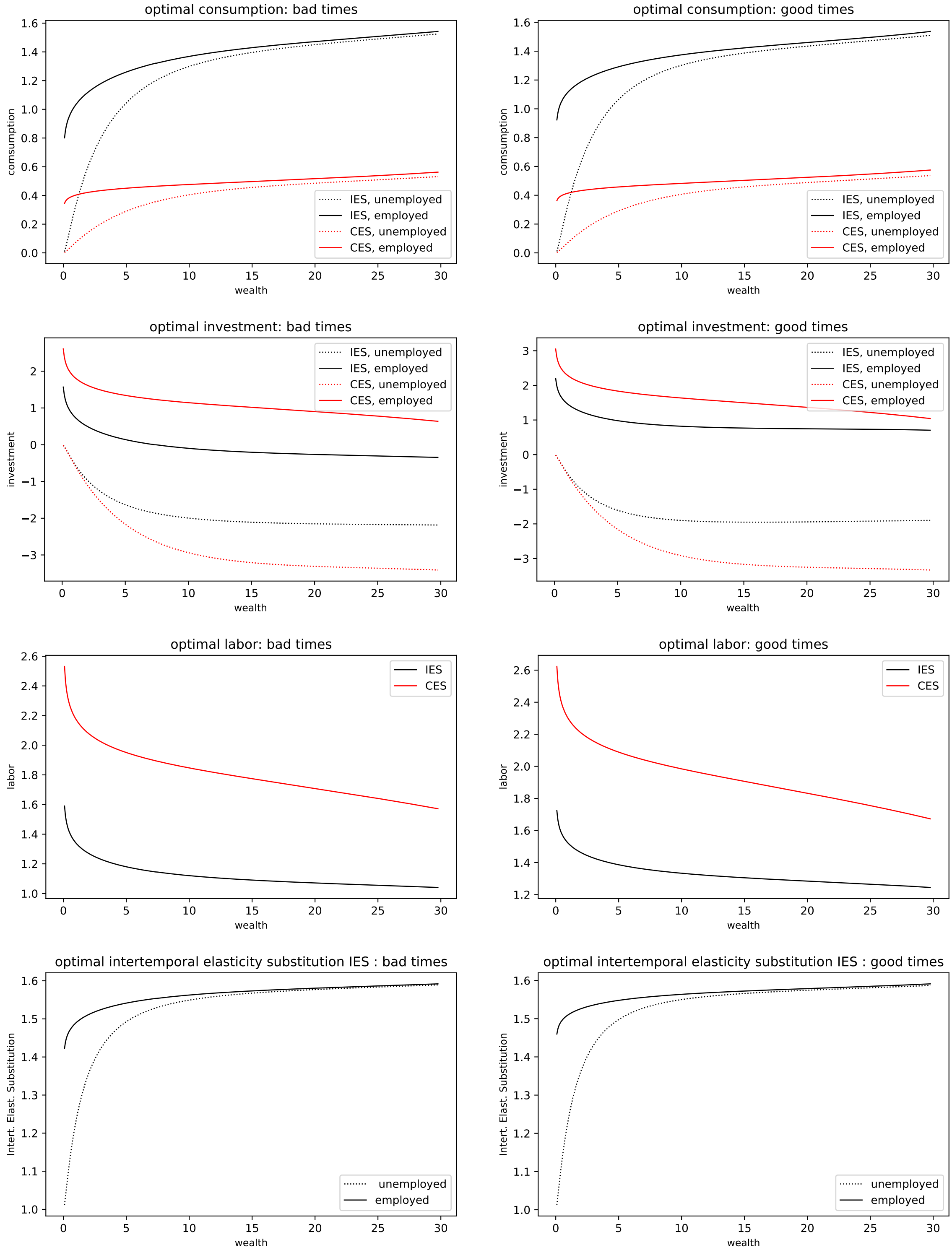

Fig. 4: Optimal Policies, and Intertemporal Elasticity of Substitution (IES vs CES). *Left (right) panels report quantities conditional on lowest (highest) log-TFP state a_1 (a_9). Red (Black) lines report policies for the heterogeneous-agents model with CES (IES) preferences, while solid (dotted) lines report policies for employed (unemployed) agents, $\varepsilon_t = 0$ ($\varepsilon_t = 1$). All quantities are plotted using the middle level of time preference rate, β_2, and using the parameters obtained in the SMM estimations.*

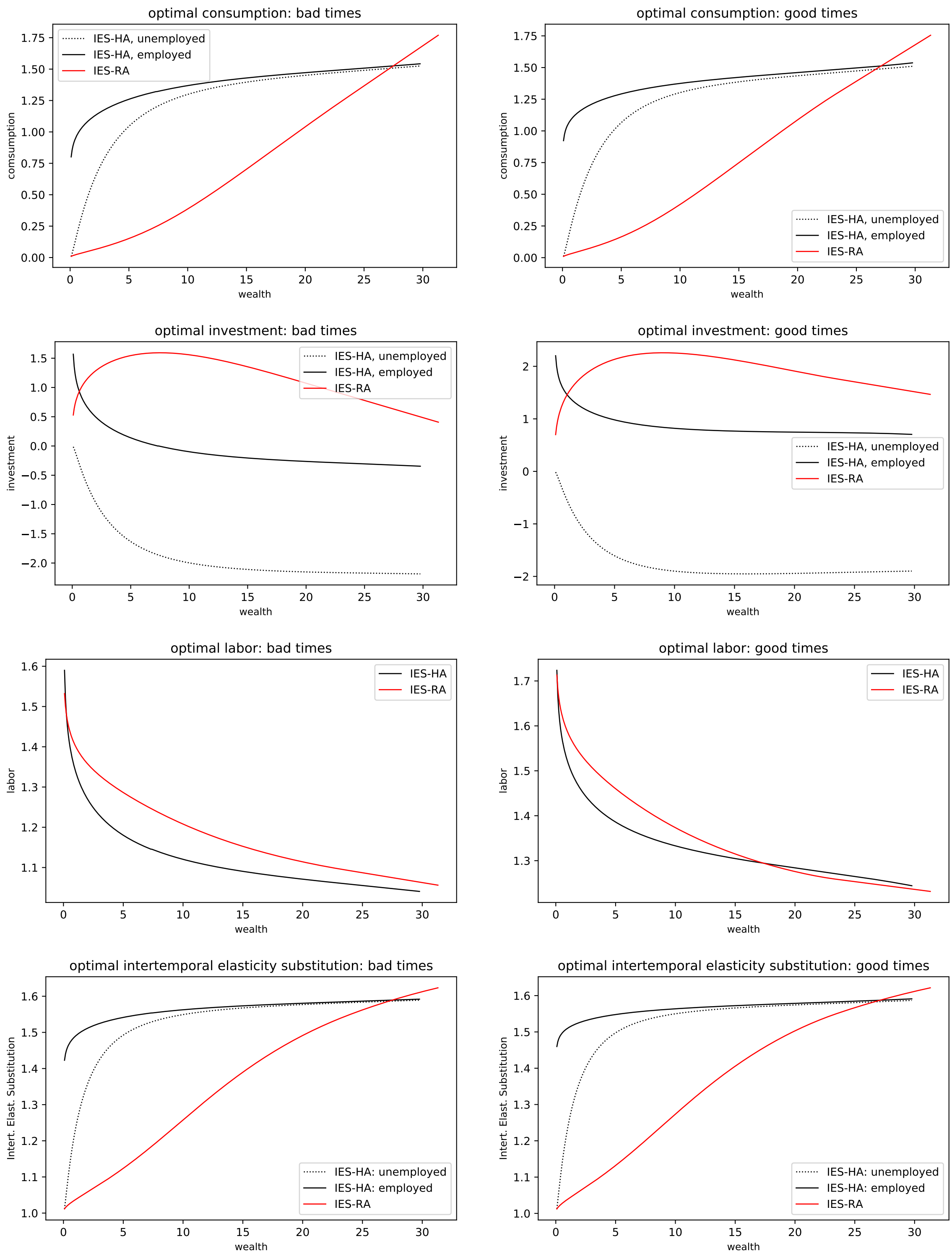

Fig. 5: Optimal Policies, and Intertemporal Elasticity of Substitution (HA vs RA). *Left (right) panels report quantities conditional on lowest (highest) log-TFP state a_1 (a_9). Red lines report policies for the representative-agent (RA) model with IES preferences, while black lines are policies for the heterogeneous-agents (HA) model with the same preferences . In the HA case, solid (dotted) lines report policies for employed (unemployed) agents, $\varepsilon_t = 0$ ($\varepsilon_t = 1$), and are plotted using the middle level of time preference rate, β_2. Parameters obtained in the SMM estimations have been used.*

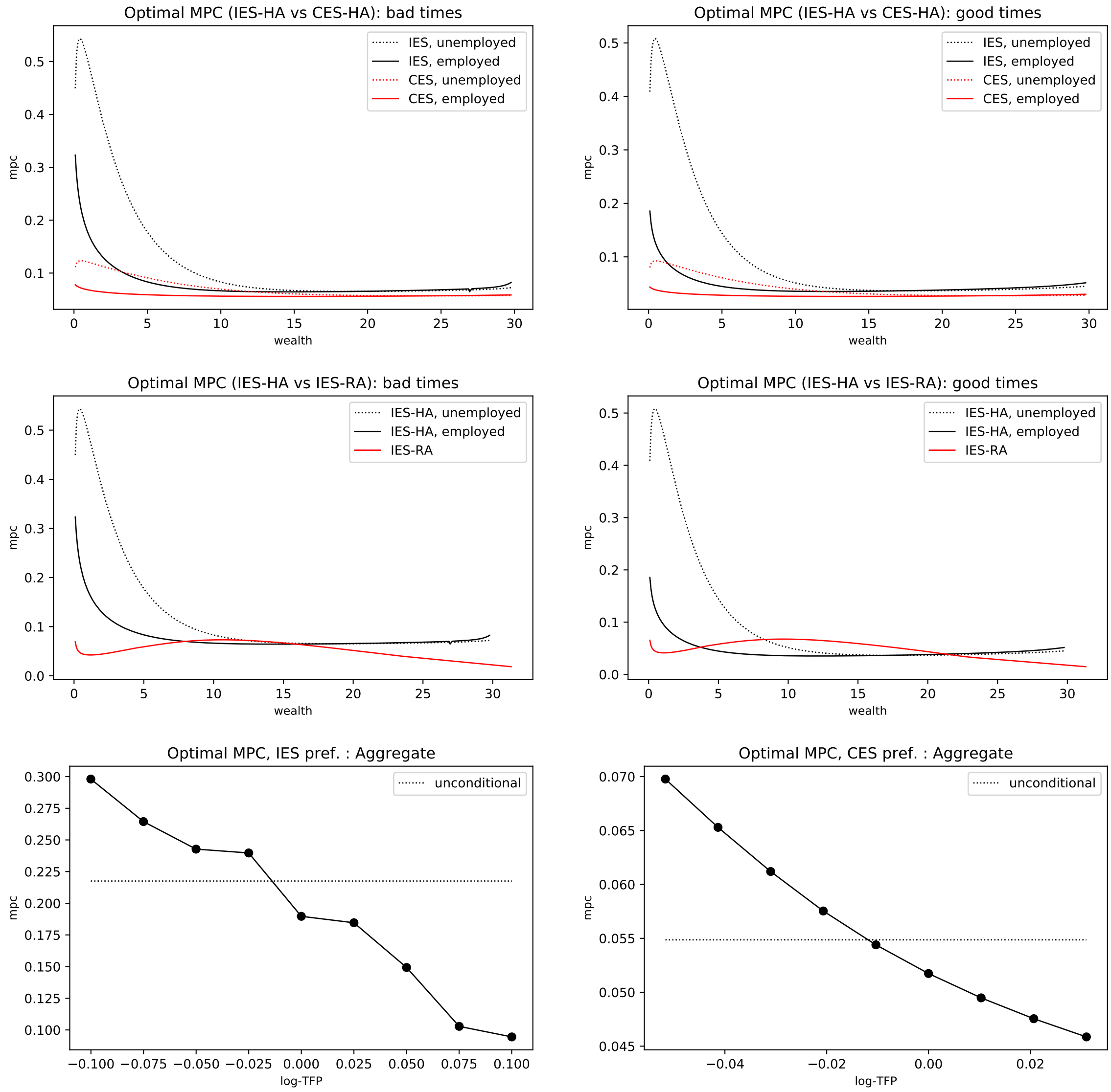

FIG. 6: Optimal Marginal Propensities to Consume. *Marginal propensities to consume (MPC) over a 1-year time horizon, corresponding to optimal consumption and saving policies, obtained with the SMM baseline parameter estimates. Top panels report conditional MPC, corresponding to the lowest (left) and highest (right) log-TFP state, for the heterogeneous-agents model with IES (black lines) and CES (red lines) preferences, employed (solid lines) and unemployed (dashed lines) agents. The middle panel reports the same quantities for the heterogeneous-agents model with IES preferences (IES-HA) and the representative-agent model with the same preferences (IES-RA). The bottom panel reports the aggregate MPC conditional on log-TFP, and the unconditional level as a dotted line, with IES (left) and CES preferences (right). All quantities are plotted using the middle level of time preference rate,* β_2.

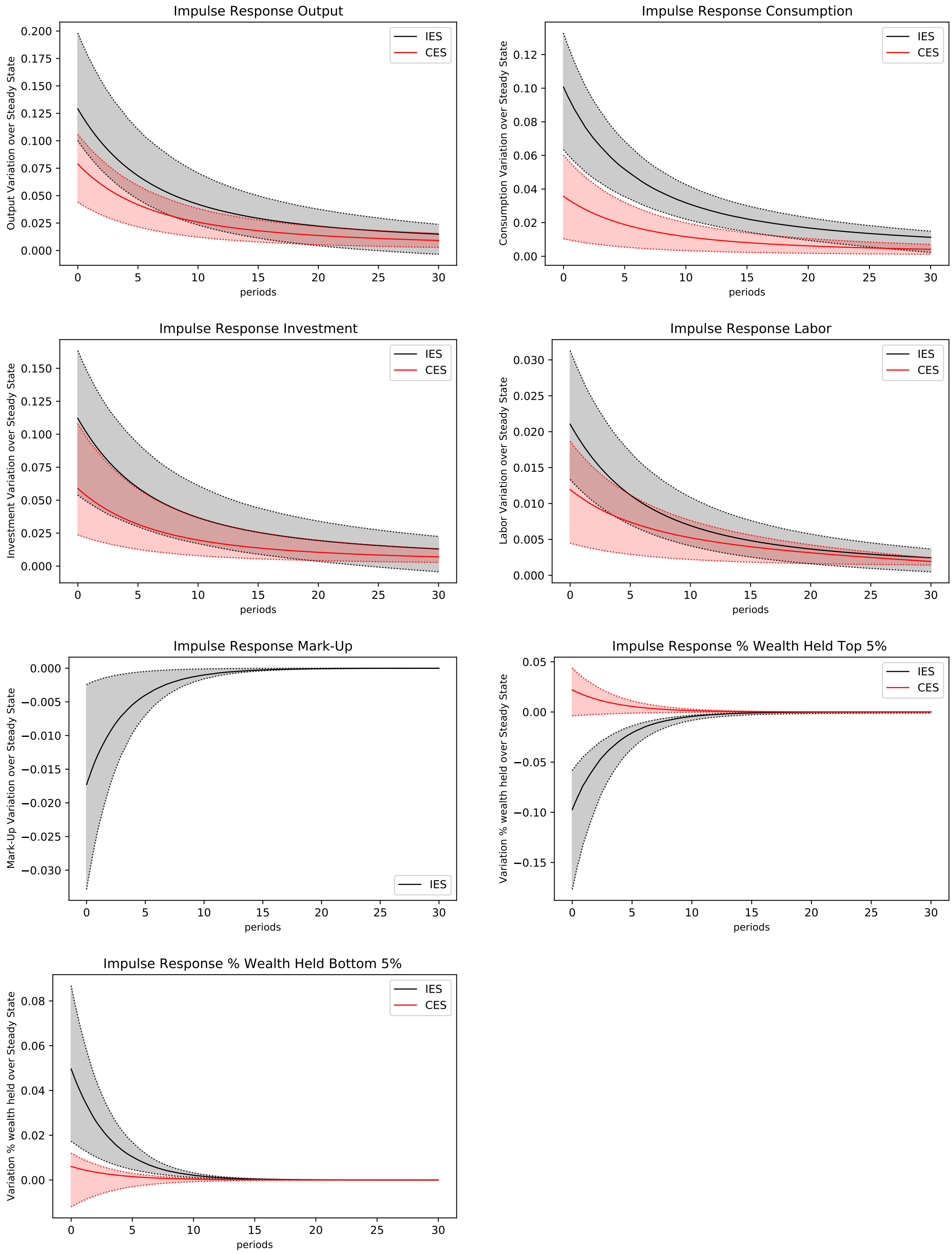

Fig. 7: ***Impulse Responses (IES vs CES)*** *. Impulse Responses of Output, Consumption, Investment, Labor Supply, Mark-Up, percentage of wealth held by top and bottom 5% of the population, presented as variations due to a positive one standard deviation TFP shock, as a fraction of steady state values. Solid lines are impulse responses corresponding to SMM point estimates of parameters, for the heterogeneous-agents model with IES (black lines) and CES (red lines) preferences, while dotted lines and the shaded area in between are 5-95% confidence bands.*

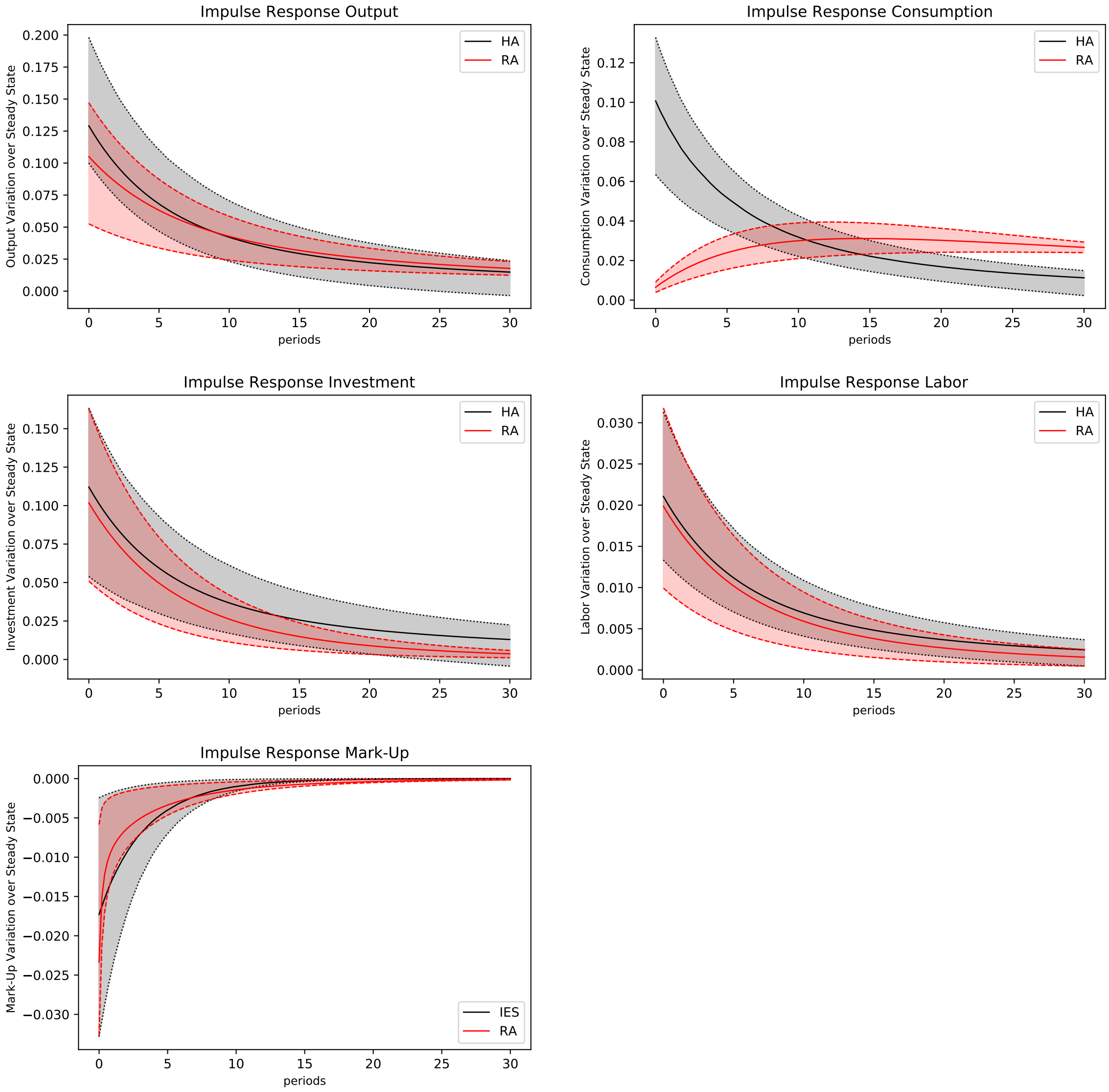

Fig. 8: Impulse Responses (HA vs RA) . *Impulse Responses of Output, Consumption, Investment, Labor Supply, Mark-Up, presented as variations due to a positive one standard deviation TFP shock, as a fraction of steady state values. Solid lines are impulse responses corresponding to SMM point estimates of parameters, for the heterogeneous-agents model with IES preferences (black lines) and the representative-agent model with the same preferences (red lines), while dotted lines and the shaded area in between are 5-95% confidence bands.*

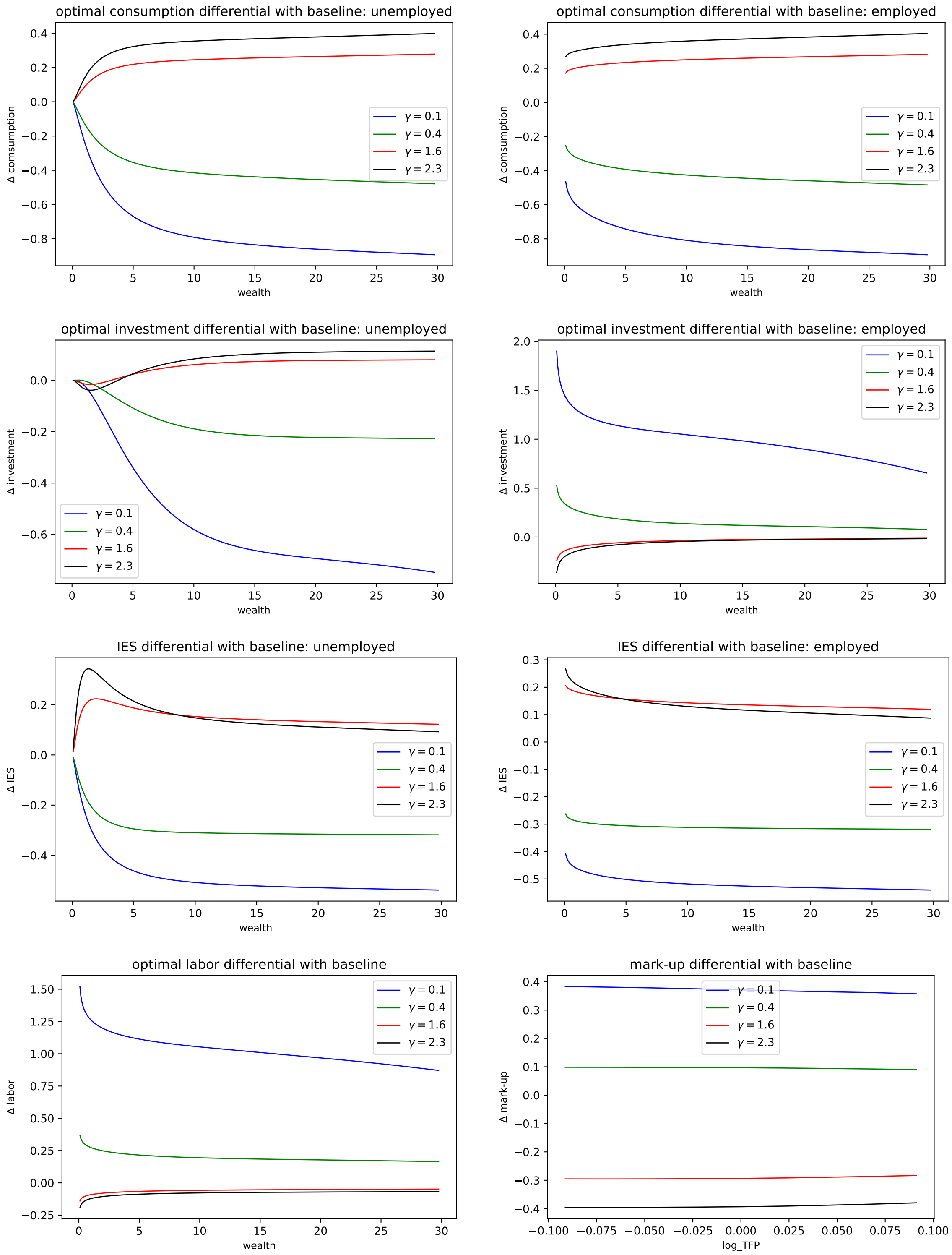

*Fig. 9: **Comparative Statics I: Optimal Policies. IES Heterogeneous Agents Model.** Optimal consumption, investment, labor policies, intertemporal elasticity of substitution, and equilibrium mark-up obtained for 4 different values of the parameter γ, and fixing others at their baseline SMM estimates. Deviation of policies from the baseline case are reported.*

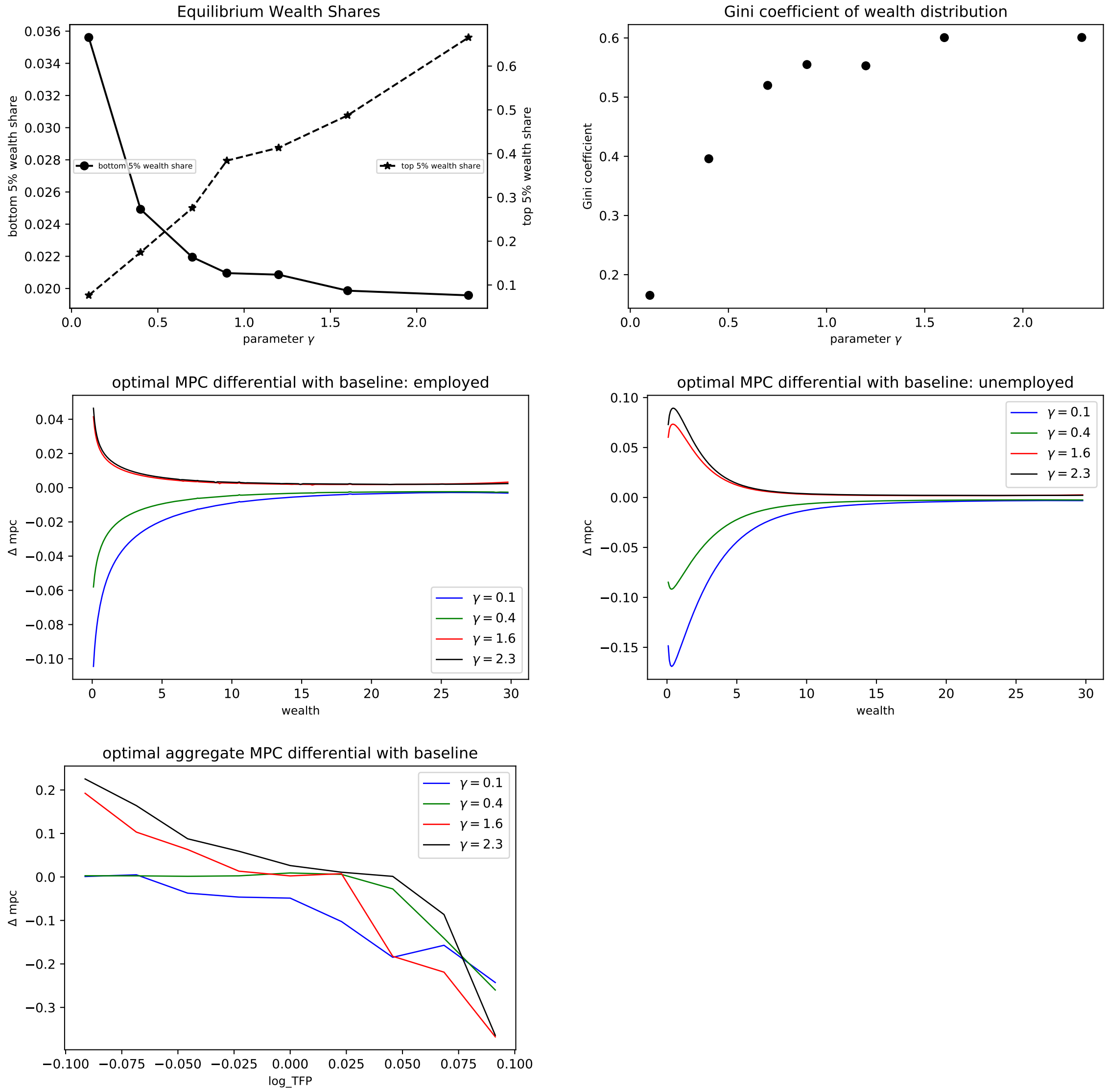

Fig. 10: Comparative Statics II: Wealth Shares and Marginal Propensities of Consumption. IES Heterogeneous Agents Model. *Top 2 panels: Wealth shares of the top and bottom 5% of the population, and Gini coefficient of the wealth distribution, obtained for 4 different values of the parameter γ, and fixing others at their baseline SMM estimates. Bottom 3 panels: Deviation of marginal propensities to consume from the baseline case, either at the individual agent level (first two panels) or at the aggregate level (last panel), in which case deviations are shown as a function of log-TFP state.*

Appendix

A Implementation Details

A.1 Numerical Solution of the HBJ equation, the KF equation, and computation of marginal propensity to consume

A.1.1 Numerical Solution of The HBJ equation (34)

In order to solve numerically the nonlinear ODE, we use an upwind finite difference scheme similar to Achdou, Han, Lasry, Lions, and Moll (2017), in conjunction with time iteration, which amounts to approximating the time invariant solution by iterating a time-varying one, after the introduction of a fictitious time dimension. We report the HBJ equation for convenience:[19]

$$
\begin{aligned}
V_t(k,\varepsilon_i,\beta_j,a_z) + \left(\beta_j + \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2} + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2} + \sum_{z_2\neq z}\lambda^{a}_{zz_2}\right) V(k,\varepsilon_i,\beta_j,a_z) = \log u(c^*_t) - v\frac{(l^*_t)^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}} \\
+ V_k(k,\varepsilon_i,\beta_j,a_z)s(k,\varepsilon_i,\beta_j,a_z) + \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2}V(k,\varepsilon_{i_2},\beta_j,a_z) + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2}V(k,\varepsilon_i,\beta_{j_2},a_z) \\
+ \sum_{z_2\neq z}\lambda^{a}_{zz_2}V(k,\varepsilon_i,\beta_j,a_{z_2})
\end{aligned}
\tag{A.1}
$$

with $u(\cdot)$ as in (13). We choose an equispaced capital grid $\{k_1,k_2,\dots,k_N\}$,[20] with $k_1=0$, k_N equal to three times capital in the deterministic steady state, and $\Delta k = (K_N - K_1)/(N-1)$. We denote by $V^n_{w,i,j,z}$ the value function at time iteration n, evaluated at k_w, $w=1,2,\dots,N$, and employment-preference-TFP state $(\varepsilon_i,\beta_j,a_z)$. We use a similar notation for the optimal consumption and labor policies $c^n_{w,i,j,z}$ and $l^n_{w,j,z}$, solving the first order conditions in (36), and the saving policy $s^n_{w,i,j,z}$ given in (35). In the upwind scheme, at each point of the grid the first derivative with respect to k can be approximated with a forward (f) or a backward (b) approximation,

$$
\frac{V^n_{w+1,i,j,z}-V^n_{w,i,j,z}}{\Delta k}, \quad \frac{V^n_{w,i,j,z}-V^n_{w-1,i,j,z}}{\Delta k} \quad \text{respectively,}
$$

depending on the sign of the drift function $s^n_{w,i,j,z}$ of the state variable k. Namely, we approximate the HBJ (A.1) with the following implicit upwind scheme:

$$
\frac{V^{n+1}_{w,i,j,z}-V^n_{w,i,j,z}}{\Delta} + \left(\beta_j + \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2} + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2} + \sum_{z_2\neq z}\lambda^{a}_{zz_2}\right) V^{n+1}_{w,i,j,z} = \log u(c^n_{w,i,j,z}) - v\frac{(l^n_{w,j,z})^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}
\tag{A.2}
$$

[19] We dropped the time dimension for ease of notation.
[20] Typically $N=500$ in our applications.

$$+\frac{V^{n+1}_{w+1,i,j,z}-V^{n+1}_{w,i,j,z}}{\Delta k}s^n_{w,i,j,z,F}\mathbf{1}_F+\frac{V^{n+1}_{w,i,j,z}-V^{n+1}_{w-1,i,j,z}}{\Delta k}s^n_{w,i,j,z,B}\mathbf{1}_B+\sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2}V^{n+1}_{w,i_2,j,z}$$
$$+\sum_{j_2\neq j}\lambda^{\beta}_{jj_2}V^{n+1}_{w,i,j_2,z}+\sum_{z_2\neq z}\lambda^{a}_{zz_2}V^{n+1}_{w,i,j,z_2}$$
$$w=1,\dots,N\quad i=1,2\quad j=1,2,3\quad z=1,\dots,n_a$$

where Δ is a time increment, $\mathbf{1}_F$ ($\mathbf{1}_B$) is the indicator function of the event $s^n_{w,i,j,z,F}>0$ ($s^n_{w,i,j,z,B}<0$), $s^n_{w,i,j,z,F}$ ($s^n_{w,i,j,z,B}$) being the saving policy with optimal consumption computed with the forward (backward) finite difference approximation of V_k. Thus when the drift of k is positive (negative) we employ a forward (backward) approximation of the derivative V_k. Notice that consumption, labor and saving policy are evaluated at time iteration n, not $n+1$, in order to preserve the linearity of the equation.

We impose boundary state constraints in k: $s^n_{1,i,j,z,B}=0$ and $s^n_{N,i,j,z,F}=0$. Since the first order conditions for consumption and labor hold at the boundaries, imposing this conditions amounts to identifying boundary values for V.

Assuming quantities at time iteration n have been solved for, expression (A.2) gives rise to a linear system of equations, the unknown of which is the $N\times 2\times 3\times n_a$ column vector $\boldsymbol{V}^{n+1}$, stacking the value function at all capital grid point and employment-preference-TFP states, at time iteration $n+1$. The linear system of equations (A.2) can be compactly written in matrix notation:

$$\boldsymbol{A}^n\boldsymbol{V}^{n+1}=\boldsymbol{b}^n \tag{A.3}$$

The $N\times 2\times 3\times n_a$ column vector $\boldsymbol{b}^n$ is

$$\boldsymbol{b}^n=\log u(\boldsymbol{c}^n)-v\frac{(\boldsymbol{l}^n)^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}+\frac{\boldsymbol{V}^n}{\Delta}$$

where we have used $\boldsymbol{c}$ and $\boldsymbol{l}$ to denote the optimal policy vectors, at all grid points and states, solving the first order conditions in (36) at iteration n. The $N\times 2\times 3\times n_a$-dimensional square matrix $\boldsymbol{A}^n$ can easily be deducted from (A.2). It is a highly sparse matrix, thus simplifying tremendously the numerical implementation of the problem. Time iteration mandates solving (A.3) iteratively until its solution reaches time invariance. Therefore, starting from an initial guess $\boldsymbol{V}^0$:

1. compute $c^n_{w,i,j,z}$ and $l^n_{w,j,z}$, solving the first order conditions in (36), and the saving policy $s^n_{w,i,j,z}$ given in (35). They all depend on $\boldsymbol{V}^n$. This provides us with $\boldsymbol{A}^n$ and $\boldsymbol{b}^n$.
2. Solve the linear system (A.3) for $\boldsymbol{V}^{n+1}$, and stop if $\boldsymbol{V}^{n+1}\approx\boldsymbol{V}^n$.

A.1.2 The KF equation

In the full model, the sample path of the distribution of wealth required by the algorithm is obtained by integrating the ODE (37) forward in time with a standard implicit scheme. To illustrate the discretization procedure, we instead show the numerical solution of the steady-state KF equation (32). We report it for convenience:

$$0 = -\frac{d}{dk}\left[g(k,\varepsilon_i,\beta_j)s(k,\varepsilon_i,\beta_j)\right] + \sum_{i_2\neq i}\lambda^{\varepsilon}_{i_2 i}g(k,\varepsilon_{i_2},\beta_j) + \sum_{j_2\neq j}\lambda^{\beta}_{j_2 j}g(k,\varepsilon_i,\beta_{j_2}) \tag{A.4}$$

$$-\left(\sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2} + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2}\right)g(k,\varepsilon_i,\beta_j) \tag{A.5}$$

$$1 = \sum_{\varepsilon=0,1}\sum_{\beta=1,2,3}\int g(k,\varepsilon_i,\beta_j;a_z)dk \tag{A.6}$$

$$i = 1,2 \quad j = 1,2,3 \tag{A.7}$$

We denote by $g_{w,i,j}$ the joint density at capital grid point k_w and employment-preference state (ε_i,β_j). Following Achdou, Han, Lasry, Lions, and Moll (2017), we use the an upwind finite-difference scheme which employs the following approximation

$$\frac{d}{dk}\left[g(k,\varepsilon_i,\beta_j)s(k,\varepsilon_i,\beta_j)\right] \approx \frac{g_{w,i,j}\max(s^*_{w,i,j,F},0) - g_{w-1,i,j}\max(s^*_{w-1,i,j,F},0)}{\Delta k} + \frac{g_{w+1,i,j,z}\min(s^*_{w+1,i,j,F},0) - g_{w,i,j,z}\min(s^*_{w,i,j,F},0)}{\Delta k} \tag{A.8}$$

Letting $\boldsymbol{g}$ denote the $N \times 2 \times 3 \times n_a$ column vector of densities at all capital grid points and states, we can substitute (A.8) in (A.4) to obtain a system of linear equations:

$$0 = \widetilde{\boldsymbol{A}}\boldsymbol{g} \tag{A.9}$$

$$1 = \sum_{\varepsilon=0,1}\sum_{\beta=1,2,3}\sum_{w=1}^{N} g(k_w,\varepsilon_i,\beta_j)\Delta k \tag{A.10}$$

The matrix $\widetilde{A}$ is singular, therefore this is an eigenvalue problem. Again, the sparsity of the matrix enhances the numerical tractability a great deal.

A.1.3 Marginal Propensity to Consume

The above solution procedure delivers an optimal consumption policy $c^*(k,\epsilon,\beta,a)$, for an individual with wealth k, employment status ϵ, subjective discount rate β, and conditional on aggregate log-TFP state a. The slope of this function with respect to k provides the consumption (rate) gain due to a small windfall in wealth over the infinitesimal time interval. Empirically, the relevant notion of Marginal Propensity to Consume (MPC) measures consumption gains over finite time-horizons, typically a year. As suggested in Achdou, Han, Lasry, Lions, and Moll (2017), the continuous -time counterpart of the finite MPC is:

$$mpc_\tau(k,\epsilon,\beta,a) = \frac{d}{dk}C(0,k,\epsilon,\beta,a) \tag{A.11}$$

$$C(0,k,\epsilon,\beta,a) = \mathbb{E}\left[\int_0^\tau c^*(k_t,\epsilon_t,\beta_t,a_t)dt \,\middle|\, (k_0,\epsilon_0,\beta_0,a_0) = (k,\epsilon,\beta,a)\right] \tag{A.12}$$

By Feyman-Kac theorem, the cumulative expected optimal consumption over horizon τ solves the system of partial differential equations

$$
\begin{aligned}
C_t(t,k,\varepsilon_i,\beta_j,a_z) - \left(\sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2} + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2} + \sum_{z_2\neq z}\lambda^{a}_{zz_2}\right) C(t,k,\varepsilon_i,\beta_j,a_z) + c^*(k,\varepsilon_i,\beta_j,a_z) \\
+ C_k(t,k,\varepsilon_i,\beta_j,a_z)s^*(k,\varepsilon_i,\beta_j,a_z) + \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2}C(t,k,\varepsilon_{i_2},\beta_j,a_z) + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2}C(t,k,\varepsilon_i,\beta_{j_2},a_z) \\
+ \sum_{z_2\neq z}\lambda^{a}_{zz_2}C(t,k,\varepsilon_i,\beta_j,a_{z_2}) = 0
\end{aligned}
\tag{A.13}
$$

with the terminal condition $C(\tau,k,\varepsilon_{i_2},\beta_j,a_z)=0$, where $s^*(k,\varepsilon,\beta,a)$ denotes the optimal savings (investment) function. Analogously to the HBJ equation above, we apply an upwind finite difference scheme to obtain a numerical solution of (A.13), so that:

$$
\begin{aligned}
\frac{C^{n+1}_{w,i,j,z} - C^{n}_{w,i,j,z}}{\Delta} - \left(\sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2} + \sum_{j_2\neq j}\lambda^{\beta}_{jj_2} + \sum_{z_2\neq z}\lambda^{a}_{zz_2}\right) C^{n}_{w,i,j,z} + c^*_{w,i,j,z} \\
+ \frac{C^{n}_{w+1,i,j,z} - C^{n}_{w,i,j,z}}{\Delta k}s^*_{w,i,j,z}\mathbf{1}_{s^*>0} + \frac{C^{n}_{w,i,j,z} - C^{n}_{w-1,i,j,z}}{\Delta k}s^*_{w,i,j,z}\mathbf{1}_{s^*<0} + \sum_{i_2\neq i}\lambda^{\varepsilon}_{ii_2}C^{n}_{w,i_2,j,z} \\
+ \sum_{j_2\neq j}\lambda^{\beta}_{jj_2}C^{n}_{w,i,j_2,z} + \sum_{z_2\neq z}\lambda^{a}_{zz_2}C^{n}_{w,i,j,z_2} = 0 \\
w=1,\dots,N \quad i=1,2 \quad j=1,2,3 \quad z=1,\dots,n_a
\end{aligned}
\tag{A.14}
$$

Notice that Δ has now the interpretation of a true time step, so that we have partitioned $[0,\tau]$ into a grid of nt points, with step length $\Delta = \tau/(nt-1)$. The linear system of equations (A.14) can be compactly written in matrix notation:

$$
\boldsymbol{A}\boldsymbol{C}^n = \boldsymbol{b}^{n+1} \tag{A.15}
$$

where the matrix $\boldsymbol{A}$ is the same as in (A.3), except it is time invariant here. The $N\times 2\times 3\times n_a$ column vector $\boldsymbol{b}^{n+1}$ is

$$
\boldsymbol{b}^n = \boldsymbol{c}^n + \frac{\boldsymbol{C}^{n+1}}{\Delta}
$$

Starting from the terminal condition $\boldsymbol{C}^{nt}=0$, and given $\boldsymbol{C}^{n+1}$, we can solve (A.15) for $\boldsymbol{C}^n$, and iterate to find $\boldsymbol{C}^1 \approx C(0,k,\epsilon,\beta,a)$. We then approximate the marginal propensity to consume as:

$$
mpc_\tau(k,\epsilon_i,\beta_j,a_z) = \frac{C^1_{w+1,i,j,z} - C^1_{w,i,j,z}}{\Delta k}
$$

A.2 Estimation Procedure

In this appendix we provide details of the combined estimation and calibration procedure outlined in Section 5 of the main text.

In the deterministic steady-state – with log-TFP $a_t = 0$ and no idiosyncratic employment shock – we obtain:

$$\begin{aligned}
\bar{r} &= \bar{\beta} + \delta \\
\bar{w} &= (1-\alpha)(\bar{r}/\alpha)^{\frac{\alpha}{\alpha-1}} \\
\bar{k} &= (\bar{r}/\alpha)^{\frac{1}{\alpha-1}} \\
\bar{c} &= \bar{k}^{\alpha} - \delta\bar{k} \\
\bar{p}^{*} &= \frac{1}{1-\frac{1}{\theta(1+\gamma\bar{c}^{1/\theta})}}
\end{aligned}$$

where we set $\bar{\beta}$ to the unconditional mean of the process β_t obtained from its steady-state distribution, after the estimation procedure detailed below. In order to have $\bar{l} = 1$, we impose the following value to the disutility of labor parameter:

$$\nu = \frac{\bar{w}}{\bar{p}^{*}} \frac{\gamma + \bar{c}^{-1/\theta}}{\gamma\bar{c} + \frac{\bar{c}^{1-1/\theta}}{1-1/\theta}} \tag{A.16}$$

We calibrate the intensity matrix Λ^{ε} of the binary employment state ε_t to match the discrete-time, yearly transition matrix reported in Krusell and Smith (1998).[21] This implies:

$$\exp(\Lambda^{\varepsilon}) = \begin{bmatrix} 0.4708 & 0.5292 \\ 0.0372 & 0.9628 \end{bmatrix} \quad \text{or} \quad \Lambda^{\varepsilon} = \log\left(\begin{bmatrix} 0.4708 & 0.5292 \\ 0.0372 & 0.9628 \end{bmatrix}\right)$$

where $\exp(\cdot)$ is intended as matrix exponential:

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!},$$

and the matrix logarithm $\log(\cdot)$ being its inverse operator.

Given the autoregressive (ρ) and volatility (σ_a) parameters of AR-1 process (43) for log-TFP – estimated on the time series of 'Total Factor Productivity at Constant National Prices for United States', obtained from the FRED database maintained by St. Louis FED – we apply the Rouwenhorst (1995) discretization procedure to identify the states and the transition probability matrix P^a of an approximating discrete-time Markov chain at yearly frequency. As seen above, for the transition probability of the continuous-time Markov chain to match this probability at a 1 year horizon, we must set:

$$\Lambda^a = \log(P^a)$$

The parameters which have not been calibrated are then collected in the 9-dimensional vector

$$\Omega = (\lambda_{12}^{\beta}, \lambda_{21}^{\beta}, \lambda_{23}^{\beta}, \lambda_{32}^{\beta}, \rho_a, \sigma_a, \varphi, \gamma, \theta),$$

[21] In that reference, a joint Markov chain for the unemployment and the business cycle state is provided. We use the implied marginal probabilities.

which we estimate with the Simulated Method of Moments of Duffie and Singleton (1993) and Lee and Ingram (1991). We first select a set of 11 moments of aggregate consumption, output, investment and labor supply, namely quarterly variances and one-quarter autocovariances of each variable, and the covariance between output and the other three. We obtain quarterly series from 1947 to 2017 on real GDP per capita, real non-durable and services consumption expenditure per capita, real total private fixed investment (including durable consumption) per capita, and total hours in the non-farm business sector, all of which from the BAE or the BLS, provided by the FRED database. We augment this set with two moments of the wealth distribution extracted from the US *Survey of Consumer Finances*: the fraction of total wealth owned by the wealthiest 5% and the poorest 5% of the population. We denote by $\mathbb{E}[M(x)]$ the 13-dimensional vector of real unconditional moments computed on these (HP-filtered) series, in which x is an i.i.d. data sample.[22] Given a parameter set Ω, we compute the model-implied counterpart of these unconditional moments by Monte-Carlo simulation. This is achieved in the following steps:

1. We solve for the equilibrium using the iterative procedure described in Section 4. This leaves us with: i) an equilibrium joint density of wealth, employment status and preference state, conditional on the aggregate TFP state, $g(k, \varepsilon, \beta; a)$. ii) Equilibrium consumption, labor, and saving policy functions, $c(k, \varepsilon, \beta, a)$, $l(k, 1, \beta, a)$, $s(k, \varepsilon, \beta, a)$. Aggregate policies contingent on the TFP state are then obtained. For instance:

$$C(a_z) = \sum_{i=1,2} \sum_{j=1,2,3} \int c(k, \varepsilon_i, \beta_j, a_z)\, g(k, \varepsilon_i, \beta_j; a_z)\, dk \tag{A.17}$$

2. We simulate at high frequency a very long trajectory for a_t,[23] using standard simulation methods for Continuous-Time Markov Chains, from which we obtain quarterly simulated series of all the aggregate policies, by evaluating equilibrium policies (A.17) at the simulated log-TFP series. For instance:

$$C_t^{(0.25)} = \int_{t-0.25}^{t} C(\tilde{a}_s) ds \approx \frac{1}{n_s - 1} \sum_{i=1}^{\lceil n_s \times 0.25 \rceil} C(\tilde{a}_{t - \lceil n_s \times 0.25 \rceil + i})$$

   where $\tilde{a}_t$ denotes the simulated log-TFP trajectory and n_s is the number of discretization time points per year. Let $\mathbb{E}[m(\Omega, \tilde{a})]$ denote the 13-dimensional vector of model-implied unconditional moments computed on these simulated series, where we have emphasized the dependence on the parameter vector. It is worthwhile to point out that the two moments pertaining the distribution of wealth – fraction of wealth owned by wealthiest and poorest 5% of population – have been obtained as:

$$\mathbb{E}\left[WH(a_t)\right] \qquad \text{and} \qquad \mathbb{E}\left[WL(a_t)\right] \tag{A.18}$$

[22] Concerning the two moments of the wealth distribution, we treat values extracted from the 2004 US *Survey of Consumer Finances* as unconditional moments, as in Carroll et al. (2017).

[23] 50000 years, with 1000 time points per year. Results do not change by increasing any of these two quantities.

where (denoting by k_5 and k_{95} the 5 and 95 percentiles of the wealth distribution)

$$WL(a_t) = \frac{\sum_{i=1,2}\sum_{j=1,2,3}\int_{\underline{k}}^{k_5} k\, g(k,\varepsilon_i,\beta_j;a_t)\, dk}{\sum_{i=1,2}\sum_{j=1,2,3}\int_{\underline{k}}^{\infty} k\, g(k,\varepsilon_i,\beta_j;a_t)\, dk} \tag{A.19}$$

$$WH(a_t) = \frac{\sum_{i=1,2}\sum_{j=1,2,3}\int_{k_{95}}^{\infty} k\, g(k,\varepsilon_i,\beta_j;a_t)\, dk}{\sum_{i=1,2}\sum_{j=1,2,3}\int_{\underline{k}}^{\infty} k\, g(k,\varepsilon_i,\beta_j;a_t)\, dk} \tag{A.20}$$

The Simulated Method of Moment estimator is:

$$\widehat{\Omega} = arg\min_{\Omega} \left(\mathbb{E}[m(\Omega,\tilde{a})] - \mathbb{E}[M(x)]\right)' \; W \; \left(\mathbb{E}[m(\Omega,\tilde{a})] - \mathbb{E}[M(x)]\right) \tag{A.21}$$

where the weighting matrix W is the inverse of the covariance matrix of the empirical moments $M(x)$, computed with the Newey and West (1987) estimator. We evaluate the standard errors of parameter estimates using the SMM asymptotic theory:

$$\sqrt{N_D}(\widehat{\Omega}-\Omega) \xrightarrow{D} N(0,V), \quad V = \left(1+\frac{1}{N_S}\right)\left(\frac{\partial\mathbb{E}[m(\Omega,\tilde{a})]'}{\partial\Omega} W \frac{\partial\mathbb{E}[m(\Omega,\tilde{a})]}{\partial\Omega}\right)^{-1} \tag{A.22}$$

where N_D is the number of true data points, N_S the number of simulated data points, and $\frac{\partial\mathbb{E}[m(\Omega,\tilde{a})]}{\partial\Omega}$ is the Jacobian matrix of the vector of model-implied moments, which we compute by finite difference. In order to test the goodness-of-fit, we test the equality of each individual moment. In particular, the t-statistics on the individual moment conditions are calculated as pricing errors in a standard GMM framework, using the asymptotic result:

$$\sqrt{N_D}\left(\mathbb{E}[m(\Omega,\tilde{a})] - \mathbb{E}[M(x)]\right) \xrightarrow{D} N\left(0, W^{-1} - \frac{\partial\mathbb{E}[m(\Omega,\tilde{a})]'}{\partial\Omega} V \frac{\partial\mathbb{E}[m(\Omega,\tilde{a})]}{\partial\Omega}\right) \tag{A.23}$$

A.3 Impulse Responses

Given an aggregate equilibrium policy $X(a_t)$ – such as the aggregate consumption displayed in (A.17), which follows a continuous-time markov chain with Λ^a transition matrix – its impulse response at horizon τ following a positive one-standard deviation log-TFP shock, is defined as:[24]

$$IR_\tau(X) = \sum_{i=1}^{n_a}\mathbb{P}(a_i)\left(\frac{1}{\tau}\mathbb{E}\left[\int_t^{t+\tau} X(a_s)ds \middle| a_t = a_{i+}\right] - \frac{1}{\tau}\mathbb{E}\left[\int_t^{t+\tau} X(a_s)ds \middle| a_t = a_i\right]\right) \tag{A.24}$$

where a_{i+} denotes the log-TFP state that best approximates $a_i + \sigma_a/\sqrt{1-\rho_a^2}$ and $\mathbb{P}(a_i)$ denotes the invariant probability distribution of the continuous-time markov chain for a_t. In other words, the impulse response is the incremental (normalized) cumulative expectation of the policy following a jump in log-TFP. Since we average over all possible states that might be perturbed, using the invariant probability, we are dealing with an unconditional value.

Expression (A.24) can be computed in close-form. By Feyman-Kac theorem, the expecta-

[24]One standard deviation of the stationary distribution ot the approximated AR-1.

tion $J(t,\tau,a_i) = \mathbb{E}\left[\int_t^{t+\tau} X(a_s)ds \middle| a_t = a_i\right]$ satisfies the ODE

$$\frac{\partial J}{\partial t} + \sum_{j\neq i} \lambda_{ij}^a J(t,\tau,a_j) - J(t,\tau,a_i)\sum_{j\neq i}\lambda_{ij}^a + X(a_i) = 0$$

with the terminal condition $J(\tau,\tau,a_i) = 0$. The solution can easily be expressed in terms of matrix exponential as follows:

$$J(t,\tau,a_i) = e_{n_a}(i)' e^{-\Lambda^a \tau} \int_0^\tau e^{\Lambda^a s} \boldsymbol{X}\, ds \tag{A.25}$$

where $\boldsymbol{X}$ is the n_a-dimensional column vector of all conditional policy values, and $e_{n_a}(i)$ is a n_a-dimensional column vector of zeros, except 1 in the $i-$th entry.

We also provide impulse responses for the equilibrium mark-up and the percentage of wealth in possession of the top or bottom 5-th percentile of the wealth distribution. With this choice of X, the impulse response is defined as

$$\sum_{i=1}^{n_a} \mathbb{P}(a_i)\left(\mathbb{E}\left[X(a_{t+\tau})|a_t = a_{i+}\right] - \mathbb{E}\left[X(a_{t+\tau})|a_t = a_i\right]\right) \tag{A.26}$$

Solving the appropriate Feyman-Kac equation, similarly to above, the expectations appearing in this expression are given explicitly by:

$$\mathbb{E}\left[X(a_{t+\tau})|a_t = a_i\right] = e_{n_a}(i)' e^{-\Lambda^a \tau} \boldsymbol{X} \tag{A.27}$$

The percentage of total wealth in possession of the top and bottom 5-th percentile of the wealth distribution are, respectively:

$$PW_{95}(a_t) = \frac{\sum_{i=1,2}\sum_{j=1,2,3}\int_{\overline{\alpha}(a_t)}^{\infty} k\, g(k,\varepsilon_i,\beta_j;a_t)\, dk}{K(a_t)} \tag{A.28}$$

$$PW_{5}(a_t) = \frac{\sum_{i=1,2}\sum_{j=1,2,3}\int_0^{\underline{\alpha}(a_t)} k\, g(k,\varepsilon_i,\beta_j;a_t)\, dk}{K(a_t)} \tag{A.29}$$

$$K(a_t) = \sum_{i=1,2}\sum_{j=1,2,3}\int k\, g(k,\varepsilon_i,\beta_j;a_t)\, dk \tag{A.30}$$

where $\overline{\alpha}(a_t)$ and $\underline{\alpha}(a_t)$ are the 95 and 5 percentiles of the conditional (on log-TFP) wealth distribution $\sum_{i=1,2}\sum_{j=1,2,3}\int g(k,\varepsilon_i,\beta_j;a_t)$.

The confidence bands of the impulse response functions are computed by Monte-Carlo simulation, sampling from the asymptotic distribution of parameter estimates. Namely:

1. We extract 5000 *iid* parameter samples $\widetilde{\Omega}$ from the asymptotic distribution $N(\widehat{\Omega}, V)$, where V is given in (A.22).

2. For each sample $\widetilde{\Omega}$, we solve for the equilibrium using the iterative procedure described in Section 4, and compute the corresponding impulse response $\widetilde{IR}_\tau(X)$ as detailed above.

3. The confidence interval is given by the 5 and 95 percentile of the empirical distribution of the $\widetilde{IR}_\tau(X)$.

B Representative-Agent Model

In this Appendix we discuss the symmetric-good, homogeneous-households version of the model, which we label 'Representative-Agent' (RA) model for brevity, a terminology justified by the observation that households do not face idiosyncratic employment or preference shocks anymore, thus being *de-facto* identical. The remaining elements of the model are unaltered, including the dynamics of log-TFP a_t and agents' IES preferences. In the end, we are discussing the continuous-time analog of the framework proposed by Cavallari and Etro (2020).

Letting p_t denote the price of the final good,[25] and $\Gamma_t = (r_t, w_t, p_t)$ the vector of endogenous state-variables, the RA solve the following consumption-savings program:

$$V(K_t, \Gamma_t) = \sup_{L_t \geq 0, C_t} \mathbb{E}\left[\int_t^\infty e^{-\beta(s-t)} \left(\log\left(u(C_s)\right) - v\frac{L_s^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}\right) ds\right] \tag{B.1}$$

subject to the budget constraint:

$$dK_t = \left[K_t(r_t - \delta) + w_t L_t - p_t C_t + \Pi_t\right] dt \tag{B.2}$$
$$K_t \geq \underline{K} \tag{B.3}$$

The notation has already been established in the text, except for the constant time preference rate β. Notice that the final goods sector's profits Π_t are rebated to the household. Mimicking the reasoning of Section 2.3, and taking into account the homogeneity of households, the equilibrium price of the final good reads:

$$p_t^* = \frac{1}{1-\vartheta_t}, \qquad \vartheta_t = -\frac{C_t u''(C_t)}{u'(C_t)} \tag{B.4}$$

where C_t is the optimal consumption policy. The household solves program (B.1) taking the dynamics of the endogenous variables Γ_t and Π_t as given. Since, in equilibrium, the interest rate and wage are determined by their marginal productivity,

$$w_t = (1-\alpha)e^{a_t}(K_t/L_t)^\alpha \tag{B.5}$$
$$r_t = \alpha e^{a_t}(K_t/L_t)^{\alpha-1}, \tag{B.6}$$

the two relevant state variables for the dynamic program are capital and log-TFP. Therefore, denoting by $V(K, a_z)$ the household's value function, for $z = 1, \ldots, n_a$, the latter solves the HBJ equation

$$\left(\beta + \sum_{z_2 \neq z} \lambda^a_{zz_2}\right) V(K, a_z) = \log u(C_t^*) - v\frac{(L_t^*)^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}} + V_K(K, a_z)s(K, a_z) + \sum_{z_2 \neq z} \lambda^a_{zz_2} V(K, a_{z_2}) \tag{B.7}$$

[25] We remind that we confine ourselves to a symmetric equilibrium

where we have used the optimal saving policy:

$$s(K,a_z) = (r_t-\delta)K + w_t L_t^* - p_t^* C_t^* + \Pi_t \tag{B.8}$$
$$= e^{a_z} K_t^\alpha (L_t^*)^{1-\alpha} - \delta K_t - C_t^* \tag{B.9}$$

Notice that in the last expression we have used the equilibrium expressions of r_t, w_t and the symmetric equilibrium total profits $\Pi_t = (p_t^* - 1)C_t^*$. Moreover, optimal policies C_t^* and L_t^* are derived from the first order conditions

$$\frac{u'(C_t)}{u(C_t)} = V_K(K,a_z)p_t^*, \qquad L_t = \left(\frac{w_t V_K(K,a_z)}{v}\right)^\varphi \tag{B.10}$$

with $u(\cdot)$ as in (13). Notice that p_t^* is itself a function of C_t, and w_t is a function of L_t.

In the numerical solution of the model, we noticed that some care must be taken in order to insure a robust convergence to equilibrium, in accordance with the assumption that the agent takes as exogenously given factor prices (w_t and r_t), good prices (p_t^*) and redistributed profits Π_t, when solving the consumption-saving problem. The algorithm is as follows:

1. At step 0, guess generic functions $r_0(K,a_z)$, $w_0(K,a_z)$, $p_0(K,a_z)$, and $\Pi_0(K,a_z)$
2. At step n, solve the HBJ equation (B.7) numerically using a finite-difference upwind scheme with time iteration, using as input $r_{n-1}(K,a_z)$, $w_{n-1}(K,a_z)$, $p_{n-1}(K,a_z)$, and $\Pi_{n-1}(K,a_z)$ The procedure is a straightforward adaptation of that outlined in Appendix A.1. Notice that (B.8) must be used for the savings function, and not (B.9). Let $C_n(K,a_z)$ and $L_n(K,a_z)$ denote the optimal consumption and labor policy just obtained, and use them in expressions (B.4), (B.5) and (B.6) to define $\tilde{p}_n(K,a_z)$, $\tilde{w}_n(K,a_z)$ and $\tilde{r}_n(K,a_z)$, respectively. Also, set $\tilde{\Pi}_n(K,a_z) = (\tilde{p}_n - 1)C_n$.
3. Set
$$x_n(K,a_z) = \omega x_{n-1}(K,a_z) + (1-\omega)\tilde{x}_n(K,a_z)$$
where x is any of p^*, w, r and Π.[26]
4. Iterate until $x_{n+1} \approx x_n$ (according to some norm-based convergence criteria) point-wise for every K and a_z.

The parameter set we use in every numerical exercise is obtained with a combined calibration and estimation procedure very similar to the heterogeneous-agents', described in Appendix A.2. In the absence of employment and preference shocks, the parameter vector to be estimated by SMM is $\Omega = (\rho_a, \sigma_a, \varphi, \gamma, \sigma)$. Clearly, the set of moment conditions now excludes those pertaining the distribution of wealth. Notice that the SMM estimation now entails simulating also a capital trajectory by Euler discretization.

Similarly, impulse response functions are also obtained as in Appendix A.3, with the exception that expectations of integrated future policy values are obtained by Feyman-Kac theorem, solving by finite-difference a system of linear PDEs identical to A.14, excluding time-preference and employment state variables, and substituting consumption with any policy we might be considering.

[26] ω close to 1 guarantees convergence, albeit slowly.

Given that estimation and impulse-responses are obtained by adapting familiar procedures, we omit further details.

In Figure 2, we also report the stationary density of capital conditional on the lowest or highest aggregate log-tfp states, obtained using SMM parameter estimates. It is easy to show that this density satisfy the following Kolmogorov Forward equation:

$$0 = -\frac{d}{dk}\left[g(k;a_z)s(k,a_z)\right] + \sum_{z_2 \neq z} \lambda^a_{zz_2} g(k;a_{z_2}) - \sum_{z_2 \neq z} \lambda^a_{zz_2} g(k;a_z) \tag{B.11}$$

$$1 = \int g(k;a_z)dk \qquad z = 1,\dots,n_a \tag{B.12}$$

Again, we solve this system of differential equations numerically by adapting the approach described in Appendix A.1.2, hence we omit further details.

C Additional Material

C.1 Martingale Solution of Household Consumption-Saving Problem

In Section 3.1 we mentioned the possibility of solving the individual consumption-investment approach using martingale methods. This can be accomplished with the methodology first developed in He and Pagès (1993), who extend the martingale approach of Cox and Huang (1989) to the case of labor income and nonnegative wealth constraints. The static formulation of problem (5) reads:

$$V(k_t,\varepsilon_t,\beta_t,\Gamma_t) = \sup_{l_t \geq 0, c_t} \mathbb{E}_t\left[\int_t^\infty B_{t,s}\left(\log u(c_s) - v\,\frac{l_s^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}\right) ds\right] \tag{C.1}$$

$$\mathbb{E}_t\left[\int_t^\infty \xi_{t,s}\, p_s c_s\right] \leq k_t + \mathbb{E}_t\left[\int_t^\infty \xi_{t,s}\, w_s l_s \varepsilon_s\right] \tag{C.2}$$

where the subjective discount factor and the state-price density are, respectively

$$\begin{aligned} B_{t,s} &= e^{-\int_t^s \beta_u\, du} \quad \Rightarrow \quad dB_{t,s} = -B_{t,s}\beta_t dt \\ \xi_{t,s} &= e^{-\int_t^s (r_u - \delta)\, du} \quad \Rightarrow \quad d\xi_{t,s} = -\xi_{t,s}(r_t - \delta)dt \end{aligned}$$

and the state constraint is $k_t \geq \underline{k}$.

He and Pagès (1993) characterize the solution to this problem as follows:

$$V(k_t,\varepsilon_t,\beta_t,\Gamma_t) = \inf_{X_t} \sup_{l_t \geq 0, c_t} \mathbb{E}_t\left[\int_t^\infty B_{t,s}\left(\log u(c_s) - v\,\frac{l_s^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}\right) ds\right] \tag{C.3}$$

$$\mathbb{E}_t\left[\int_t^\infty \xi_{t,s} X_s\, p_s c_s\, ds\right] \leq k_t + \mathbb{E}_t\left[\int_t^\infty \xi_{t,s} X_s\, w_s l_s \varepsilon_s\, ds\right] \tag{C.4}$$

where X_t is a nonnegative and nonincreasing process interpreted as lagrange multiplier for the infinite sequence of borrowing constraints. Define $\zeta_{t,s} = \xi_{t,s} X_s$. Applying Lagrangian

Theory, consider the convex conjugate:

$$\tilde{u}(\zeta_{t,s}/B_{t,s};\varepsilon_s) = \sup_{l_s\geq 0, c_s} \left[B_{t,s}\left(\log u(c_s) - v\frac{l_s^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}\right) - \zeta_{t,s}(p_s c_s - w_s l_s \varepsilon_s)\right] \tag{C.5}$$

$$= \begin{cases} B_{t,s}\log u(f(p_s\zeta_{t,s}/B_{t,s})) - \zeta_{t,s}p_s f(p_s\zeta_{t,s}/B_{t,s}) \\ +\frac{1/\varphi}{1+1/\varphi}(\zeta_{t,s}w_s)^{1+\varphi}(B_{t,s}v)^{-\varphi} & \text{if} \quad \varepsilon = 1 \\ B_{t,s}f(p_s\zeta_{t,s}/B_{t,s}) - \zeta_{t,s}p_s f(p_s\zeta_{t,s}/B_{t,s}) & \text{if} \quad \varepsilon = 0 \end{cases} \tag{C.6}$$

where $f(\cdot)$ is the inverse function of $u'(\cdot)/u(\cdot)$. Problem (C.3) then admits the dual formulation

$$\widetilde{V}(\zeta;\varepsilon_t,\beta_t) = \inf_{X_t}\mathbb{E}_t\left[\int_t^\infty \tilde{u}(\zeta_{t,s}/B_{t,s};\varepsilon_s)\,ds\right] \tag{C.7}$$

Assuming that X_t is absolutely continuous, so that $dX_t = -X_t\psi_t dt$, for $\psi_t \geq 0$, and

$$d\zeta_{t,s} = -\zeta_{t,s}(r_t - \delta + \psi_t)dt,$$

the Hamilton-Bellman-Jacobi equation satisfied by the dual value function $\widetilde{V}$ is:

$$\inf_{\psi}\left[\tilde{u}(\zeta/B_{t,s};\varepsilon_s) - \widetilde{V}_z(\zeta;\varepsilon_s,\beta_s)\zeta\psi + \mathcal{A}\widetilde{V}(\zeta;\varepsilon_s,\beta_s)\right] = 0 \tag{C.8}$$

where $\mathcal{A}\cdot$ is the infinitesimal generator of the processes $(\zeta_t,\varepsilon_t,\beta_t)$ (excluding the $\psi-$term in the drift of ζ). This is indeed a singular control problem, which can be alternatively characterized as the (system of) variational inequality:

$$\min\left(\tilde{u}(\zeta/B_{t,s};\varepsilon_s) + \mathcal{A}\widetilde{V}(\zeta;\varepsilon_s,\beta_s), -\widetilde{V}_z(\zeta;\varepsilon_s,\beta_s)\right) = 0 \tag{C.9}$$

In our framework this approach does not yield a fully explicit characterization, since we would still need to solve numerically for the free-boundary of the ζ state variable, where the state constraint $k_t \geq \underline{k}$ becomes binding and the state variable ζ is reflected in the continuation region. In other contexts though, such a CES subutility, the outcome might be different.