Inequality, Business Cycles, and Monetary-Fiscal Policy*

Anmol Bhandari U of Minnesota

David Evans U of Oregon

Mikhail Golosov U of Chicago

Thomas J. Sargent NYU

March 15, 2021

Abstract

We study optimal monetary and fiscal policies in a New Keynesian model with heterogeneous agents, incomplete markets, and nominal rigidities. Our approach uses small-noise expansions and Fréchet derivatives to approximate equilibria quickly and efficiently. Responses of optimal policies to aggregate shocks differ qualitatively from what they would be in a corresponding representative agent economy and are an order of magnitude larger. A motive to provide insurance that arises from heterogeneity and incomplete markets outweighs price stabilization motives.

KEY WORDS: Sticky prices, heterogeneity, business cycles, monetary policy, fiscal policy

*We thank the editor and referees as well as Adrien Auclert, Benjamin Moll, and participants at numerous seminars and conferences for helpful criticisms.

1 Introduction

We approximate recursive representations of optimal monetary and fiscal policies in an incomplete markets economy with agents who face both aggregate and idiosyncratic risks. Agents differ in wages, exposures to aggregate shocks, holdings of financial assets, and abilities to trade assets. They cannot fully insure themselves because financial markets are incomplete. Firms are monopolistically competitive. Price adjustments are costly. We examine how a Ramsey planner's policies for nominal interest rates, transfers, and flat-rate taxes on labor earnings, dividends, and interest income respond to aggregate shocks.

It is challenging to approximate a Ramsey plan in our setting. The aggregate state in a recursive formulation of the Ramsey problem includes the joint distribution of individual asset holdings and auxiliary promise-keeping variables chosen earlier by a planner. The law of motion for that high-dimensional object must be determined jointly with the optimal policies, and the distributions along the transition path differ substantially from the invariant distribution without aggregate shocks. These aspects render inapplicable common computational strategies that approximate policy functions after summarizing cross-sectional distributions with a small number of moments or that linearize policy functions around some time-invariant distribution.

We forge a new computational approach that can be applied to economies with substantial heterogeneity and that does not require knowing long-run properties in advance. Our approach builds on a perturbation theory that constructs a sequence of small-noise expansions with respect to a one-dimensional parameterization of uncertainty along simulations of sample paths of our economy. The procedure is recursive and unfolds over time. At each time period along a simulated sample path, we approximate policy functions by applying a perturbation algorithm at that period's cross-section distribution. We use these approximate decision rules for that period to determine outcomes that include government policy decisions and the cross-section distribution next period. Then we move forward one period and perturb around next period's cross-section distribution to approximate next period's government decision rules and other outcomes. In this way, along an equilibrium sample path we sequentially update cross-section distributions around which we approximate policy functions.

Our perturbation approach requires that each period we compute derivatives of policy functions with respect to all state variables, one of which is a Fréchet derivative with respect to a distribution over a multi-dimensional vector of agents' characteristics. It is usually hard to compute this Fréchet derivative directly. What makes our approach practical is that, in an interesting class of heterogeneous agent competitive equilibrium models, we can streamline this computationally-intensive step by representing parameters of approximate policy functions as a collection of low-dimensional linear equations that are independent of each other. This is possible because in standard competitive environments, conditional on prices and aggregate quantities, different agents' optimal choices can be solved separately. These systems of linear equations are independent across agents and thus easily parallelizable. This helps us manage the ample heterogeneity present in our model. Agents' optimal choices can then be aggregated into a set of equilibrium conditions that lead to a low-dimensional fixed point problem whose solution determines prices and aggregate quantities. A similar computationally convenient linear structure prevails for second- and higher-order expansions, making our approach applicable to many optimal policy problems in which aggregate risks have important effects on equilibrium dynamics.

We apply our approach to a textbook sticky price model (see, e.g., Galí, 2015) augmented with heterogeneous agents. In the tradition of Bewley (1977, 1980), Huggett (1993), and Aiyagari (1994), financial markets are incomplete: agents can trade only non-state-contingent nominal debt. Agents' wages are subject to idiosyncratic and aggregate shocks that we calibrate to match U.S. business cycles and cross-sectional properties of labor earnings. We set the initial joint distribution of nominal and real claims, and wages to match cross-sectional moments in the Survey of Consumer Finances. We posit two aggregate shocks: one to productivity, and another to the elasticity of substitution between differentiated intermediate goods that affects firms' optimal markups. We pose two Ramsey problems. In the first, a "purely monetary policy" planner can adjust only nominal interest rates and transfers in response to shocks, while keeping all other tax rates at fixed levels chosen in period 0. This is a common assumption used in New Keynesian models. In addition to interest rates and transfers, our second Ramsey planner has more tools and can adjust tax rates on all sources of income. Since standard calibrations of Bewley-Aiygari economies imply a slow drift towards a long-run distribution with much smaller asset heterogeneity than observed in the U.S. data, we focus our attention on optimal policy responses in the initial 100 periods, when the cross-sectional distribution of earnings and assets is similar to the initial one.

We find that inflation, nominal interest rates, and taxes are substantially more volatile in our calibrated heterogeneous agent (HANK) economy than in its representative agent (RANK) counterpart. For example, the standard deviation of inflation chosen by the "purely monetary policy" planner is an order of magnitude higher in HANK. Moreover, the magnitudes and signs of correlations between nominal and real variables differ in HANK and RANK. We run diagnostics that show that providing insurance against aggregate shocks to heterogeneous agents accounts for most of the differences between optimal monetary and fiscal policies in the two economies.

To understand how insurance concerns shape optimal policies, consider the optimal monetary response to a one-time positive markup shock that motivates firms to increase their prices. A standard optimal response in New Keynesian models is to stabilize the price level. The planner increases nominal interest rates and in that way decreases firms' marginal costs by lowering aggregate demand. This response rationalizes the prescription to "lean against the wind" by raising interest rates when firms' desired markups increase (Galí, 2015). The markup shock also changes relative shares of payments to labor and owners of equity. When firm owners and wage earners are the same people, such movements in factor shares have no welfare consequences, but they can have adverse risk-sharing consequences if different agents have different sources of incomes. When agents are heterogeneous and cannot trade Arrow securities, the Ramsey planner can use monetary and fiscal policy to compensate for missing insurance markets. Since a positive markup shock creates an unexpected drop in wage income and a rise in profits, a Ramsey planner can provide insurance payments to workers by lowering nominal rates to boost wages. A negative markup shock makes the planner want to synthesize insurance payouts to equity owners, making the optimal response a mirror image of the response to a positive markup shock.

Quantitatively, the strength of the insurance motive depends on the correlation between labor and capital incomes: less positive correlations call for more insurance. That the distribution of stock ownership is much more skewed to the right than is the distribution of labor earnings implies that there are potentially large welfare gains from supplying insurance. As a result, in our calibrated economy optimal monetary responses to markup shocks are an order of magnitude larger and opposite in direction from those in a corresponding representative agent economy.

An insurance motive also shapes a Ramsey planner's responses to TFP shocks. While TFP shocks push profits and wages in the same direction, the consequences of TFP shocks are not shared equally by borrowers and lenders. When agents can trade only non-state-contingent bonds, a TFP shock changes total output while keeping nominal obligations unchanged. So a negative TFP shock hurts borrowers while a positive TFP shock hurts lenders. A Ramsey planner can provide insurance and improve welfare by lowering (raising) the real return on debt in response to a negative (positive) TFP shock. That optimal response contrasts with the standard New Keynesian prescription of adjusting the nominal interest rate one-for-one with the "natural" rate of interest (i.e., the interest rate that would prevail if nominal prices were perfectly flexible). That we observe substantial dispersion in ownership of nominal claims indicates that the planner's insurance motive is strong in our calibrated economy.

We also consider a number of extensions and robustness checks to explore the implications of different assumptions about the redistributive objective of the planner, price stickiness, asset trading frictions, and heterogeneity in marginal propensities to consume on optimal policies. In all cases, that we considered, insurance considerations account for most differences between optimal policies in HANK and RANK economies.

1.1 Related literature

Our paper contributes to two literatures: one that approximates equilibria of incomplete markets economies with heterogeneous agents, and another that computes Ramsey plans for fiscal and monetary policies.

We compute small-noise expansions around transition paths like those deployed by Fleming (1971), Fleming and Souganidis (1986), and Anderson et al. (2012), all of whom study problems with state vectors that are much smaller than ours. That makes direct applications of approaches in those papers computationally impractical for us. Instead, we use functional derivatives techniques[1] to cope with the most computationally intensive step and reformulate the problem of computing approximate policy functions as a manageable collection of low-dimensional linear equations. Our techniques can be used to construct second- and higher-order approximations via a convenient set of recursions.

Relative to popular approaches such as Krusell and Smith (1998) or Reiter (2009), our method brings benefits and costs. In the Krusell and Smith approach, one summarizes the distribution of agents' characteristics with a small number of moments and approximates the law of motion of those moments. In contrast, our method allows for complicated multidimensional distributions that are hard to summarize with few parameters but relies on local expansions with respect to shocks. In Reiter's approach, while policy functions are accurate with respect to idiosyncratic shocks, they are approximated only to the first order with respect to aggregate shocks around the invariant distribution of the no-aggregate-shock economy. In comparison, our method is not constrained to linear approximations, and we repeatedly update the approximation as the state of the economy moves along an equilibrium path. Our method is particularly well suited for economies with possibly non-stationary transition dynamics or when impulse responses depend on past shocks or when higher-order moments of aggregate variables play important roles. Despite its accuracy in approximating optimal responses to aggregate shocks, our method is less accurate in approximating the dependence of optimal policies on idiosyncratic shocks.

We assess the numerical accuracy of our method by computing the competitive equilibrium for a special case of our model that corresponds to an economy studied by Acharya and Dogra (2018) under fixed government policies. An advantage of studying their economy is that we can analytically compute an equilibrium. This allows us to compare both

[1] Childers (2018) combines related functional derivative techniques with a Reiter (2009) method, but unlike our approach, his still requires that distributions remain close to the invariant distribution of a no-aggregate-shock economy. We build on and extend Evans (2015).

our approximation and Reiter's to a pencil-and-paper answer. Under Acharya and Dogra's calibration, maximum numerical errors in our approximated policy functions are less than 0.05%. Moreover, approximation errors in policy functions of aggregate variables are *two orders of magnitude* smaller than those obtained with Reiter's method. While our method is less accurate in approximating responses to idiosyncratic shocks, those errors average out in the aggregate. (However, second-order errors in approximating responses to aggregate shocks under Reiter's method do not average out). We also show that because errors in approximating responses to aggregate shocks compound over time under Reiter's approach, they adversely affect approximating long-run distributions. Finally, we also illustrate how the drift of the distribution of assets away from the point of approximation leads to errors in impulse responses under a Reiter method that do not emerge with our method.

A substantial literature on optimal monetary and fiscal policies in the Ramsey tradition has mostly studied economies with few if any sources of heterogeneity. For treatments of optimal monetary policies in representative agent New Keynesian models, see Galí (2015) and Woodford (2003).[2] Bilbiie and Ragot (2017), Challe (2017), Bilbiie (2019), and Debortoli and Galí (2017) study optimal monetary policy in economies with limited heterogeneity and in which a cross-sectional distribution disappears from the formulation of a Ramsey problem and analysis can be done using traditional techniques. Like us, those papers emphasize that uninsurable aggregate shocks create reasons for the planner to sacrifice price stability. Notable recent papers by Nuno and Thomas (2016), LeGrand and Ragot (2017) and LeGrand et al. (2020) develop alternative methods to approximate Ramsey allocations in incomplete market economies with heterogeneity. Nuno and Thomas (2016) study dynamics of a Ramsey allocation in a small open economy using continuous time methods. LeGrand and Ragot (2017) and LeGrand et al. (2020) truncate idiosyncratic histories and then linearize with respect to aggregate shocks. Their approximations to optimal policies for a standard Aiyagari-Bewley incomplete markets model with nominal rigidities indicate that inflation contributes little to shaping an optimal allocation in the long-run. However, standard calibrations of Aiyagari-Bewley economies like theirs display a much smaller dispersion in the distribution of wealth than those presented in the U.S. data to which we calibrate our model. That means that there are fewer gains that a Ramsey planner can reap from providing insurance using state-contingent movements in inflation in a long-run steady state. More broadly, the scope for insurance depends crucially on the joint distribution of earnings, nominal and real assets. It is important to study optimal monetary policy in models that closely match those distributions in the data because they pin down the heterogeneity in

[2] There are several papers that study optimal monetary and fiscal policies in calibrated representative agent settings. For instance, see Chari and Kehoe (1999) for a neoclassical setup, and see Schmitt-Grohe and Uribe (2004a) and Siu (2004) for optimal responses to government spending shocks in setups with nominal rigidities.

the unhedged exposures to aggregate shocks.

2 Environment

There is a continuum of infinitely lived households indexed by $i \in [0,1]$. Individual i's preferences over final consumption good $\{c_{i,t}\}_t$ and hours $\{n_{i,t}\}_t$ are ordered by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u\left(c_{i,t}, n_{i,t}\right), \tag{1}$$

where $\mathbb{E}_t$ is a mathematical expectation operator conditioned on time t information, $\beta \in (0,1)$ is a time discount factor, and u is an infinitely differentiable utility function that is concave in c and $-n$ and satisfies Inada conditions. Partial derivatives are denoted by $u_{c,i,t} \equiv u_c\left(c_{i,t}, n_{i,t}\right)$, $u_{n,i,t} \equiv u_n\left(c_{i,t}, n_{i,t}\right)$, and so on. A random variable with subscript t is measurable with respect to time t information.

Agent i supplies $\epsilon_{i,t} n_{i,t}$ units of effective labor, where $\epsilon_{i,t}$ is an exogenous productivity process. A unit of effective labor receives nominal wage $P_t W_t$, where P_t is the nominal price of the final consumption good at time t. Agents trade a one-period risk-free nominal bond. The price of the bond is denoted by Q_t, which equals the inverse of the gross nominal rate between periods t and $t{+}1$. We use $P_t b_{i,t}$ to denote the face value of nominal bonds owned by agent i at end of period t , and $P_t d_{i,t}$ to denote nominal dividends received from intermediate goods producers during period t. In what will serve as our *baseline* specification, we assume that agent i's dividends in period t are $d_{i,t} = s_i D_t$, where s_i is fixed over time, a specification that restricts agents not to trade equity.

Let $\Pi_t = \frac{P_t}{P_{t-1}} - 1$ denote the net inflation rate. Households receive a uniform lump-sum transfer T_t and face a linear tax Υ_t^n on their labor earnings, a tax Υ_t^d on their dividends, and a tax Υ_t^b on their interest income.[3] The budget constraint of household i at date t in units of final goods is

$$c_{i,t} + Q_t b_{i,t} = \left(1 - \Upsilon_t^n\right) W_t \epsilon_{i,t} n_{i,t} + T_t + \left(1 - \Upsilon_t^d\right) d_{i,t} + \left(1 - \Upsilon_t^b\right) \frac{b_{i,t-1}}{1 + \Pi_t}. \tag{2}$$

The government's budget constraint at time t is

$$\bar{G} + T_t + \frac{B_{t-1}}{1 + \Pi_t} = \int \left[\Upsilon_t^n W_t \epsilon_{i,t} n_{i,t} + \Upsilon_t^d d_{i,t} + \frac{\Upsilon_t^b b_{i,t-1}}{1 + \Pi_t}\right] di + Q_t B_t,$$

[3] Although Υ_t^b multiplies $b_{i,t-1}$, we refer to it as a tax on the interest income because it is equivalent to a tax on the return on a one-period bond. To see this, rewrite the budget constraint using the market value of nominal debt $\mathrm{b}_{i,t} = Q_t b_{i,t}$ and notice that $\left(1 - \Upsilon_t^b\right) \frac{b_{i,t-1}}{1+\Pi_t} = \left(1 - \Upsilon_t^b\right) \left(R_{t-1,t}\right) \mathrm{b}_{i,t-1}$, where $R_{t-1,t} = \left(\frac{1}{Q_{t-1}}\right) \left(\frac{1}{1+\Pi_t}\right)$ is the real return from holding a nominal bond from $t-1$ to t.

where $\bar{G}$ is a time-invariant level of non-transfer government expenditures. We denote $\Upsilon_t \equiv \left(\Upsilon_t^n, \Upsilon_t^d, \Upsilon_t^b\right)$.

A final good Y_t is produced by competitive firms that use a continuum of intermediate goods $\{y_t(j)\}_{j\in[0,1]}$ as inputs into a production function

$$Y_t = \left[\int_0^1 y_t(j)^{\frac{\Phi_t - 1}{\Phi_t}} dj\right]^{\frac{\Phi_t}{\Phi_t - 1}},$$

where the elasticity of substitution Φ_t is stochastic. Final good producers take the final good price P_t and the intermediate goods prices $\{p_t(j)\}_j$ as given and solve

$$\max_{\{y_t(j)\}_{j\in[0,1]}} P_t \left[\int_0^1 y_t(j)^{\frac{\Phi_t - 1}{\Phi_t}} dj\right]^{\frac{\Phi_t}{\Phi_t - 1}} - \int_0^1 p_t(j) y_t(j) dj. \tag{3}$$

Outcomes of optimization problem (3) are a demand function for intermediate goods

$$y_t(j) = \left(\frac{p_t(j)}{P_t}\right)^{-\Phi_t} Y_t \tag{4}$$

and a final goods price that satisfies

$$P_t = \left(\int_0^1 p_t(j)^{1-\Phi_t}\right)^{\frac{1}{1-\Phi_t}}.$$

Intermediate goods $y_t(j)$ are produced by monopolists with production functions

$$y_t(j) = \left[n_t^D(j)\right]^{\alpha} \left[h_t(j)\right]^{1-\alpha}, \tag{5}$$

where $n_t^D(j)$ is effective labor hired by firm j and $h_t(j)$ is an intermediate input measured in units of the final good. Intermediate goods monopolists face downward sloping demand curves $\left(\frac{p_t(j)}{P_t}\right)^{-\Phi_t} Y_t$ and choose prices $p_t(j)$, while bearing quadratic Rotemberg (1982) price adjustment costs $\frac{\psi}{2}\left(\frac{p_t(j)}{p_{t-1}(j)} - 1\right)^2$ measured in units of the final consumption good. Intermediate goods producing firm j chooses prices $\{p_t(j)\}_t$ and factor inputs $\left\{h_t(j), n_t^D(j)\right\}_t$ that solve

$$\max_{\left\{p_t(j), n_t^D(j), h_t(j)\right\}_t} \mathbb{E}_0 \sum_t S_t (1 - \Upsilon_t^d) \left\{\frac{p_t(j)}{P_t} y_t(j) - W_t n_t^D(j) - h_t(j) - \frac{\psi}{2}\left(\frac{p_t(j)}{p_{t-1}(j)} - 1\right)^2\right\}, \tag{6}$$

subject to (4) and (5), where W_t is the real wage per unit of effective labor and S_t is a stochastic discount factor (SDF) process defined recursively via

$$S_t = S_{t-1}Q_{t-1}(1+\Pi_t)/\left(1-\Upsilon_t^b\right), \tag{7}$$

with $S_{-1} = 1$.[4] In a symmetric equilibrium, $p_t(j) = P_t$, $y_t(j) = Y_t$, $h_t\,(j) = H_t$, and $n_t^D(j) = N_t$ for all j. Market clearing conditions in labor, goods, and bond markets are

$$C_t = \int c_{i,t}di, \qquad N_t = \int \epsilon_{i,t}n_{i,t}di, \qquad D_t = Y_t - H_t - W_tN_t - \frac{\psi}{2}\Pi_t^2, \tag{8}$$

$$Y_t = N_t^{\alpha}H_t^{1-\alpha}, \qquad \Pi_t = P_t/P_{t-1} - 1 \tag{9}$$

$$C_t + \bar{G} = Y_t - H_t - \frac{\psi}{2}\Pi_t^2, \tag{10}$$

$$\int b_{i,t}di = B_t. \tag{11}$$

There are aggregate and idiosyncratic shocks. Aggregate shocks are a "markup" shock Φ_t and an aggregate productivity shock Θ_t that follow AR(1) processes

$$\begin{aligned}\ln\Phi_t =& \rho_\Phi \ln\Phi_{t-1} + (1-\rho_\Phi)\ln\bar{\Phi} + \mathcal{E}_{\Phi,t},\\ \ln\Theta_t =& \rho_\Theta \ln\Theta_{t-1} + (1-\rho_\Theta)\ln\bar{\Theta} + \mathcal{E}_{\Theta,t},\end{aligned}$$

where $\mathcal{E}_{\Phi,t}$ and $\mathcal{E}_{\Theta,t}$ are mean-zero random variables that are i.i.d. over time and uncorrelated with each other at all times.

Individual productivity $\epsilon_{i,t}$ follows a stochastic process described by

$$\ln\epsilon_{i,t} = \ln\Theta_t + \ln\theta_{i,t} + \varepsilon_{\epsilon,i,t}, \tag{12}$$

$$\ln\theta_{i,t} = \rho_\theta \ln\theta_{i,t-1} + \varepsilon_{\theta,i,t}, \tag{13}$$

where innovations $\varepsilon_{\epsilon,i,t}$ and $\varepsilon_{\theta,i,t}$ are mean-zero, uncorrelated with each other, and i.i.d. across time.

We set the initial price level $P_{-1} = 1$. In period 0, agent i is characterized by a triple $(\theta_{i,-1}, b_{i,-1}, s_i)$, where $\theta_{i,-1}$ is agent i's initial persistent component of productivity, $b_{i,-1}$ denotes the bonds that agent i initially owns, and s_i denotes agent i's initial ownership of equity. Initial conditions include the set $\{\theta_{i,-1}, b_{i,-1}, s_i\}_i$ for individuals states and a vector (Φ_{-1}, Θ_{-1}) for the aggregate shocks.

[4]In economies with heterogeneous agents and incomplete markets, a stand must be taken on how firms are valued. To explain our numerical methods most transparently, we chose a simple specification of the SDF that discounts future profits at the after-tax real risk-free rate. Our quantitative results are virtually identical when we use other choices of SDFs: equally and asset-weighted averages of individual intertemporal marginal rates of substitutions as well as a risk-neutral SDF.

2.1 Ramsey problems

Before diving into details about how we approximate a Ramsey plan in section 3, it is useful here to provide definitions of a Ramsey plan.

Definition 1. Given initial conditions and a monetary-fiscal policy $\{Q_t, \Upsilon_t, T_t\}_t$, a competitive equilibrium is a stochastic process $\left\{\{c_{i,t}, n_{i,t}, b_{i,t}\}_i, C_t, N_t, B_t, W_t, P_t, Y_t, H_t, D_t, \Pi_t, S_t\right\}_t$ that satisfies: (i) $\{c_{i,t}, n_{i,t}, b_{i,t}\}_{i,t}$ maximize (1) subject to (2) and natural debt limits;[5] (ii) final goods firms choose $\{y_t(j)\}_j$ to maximize (3); (iii) intermediate goods producers' prices and factor inputs solve (6) and satisfy $p_t(j) = P_t,\ y_t(j) = Y_t,\ h_t(j) = H_t,$ and $n_t^D(j) = N_t$ for all j; and (iv) market clearing conditions (8)-(11) are satisfied.

We can characterize competitive equilibria by feasibility constraints (7), (8), (9), and (10); consumers' and firms' optimality conditions

$$(1-\Upsilon_t^n)W_t\epsilon_{i,t}u_{c,i,t} = -u_{n,i,t}, \tag{14}$$

$$Q_t u_{c,i,t} = \beta\mathbb{E}_t u_{c,i,t+1}\left(1-\Upsilon_{t+1}^b\right)/\left(1+\Pi_{t+1}\right), \tag{15}$$

$$\begin{aligned}0 = \frac{1}{\psi}Y_t\left[1-\Phi_t\left(1-\frac{1}{1-\alpha}\left(\frac{1-\alpha}{\alpha}W_t\right)^{\alpha}\right)\right] - \Pi_t(1+\Pi_t)\\ +\mathbb{E}_t\frac{S_{t+1}}{S_t}\left(\frac{1-\Upsilon_{t+1}^d}{1-\Upsilon_t^d}\right)\Pi_{t+1}(1+\Pi_{t+1}),\end{aligned} \tag{16}$$

$$\frac{1-\alpha}{\alpha}W_t = \frac{H_t}{N_t}, \tag{17}$$

and agents' budget constraints that, by using equation (14) to eliminate $(1-\Upsilon_t^n)W_t\epsilon_{i,t}$, we can represent as

$$c_{i,t} - T_t - (1-\Upsilon_t^d)s_iD_t - \frac{(1-\Upsilon_t^b)b_{i,t-1}}{1+\Pi_t} = \left(\frac{u_{n,i,t}}{u_{c,i,t}}\right)n_{i,t} + \mathbb{E}_t\left(\frac{u_{c,i,t+1}}{u_{c,i,t}}\right)\frac{(1-\Upsilon_{t+1}^b)b_{i,t}}{1+\Pi_{t+1}}. \tag{18}$$

A Ramsey planner orders allocations by

$$\mathbb{E}_0\int\sum_{t=0}^{\infty}\beta^t\vartheta_i u\left(c_{i,t}, n_{i,t}\right)di, \tag{19}$$

[5]We impose natural debt limits by imposing for all t

$$\lim_{s\to\infty}\mathbb{E}_t\left(\prod_{k=1}^{s}Q_{t+k}\left(1-\Upsilon_{t+k+1}^b\right)\right)P_{t+s}b_{i,t+s} = 0.$$

where $\vartheta_i \geq 0$ is a Pareto weight attached to agent i and $\int \vartheta_i di = 1$.

Definition 2. Given initial conditions and a time-invariant tax policy satisfying $\Upsilon_t = \bar{\Upsilon}$ for some $\bar{\Upsilon}$, *an optimal monetary policy* is a stochastic process $\{Q_t, T_t\}_t$ that brings about a competitive equilibrium allocation that maximizes (19). Given initial conditions, *an optimal monetary-fiscal policy* is a stochastic process $\{Q_t, \Upsilon_t, T_t\}_t$ that implements a competitive equilibrium allocation that maximizes (19). A maximizing monetary or monetary-fiscal stochastic process is called a *Ramsey plan*; an associated allocation is called a *Ramsey allocation.*

We construct an optimal monetary-fiscal policy and an associated competitive equilibrium by maximizing the welfare criterion (19) subject to constraints (7)-(10) and (14)-(18). Choice of an optimal monetary policy is subject to the constant-tax-rate constraints $\Upsilon_t = \bar{\Upsilon}$ for all $t \geq 0$.

2.2 Discussion of the environment

To bring out economic forces that shape how optimal policies respond to aggregate shocks, we use two baselines that differ in whether a Ramsey planner can adjust tax rates. In what we call our *optimal monetary-fiscal policy baseline*, a Ramsey planner can freely adjust the nominal interest rate and all tax rates in response to aggregate shocks. In what we call our *optimal monetary policy baseline*, the Ramsey planner can adjust only the nominal interest rate. Our use of a monetary-policy-only baseline follows a New Keynesian tradition that, in our notation, imposes time-invariant tax rates $\Upsilon_t = \bar{\Upsilon}$ and assumes that only the nominal interest rate Q_t^{-1} and lump sum transfers T_t can respond to shocks, with lump sum transfers adjusting to satisfy the government's budget constraint. A popular justification for this restriction is that central banks can adjust interest rates fast enough to react to shocks at business cycle frequencies, while institutional constraints prevent governments from adjusting tax rates quickly. Optimal monetary policy depends on $\bar{\Upsilon}$; following the New Keynesian tradition, our section 4 quantitative application focuses on a level of $\bar{\Upsilon}$ that maximizes welfare (19) under an optimal monetary policy associated with that $\bar{\Upsilon}$.

We extend a New Keynesian model like that of Galí (2015, ch. 3) to allow for incomplete markets and heterogeneous agents as in the models of Bewley (1977, 1980), Huggett (1993), and Aiyagari (1994). Galí's setup has the advantage that policy prescriptions for the representative agent version are widely understood. That allows us to isolate modifications of those prescriptions that heterogeneity and incomplete markets bring. In our baseline environment, we model heterogeneity with a wage process representative of ones used in the macro labor literature–for instance, Low et al. (2010). In section 6.4, we enrich the baseline process for wage dynamics to allow for some of the diverse responses of labor earnings to recessions documented by Guvenen et al. (2014).

In our two baseline models, we assume that all agents can freely trade bonds subject to natural debt limits. This means that Ricardian equivalence holds so that optimal timings of transfers are undetermined.[6] This is a natural baseline since economies with *ad hoc* debt limits often imply Ramsey plans that prescribe a non-stationary optimal fiscal policy that front-loads transfers in order to undo those debt constraints.[7] We relax that assumption in section 6.2 by including a subset of liquidity-constrained agents.

In our baseline models, we assume that agents can trade debt but not equity. We relax the nontradability of claims to dividends in section 6.3 when we introduce mutual funds that hold corporate equity and government debt and that issue mutual fund shares that households trade in a competitive market.

3 Approximating a Ramsey plan

We approximate Ramsey plans for heterogeneous agent (HA) economies that present a continuation Ramsey planner with a state vector that includes a joint probability distribution of agents' characteristics. For reasons anticipated in section 1.1, this feature prevents us from approximating with a projection method like that of Krusell and Smith (1998) or a method like that of Reiter (2009) who perturbs around an economy with no aggregate shocks.

A Krusell-Smith approach can work well when the dimension of a state vector is small and when policy functions are nearly affine over sufficiently large parts of the state space to make agents' policy rules aggregate well. Even in simple versions of our problem, the state vector includes a high dimensional joint distribution, so a Krusell-Smith approach would require tracking too many moments. Reiter designed his approach for situations in which it is easy to compute an invariant distribution when there are no aggregate shocks *and* in which the state vectors stay near the support of that invariant distribution when aggregate shocks are active. These conditions can prevail in some models operating under arbitrarily fixed government policies, but not in our setting.[8]

[6] In the general formulation of the Ramsey problem, we do not restrict lump-sum transfers T_t to be positive. However, in our section 4 quantitative application, transfers are always positive, since households are unequal and the planner cares about redistribution.

[7] Bhandari et al. (2017) study a Ramsey problem with *ad hoc* debt limits, in which a planner who enforces both debt contracts and tax liabilities can time transfers to undo the *ad hoc* debt limits. But sometimes a Ramsey planner can improve outcomes by not enforcing private debt contracts; see Yared (2013) for related results.

[8] The long-run behavior of the state variables in even the simplest Ramsey problems can differ dramatically with and without aggregate shocks in otherwise identical economies. One can readily see this from the classic tax-smoothing model of Barro (1979), in which government debt is the only endogenous state variable. Without aggregate shocks to government expenditures, it stays at its initial level, while with aggregate shocks it follows a random walk; thus, whether aggregate shocks are present has important implications

We propose an alternative method that constructs a stochastic sequence of small-noise expansions along a simulated optimal path. A key step uses functional derivative techniques to characterize how government decisions depend on a high-dimensional state vector that changes over time in response to aggregate shocks. We present fast computational techniques that work at any order of approximation.[9]

Section 3.1 considers a special case in which the state vector for a continuation Ramsey problem is a joint distribution of agents' characteristics. This example allows us to describe essential aspects of our approach. Section 3.2 then extends things to settings with additional state variables that appear in the quantitative applications presented in later sections. Section 3.3 discusses numerical accuracy, computational speed, and comparisons with earlier methods for approximating equilibria of HA economies.

3.1 An enlightening special case

In the following case, (i) utilitarian (i.e., equal) Pareto weights are imposed: (ii) equity holdings s_i are uniform across households; (iii) $\alpha = 1$, so that no intermediate goods are used as inputs; (iv) all shocks are i.i.d. These restrictions reduce the size of the state space while keeping it large enough to convey essential features of our technique. The second and third assumptions imply that the Phillips curve constraint (16) is slack in all periods and so can be omitted from the optimal monetary-fiscal policy problem. The last assumption implies that past shocks do not appear as arguments in optimal policy functions.

There are alternative ways to choose state vectors. Our approach works best when state vectors satisfy an independence property that we define below. In the simple economy under study, most popular choices of state variables satisfy this property. We purposefully adopt a recursive formulation that preserves the independence property in more general settings.

Let $M_t \equiv \int u_{c,i,t} di$ be the average marginal utility of consumption at time t, and let $m_{i,t} \equiv u_{c,i,t}/M_t$ be the scaled marginal utility of consumption of agent i at time t. We can interpret $m_{i,t}$ as (an inverse of) a "Negishi" weight that the planner attaches to agent i at

for the long-run distribution of debt. Similarly, Aiyagari et al. (2002), Farhi (2010), and Bhandari et al. (2017) all study Ramsey policies and find that the invariant distribution of state variables, while being well defined in all cases that they consider, is discontinuous with respect to the size of aggregate shocks around a no-aggregate-shock case.

[9] While we compare our techniques to alternatives in detail in section 3.3, an informal summary might help at this point. Perturbation methods of Reiter (2009) and Kaplan et al. (2018) are exact with respect to the dependence of policy functions on idiosyncratic shocks but only first-order approximate with respect to aggregate shocks, all around a fixed distribution $\bar{\Omega}$. That means that approximation errors are on the order of $\mathcal{O}\left(\sigma_{agg}^2, \|\Omega - \bar{\Omega}\|^2\right)$, where σ_{agg} measures the size of aggregate shocks. Childers (2018) provides a formal treatment. Our approach constructs expansions with respect to both aggregate and idiosyncratic shocks around the time t distribution Ω_t in that period, and can be done to an arbitrary order of approximation. Approximation errors are of the order $\mathcal{O}\left(\sigma_{agg}^{n+1}, \sigma_{idiosync}^{n+1}\right)$ for arbitrary n. The two approaches are therefore complementary and have advantages and disadvantages in different applications.

time t.[10] Replace (15) with

$$Q_t M_t m_{i,t} = \beta \mathbb{E}_t u_{c,i,t+1} \left(1 - \Upsilon^b_{t+1}\right) / \left(1 + \Pi_{t+1}\right), \; m_{i,t} = u_{c,i,t}/M_t, \; M_t = \int u_{c,i,t} di. \quad (20)$$

Let $\beta^t \mu_{i,t}$ be a Lagrange multiplier on constraint (18) for agent i. Following Marcet and Marimon (2019), the monetary-fiscal Ramsey planner's Lagrangian is

$$\begin{aligned}\inf \sup \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int \Big[& u\left(c_{i,t}, n_{i,t}\right) + \left(u_{c,i,t} c_{i,t} + u_{n,i,t} n_{i,t} - u_{c,i,t}(T_t + \left(1 - \Upsilon^d_t\right) D_t)\right) \mu_{i,t} \\ & + \left(1 - \Upsilon^b_t\right) \frac{b_{i,t-1}}{1 + \Pi_t} u_{c,i,t} \left(\mu_{i,t-1} - \mu_{i,t}\right) \Big] di \end{aligned} \quad (21)$$

subject to $\mu_{i,-1} = 0$; the infimum is with respect to $\{\mu_{i,t}\}_{i,t}$ and the supremum is with respect to the stochastic process

$$\{c_{i,t}, n_{i,t}, b_{i,t}, m_{i,t}, C_t, N_t, D_t, T_t, M_t, Q_t, \Pi_t, \Upsilon_t\}_{i,t}$$

subject to constraints (8)-(10), (14), and (20).

3.1.1 Computational strategy

In the online appendix, we show that the solution to (21) can be conveniently split into a set of functions that describe the $t = 0$ choices of the Ramsey planner, and a set of functions for $t \geq 1$ choices. We also show that the $t \geq 1$ allocation is a function of the bivariate distribution over $\boldsymbol{z}_{i,t-1} \equiv (m_{i,t-1}, \mu_{i,t-1})$. We denote this distribution by Ω and use $\boldsymbol{z}$ to denote a typical value in the support of Ω. Working backwards, we solve problem (21) in two steps. First, we solve a typical continuation Ramsey planner's problem for a $t \geq 1$. Second, we solve a time $t = 0$ Ramsey problem to obtain an allocation $\{c_{i,0}, n_{i,0}\}_i$ and a distribution Ω_0, both as functions of the initial state $\{\theta_{i,-1}, s_{i,-1}, b_{i,-1}\}_i$ confronting the Ramsey planner. We devote most of the text to the continuation Ramsey plan and focus on how policy functions depend on the cross-section distribution of agent's characteristics.[11]

3.1.2 Computing a continuation Ramsey plan

We use tildes to denote policy functions for a time $t \geq 1$ continuation Ramsey plan. *Aggregate* policy functions determine the time t values of all upper-case choice variables in problem (21). We denote the vector of these functions by $\tilde{\boldsymbol{X}}\left(\Omega, \boldsymbol{\mathcal{E}}\right)$, where $\boldsymbol{\mathcal{E}}$ is a vector of aggregate shocks. *Individual* policy functions determine all lower-case time t choice variables for the

[10]We call $m_{i,t}$ as "Negishi" weights to distinguish from Pareto weights ϑ_i.

[11]See the online appendix for details about the time $t = 0$ Ramsey problem.

planner in problem (21). We denote individual policy functions by $\tilde{\boldsymbol{x}}\left(\boldsymbol{z}, \Omega, \boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}\right)$, where $\boldsymbol{\varepsilon}$ is a vector of idiosyncratic shocks. Policy functions for individual states $\tilde{\boldsymbol{z}}$ are components of $\tilde{\boldsymbol{x}}$. We define p to be a matrix that selects $\tilde{\boldsymbol{z}}$ from $\tilde{\boldsymbol{x}}$ so that $\tilde{\boldsymbol{z}} = \mathsf{p}\tilde{\boldsymbol{x}}$. The law of motion for the aggregate state is $\Omega' = \widetilde{\Omega}\left(\Omega, \boldsymbol{\mathcal{E}}\right)$.

Consider the subset of first-order optimality conditions for problem (21) for $t \geq 1$. We split these conditions into two groups. The first group consists of optimality conditions for individual choices that connect current period individual and aggregate policy functions $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{X}}$; current period realizations of shocks ($\boldsymbol{\varepsilon}$, $\boldsymbol{\mathcal{E}}$); and expectations of current and next period policy functions, $\mathbb{E}\left[\tilde{\boldsymbol{x}}|\boldsymbol{z}, \Omega\right]$ and $\mathbb{E}\left[\tilde{\boldsymbol{x}}(\tilde{\boldsymbol{z}}(\boldsymbol{z}, \Omega, \boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}), \tilde{\Omega}(\Omega, \boldsymbol{\mathcal{E}}), \cdot, \cdot)|\boldsymbol{z}, \Omega, \boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}\right]$. To economize notation, we denote these two mathematical expectations by $\mathbb{E}_{-}\tilde{\boldsymbol{x}}$ and $\mathbb{E}_{+}\tilde{\boldsymbol{x}}$, respectively. The first group of conditions can be written as

$$F\left(\mathbb{E}_{-}\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}, \mathbb{E}_{+}\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{X}}, \boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}, \boldsymbol{z}\right) = \mathbf{0} \tag{22}$$

for a collection of functions F.[12] The second group of optimality conditions for a continuation Ramsey problem are various aggregate feasibility constraints and first-order conditions with respect to $\tilde{\boldsymbol{X}}$ that connect aggregate functions and averages of individual policy functions. These conditions can be written as

$$R\left(\int \tilde{\boldsymbol{x}} d\Omega, \tilde{\boldsymbol{X}}, \boldsymbol{\mathcal{E}}\right) = \mathbf{0} \tag{23}$$

for some mapping R. The law of motion for measure Ω is

$$\Omega'\left(\boldsymbol{z}\right) = \widetilde{\Omega}\left(\Omega, \boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{z}\right) = \int \iota\left(\tilde{\boldsymbol{z}}\left(\boldsymbol{y}, \Omega, \boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}\right) \leq \boldsymbol{z}\right) d\Pr\left(\boldsymbol{\varepsilon}\right) d\Omega\left(\boldsymbol{y}\right) \quad \forall \boldsymbol{z} \tag{24}$$

where $\iota\left(\tilde{\boldsymbol{z}} \leq \boldsymbol{z}\right)$ is 1 if all elements of $\tilde{\boldsymbol{z}}$ are less than or equal to all elements of $\boldsymbol{z}$ and zero otherwise.

At each point in time $t \geq 1$, we use perturbation methods to approximate how continuation Ramsey policy functions depend on $\boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}$ shocks as the cross-section distribution Ω_t of individual characteristics evolves through a simulated history. From these approximations, we can deduce how the aggregate shock $\boldsymbol{\mathcal{E}}_t$ affects the time $t+1$ distribution Ω_{t+1}. See the online appendix for the list of equations that constitute the F and R mappings for the Ramsey problem (21).

To construct small-noise approximations of policy functions, we consider a family of economies parameterized by a positive scalar σ that multiplies *all* shocks $\boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}$, so that

[12] Strictly speaking, if $\tilde{\boldsymbol{x}}$ consists of all lower-case choice variables and multipliers in problem (21), then the relevant objects are $\mathbb{E}_{-}f(\tilde{\boldsymbol{x}})$ and $\mathbb{E}_{+}g(\tilde{\boldsymbol{x}})$ for some transformations f and g. Our exposition would become more general if we were to extend the definition of $\tilde{\boldsymbol{x}}$ to include variables $f\left(\tilde{\boldsymbol{x}}\right)$ and $g\left(\tilde{\boldsymbol{x}}\right)$, for example, by including variable $\tilde{u}_c$ in vector $\tilde{\boldsymbol{x}}$ and its definition $\tilde{u}_c = u_c\left(\tilde{c}, \tilde{n}\right)$ in mapping F.

policy functions are $\tilde{\boldsymbol{X}}(\Omega, \sigma\boldsymbol{\mathcal{E}}; \sigma)$ and $\tilde{\boldsymbol{x}}(\boldsymbol{z}, \Omega, \sigma\boldsymbol{\varepsilon}, \sigma\boldsymbol{\mathcal{E}}; \sigma)$. Let $\bar{\boldsymbol{X}}(\Omega)$ and $\bar{\boldsymbol{x}}(\boldsymbol{z}, \Omega)$ denote these functions evaluated at $\sigma = 0$. We will often suppress dependence on Ω when it is clear from the context.[13] We assume that policy functions are smooth enough to justify taking derivatives. We let $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}$, $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$, $\bar{\boldsymbol{x}}_{\varepsilon}(\boldsymbol{z})$ be gradients of policy functions with respect to aggregate and idiosyncratic shocks, and $\bar{\boldsymbol{X}}_{\sigma}$ and $\bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z})$ denote their derivatives with respect to σ, all evaluated at $\sigma = 0$. Similarly, $\bar{\Omega}_{\boldsymbol{\mathcal{E}}}$ refers to the gradient of $\widetilde{\Omega}(\Omega, \sigma\boldsymbol{\mathcal{E}}; \sigma)$ with respect to aggregate shocks at $\sigma = 0$. First-order small noise expansions of policy functions are

$$\tilde{\boldsymbol{X}}(\Omega, \sigma\boldsymbol{\mathcal{E}}; \sigma) = \bar{\boldsymbol{X}} + \sigma\left(\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}\boldsymbol{\mathcal{E}} + \bar{\boldsymbol{X}}_{\sigma}\right) + \mathcal{O}(\sigma^2) \tag{25}$$

and

$$\tilde{\boldsymbol{x}}(\boldsymbol{z}, \Omega, \sigma\boldsymbol{\varepsilon}, \sigma\boldsymbol{\mathcal{E}}; \sigma) = \bar{\boldsymbol{x}}(\boldsymbol{z}) + \sigma\left(\bar{\boldsymbol{x}}_{\varepsilon}(\boldsymbol{z})\varepsilon + \bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})\boldsymbol{\mathcal{E}} + \bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z})\right) + \mathcal{O}(\sigma^2). \tag{26}$$

Higher-order expansions are constructed analogously.

3.1.3 Zeroth-order expansions

Higher-order approximations of policy functions use inputs from lower-order approximations, so we start with a zeroth-order approximation constructed from an economy without shocks. We use bars to denote zeroth-order approximations to functions.

Lemma 1. *For any Ω and any $\boldsymbol{z}$, zeroth-order approximations to policy functions satisfy $\bar{\boldsymbol{z}}(\boldsymbol{z}, \Omega) = \boldsymbol{z}$ and therefore $\bar{\Omega}(\Omega) = \Omega$.*

Proof. The first-order condition with respect to $b_{i,t-1}$ in (21) is

$$\mathbb{E}\left[\frac{\tilde{u}_c(\boldsymbol{z}, \Omega, \cdot, \cdot)}{1 + \tilde{\Pi}(\Omega, \cdot,)}\left(\mu - \tilde{\mu}(\boldsymbol{z}, \Omega, \cdot, \cdot)\right)\right] = 0,$$

which implies $\bar{\mu}(\boldsymbol{z}, \Omega) = \mu$ for all $\boldsymbol{z}, \Omega$. To the zeroth-order equation (20) is

$$\bar{Q}(\Omega)\bar{M}(\Omega)m = \beta\bar{m}(\boldsymbol{z}, \Omega)\bar{M}\left(\bar{\Omega}(\Omega)\right)\left(1 + \bar{\Pi}\left(\bar{\Omega}(\Omega)\right)\right)^{-1}.$$

Since Negishi weights $m(z, \Omega)$ integrate to one, this equation implies

$$\bar{Q}(\Omega)\bar{M}(\Omega) = \beta\bar{M}\left(\bar{\Omega}(\Omega)\right)\left(1 + \bar{\Pi}\left(\bar{\Omega}(\Omega)\right)\right)^{-1}$$

and therefore that $\bar{m}(\boldsymbol{z}, \Omega) = m$ for all $\boldsymbol{z}, \Omega$. □

The cross-sectional distribution of characteristics Ω stays constant because in a $\sigma = 0$ economy a continuation Ramsey planner wants to keep Negishi weights and Lagrange

[13] For instance, $\bar{X}$ would refer to the function $\bar{X}(\Omega) \equiv \tilde{X}(\Omega, \mathbf{0}, 0)$, $\bar{x}(z)$ would refer to the function $\bar{x}(z, \Omega) \equiv \tilde{x}(z, \Omega, \mathbf{0}, \mathbf{0}, 0)$, and so on.

multipliers on individuals' budget constraints constant over time for all agents. This makes sense because the $\sigma = 0$ economy shuts down all randomness, and so a market structure with a risk-free bond only is enough to support perfect smoothing of agents' quantity choices.

Lemma 1 implies that $\bar{\boldsymbol{x}}(\bar{\boldsymbol{z}}(\boldsymbol{z}, \Omega), \bar{\Omega}(\Omega)) = \bar{\boldsymbol{x}}(\boldsymbol{z}, \Omega)$. Therefore, we can compute $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{x}}(\boldsymbol{z})$ by solving a system of non-linear equations

$$\bar{F}(\boldsymbol{z}) \equiv F\left(\bar{\boldsymbol{x}}(\boldsymbol{z}), \bar{\boldsymbol{x}}(\boldsymbol{z}), \bar{\boldsymbol{x}}(\boldsymbol{z}), \bar{\boldsymbol{X}}, \mathbf{0}, \mathbf{0}, \boldsymbol{z}\right) = \mathbf{0}, \bar{R} \equiv R\left(\int \bar{\boldsymbol{x}}(\boldsymbol{z}) d\Omega(\boldsymbol{z}), \bar{\boldsymbol{X}}, \mathbf{0}\right) = \mathbf{0}. \qquad (27)$$

From zeroth-order terms $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{x}}(\boldsymbol{z})$, we construct several objects that we use to compute higher-order terms. Let $\bar{R}_{\boldsymbol{x}}$ be the derivative of the mapping R with respect to its first argument $\int \bar{x} d\Omega$, and let $\bar{R}_{\boldsymbol{X}}$ and $\bar{R}_{\boldsymbol{\mathcal{E}}}$ be derivatives of R with respect to its second and third arguments, respectively, all evaluated at $\sigma = 0$. Similarly, let subscripts $\boldsymbol{x}-, \boldsymbol{x}$, $\boldsymbol{x}+$, $\boldsymbol{X}$, $\boldsymbol{\varepsilon}$, $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{z}$ of $\bar{F}$ denote corresponding derivatives of F with respect to each of its arguments evaluated at $\sigma = 0$. From the implicit function theorem, we have $\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z}) = \left[\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\right]^{-1} \bar{F}_{\boldsymbol{z}}(\boldsymbol{z})$. All of these objects can be constructed from $\bar{\boldsymbol{X}}, \bar{\boldsymbol{x}}(\boldsymbol{z})$.

Finally, we use $\partial\bar{\boldsymbol{x}}(\boldsymbol{z}, \Omega)$, $\partial\bar{\boldsymbol{X}}(\Omega)$ to denote Fréchet derivatives of $\bar{\boldsymbol{x}}(\boldsymbol{z}, \Omega)$ and $\bar{\boldsymbol{X}}(\Omega)$ with respect to the measure Ω.[14] Fréchet derivatives generalize the notion of gradients to infinite-dimensional variables and capture how changes in the distribution Ω affect policy functions. In principle, these Fréchet derivatives could be calculated from (27), but except for some very simple cases, that approach is impractical because of how the number of unknowns in the operators $\partial\bar{\boldsymbol{x}}(\cdot, \Omega)$ and $\partial\bar{\boldsymbol{X}}(\Omega)$ grows exponentially with the size of Ω. We show that this problem can be overcome when policy functions satisfy an *independence property* that we define in the next corollary.

Corollary 1. *Policy function* $\bar{\boldsymbol{z}}$ *satisfies the* ***independence property****:* $\partial\bar{\boldsymbol{z}}(\boldsymbol{z}, \Omega) = \mathbf{0}$ *for all* $\boldsymbol{z}, \Omega$*.*

Corollary 1 asserts that at $\sigma = 0$, the Fréchet derivative of policy functions for individual states equals zero. This property eases the task of calculating $\partial\bar{\Omega}$, a key term in our expansions. In the case studied in this section, corollary 1 and lemma 1 imply $\partial\bar{\Omega} = I$. More generally, we show that as long as the independence property is satisfied, $\partial\bar{\Omega}$ can be expressed in terms of $\bar{\boldsymbol{z}}_{\boldsymbol{z}}(\boldsymbol{z})$, which is easy to compute.

[14] A Fréchet derivative of some variable $\bar{X}(\Omega)$ is a linear operator from the space of distributions Ω to $\mathbb{R}$ with a property that $\lim_{\|\Delta\|\to 0} \frac{\left\|\bar{X}(\Omega+\Delta) - \bar{X}(\Omega) - \partial\bar{X}(\Omega)\cdot\Delta\right\|}{\|\Delta\|} = 0$. It can be found by fixing a feasible direction Δ and calculating a directional (Gateaux) derivative, since when both derivatives exist, they coincide, $\partial\bar{X}(\Omega) \cdot \Delta = \lim_{\alpha\to 0} \frac{\bar{X}(\Omega+\alpha\Delta) - \bar{X}(\Omega)}{\alpha}$. Following Luenberger (1997), we refer to $\partial\bar{X}(\Omega) \cdot \Delta$ as a Fréchet derivative of $\bar{X}$ at a point Ω with increment Δ. Think of $\partial\bar{X}(\Omega)$ as a measure and $\partial\bar{X}(\Omega) \cdot \Delta$ as an integral of function Δ with respect to $\partial\bar{X}(\Omega)$.

3.1.4 First-order expansions

We can now construct a first-order Taylor expansion of equations (22)-(24). As a preliminary step, use lemma 1 and observe that expansions of $\mathbb{E}_{-}\tilde{\boldsymbol{x}}$ and $\mathbb{E}_{+}\tilde{\boldsymbol{x}}$ are

$$\begin{aligned}
\mathbb{E}_{+}\tilde{\boldsymbol{x}} &= \bar{\boldsymbol{x}}(\boldsymbol{z}) + \left[\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z}) + \partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}}\right]\sigma\boldsymbol{\mathcal{E}} + \left[\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\bar{\boldsymbol{x}}_{\varepsilon}(\boldsymbol{z})\right]\sigma\varepsilon + \bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z})\sigma + \mathcal{O}(\sigma^2),\\
\mathbb{E}_{-}\tilde{\boldsymbol{x}} &= \bar{\boldsymbol{x}}(\boldsymbol{z}) + \bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z})\sigma + \mathcal{O}(\sigma^2).
\end{aligned}$$

This implies that the Ramsey planner's optimality conditions equations (22) and (23) satisfy, up to $\mathcal{O}(\sigma^2)$,

$$\begin{aligned}
&\bar{F}(\boldsymbol{z}) + \left[\left(\bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\,\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\right)\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\,\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}} + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\,\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}} + \bar{F}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})\right]\sigma\boldsymbol{\mathcal{E}}\\
&\qquad + \left[\left(\bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\,\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\right)\bar{\boldsymbol{x}}_{\varepsilon}(\boldsymbol{z}) + \bar{F}_{\varepsilon}(\boldsymbol{z})\right]\sigma\varepsilon\\
&\qquad + \left[\left(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\right)\bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{X}}\bar{\boldsymbol{X}}_{\sigma}\right]\sigma = \mathbf{0}
\end{aligned} \tag{28}$$

and

$$\bar{R} + \left[\bar{R}_{\boldsymbol{x}}\int\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})d\Omega + \bar{R}_{\boldsymbol{X}}\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}} + \bar{R}_{\boldsymbol{\mathcal{E}}}\right]\sigma\boldsymbol{\mathcal{E}} + \left[\bar{R}_{\boldsymbol{x}}\int\bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z})d\Omega + \bar{R}_{\boldsymbol{X}}\bar{\boldsymbol{X}}_{\sigma}\right]\sigma = \mathbf{0}. \tag{29}$$

Equations (28) and (29) must hold for all ε, $\boldsymbol{\mathcal{E}}$ and σ and characterize $\left\{\bar{\boldsymbol{x}}_{\varepsilon}(\boldsymbol{z}), \bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z}), \bar{\boldsymbol{X}}_{\sigma}, \bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z}), \bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}\right\}$. Let's consider each of these functions in turn. From (28), we immediately get

$$\bar{\boldsymbol{x}}_{\varepsilon}(\boldsymbol{z}) = -\left(\bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\,\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\right)^{-1}\bar{F}_{\varepsilon}(\boldsymbol{z}).$$

All the terms on the right-hand side are known from the zeroth-order expansion, so we can compute $\bar{\boldsymbol{x}}_{\varepsilon}(\boldsymbol{z})$ by matrix inversion. This step is easily parallelizable because the computation can be done separately for each $\boldsymbol{z}$. Terms $\bar{\boldsymbol{x}}_{\sigma}(\boldsymbol{z})$ and $\bar{\boldsymbol{X}}_{\sigma}$ can be computed in a similar way; it is straightforward to verify that they equal zero.

Calculating $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$ and $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}$ is more challenging. The aggregate shock $\boldsymbol{\mathcal{E}}$ changes next period's state by $\bar{\Omega}_{\boldsymbol{\mathcal{E}}}$ and that alters expectations of next period's policies by $\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}}$, as can be seen from the first square bracket in (28). Neither $\partial\bar{\boldsymbol{x}}(\boldsymbol{z})$ nor $\bar{\Omega}_{\boldsymbol{\mathcal{E}}}$ is known at this stage. The next theorem and proof use functional derivatives to construct $\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}}$.

Theorem 1. *From the zeroth-order expansion, we can construct matrices* $\mathsf{A}(\boldsymbol{z})$ *and* $\mathsf{C}(\boldsymbol{z})$

that satisfy

$$\partial\bar{\boldsymbol{x}}(\boldsymbol{z}) = \mathsf{C}(\boldsymbol{z})\partial\bar{\boldsymbol{X}}, \tag{30a}$$

$$\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}} = \mathsf{C}(\boldsymbol{z})\partial\bar{\boldsymbol{X}}\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}} = \mathsf{C}(\boldsymbol{z})\int\mathsf{A}(\boldsymbol{y})\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y})d\Omega\left(\boldsymbol{y}\right). \tag{30b}$$

Proof. Lemma 1 implies $\partial\bar{\Omega} = \mathbf{1}$. Fréchet derivatives of (22) and (23) with arbitrary increment Δ satisfy

$$\begin{aligned}\left(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\right)\partial\bar{\boldsymbol{x}}(z)\cdot\Delta + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\partial\bar{\boldsymbol{X}}\cdot\Delta &= \mathbf{0}, &(31\text{a})\\ \bar{R}_{\boldsymbol{x}}\partial\left(\int\bar{\boldsymbol{x}}\left(\boldsymbol{y}\right)d\Omega\left(\boldsymbol{y}\right)\right)\cdot\Delta + \bar{R}_{\boldsymbol{X}}\partial\bar{\boldsymbol{X}}\cdot\Delta &= \mathbf{0}. &(31\text{b})\end{aligned}$$

The first equation yields (30a) with $\mathsf{C}(\boldsymbol{z}) = -\left(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\right)^{-1}\bar{F}_{\boldsymbol{X}}(\boldsymbol{z})$.

Since directional and Fréchet derivatives coincide, by fixing a direction Δ and computing the directional derivative (see footnote 14), we obtain

$$\partial\left(\int\bar{\boldsymbol{x}}\left(\boldsymbol{y}\right)d\Omega\left(\boldsymbol{y}\right)\right)\cdot\Delta = \int\left(\partial\bar{\boldsymbol{x}}\left(\boldsymbol{y}\right)\cdot\Delta\right)d\Omega\left(\boldsymbol{y}\right) + \int\bar{\boldsymbol{x}}\left(\boldsymbol{y}\right)d\Delta\left(\boldsymbol{y}\right). \tag{32}$$

We want to evaluate the integral on the right side at $\Delta = \bar{\Omega}_{\boldsymbol{\mathcal{E}}}$. Differentiating (24) at any $\boldsymbol{z} = (m,\mu)$ and applying lemma 1 gives

$$\bar{\Omega}_{\boldsymbol{\mathcal{E}}}\left(m,\mu\right) = -\int_{y_2\le\mu}\bar{m}_{\boldsymbol{\mathcal{E}}}\left(m,y_2\right)\omega\left(m,y_2\right)dy_2 - \int_{y_1\le m}\bar{\mu}_{\boldsymbol{\mathcal{E}}}\left(y_1,\mu\right)\omega\left(y_1,\mu\right)dy_1,$$

where ω is the density of Ω. The density of $\bar{\Omega}_{\boldsymbol{\mathcal{E}}}\left(m,\mu\right)$, which is denoted by $\bar{\omega}_{\boldsymbol{\mathcal{E}}}\left(m,\mu\right)$, is then

$$\bar{\omega}_{\boldsymbol{\mathcal{E}}}\left(m,\mu\right) = -\frac{d}{dm}\left[\bar{m}_{\boldsymbol{\mathcal{E}}}\left(m,\mu\right)\omega\left(m,\mu\right)\right] - \frac{d}{d\mu}\left[\bar{\mu}_{\boldsymbol{\mathcal{E}}}\left(m,\mu\right)\omega\left(m,\mu\right)\right].$$

Substitute this equation and (30a) into (32) to get

$$\begin{aligned}\partial\left(\int\bar{\boldsymbol{x}}\left(\boldsymbol{y}\right)d\Omega\left(\boldsymbol{y}\right)\right)\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}} &= \int\mathsf{C}(\boldsymbol{y})\partial\bar{\boldsymbol{X}}\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}}d\Omega\left(\boldsymbol{y}\right) - \int\bar{\boldsymbol{x}}\left(\boldsymbol{y}\right)\frac{d}{dm}\left[\bar{m}_{\boldsymbol{\mathcal{E}}}\left(\boldsymbol{y}\right)\omega\left(\boldsymbol{y}\right)\right]d\boldsymbol{y}\\ &\quad - \int\bar{\boldsymbol{x}}\left(\boldsymbol{y}\right)\frac{d}{d\mu}\left[\bar{\mu}_{\boldsymbol{\mathcal{E}}}\left(\boldsymbol{y}\right)\omega\left(\boldsymbol{y}\right)\right]d\boldsymbol{y}\\ &= \left(\partial\bar{\boldsymbol{X}}\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}}\right)\int\mathsf{C}(\boldsymbol{y})d\Omega\left(\boldsymbol{y}\right) + \int\bar{\boldsymbol{x}}_{\boldsymbol{z}}\left(\boldsymbol{y}\right)\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}\left(\boldsymbol{y}\right)d\Omega\left(\boldsymbol{y}\right),\end{aligned}$$

where the second equality follows from integration by parts. Substitute this expression into

(31b) and solve for $\partial \bar{\boldsymbol{X}} \cdot \bar{\Omega}_{\boldsymbol{\mathcal{E}}}$ to obtain

$$\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}' \equiv \partial \bar{\boldsymbol{X}} \cdot \bar{\Omega}_{\boldsymbol{\mathcal{E}}} = \int \mathsf{A}(\boldsymbol{y}) \bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y}) d\Omega\left(\boldsymbol{y}\right), \tag{33}$$

where $\mathsf{A}(\boldsymbol{z}) = -\left(\bar{R}_{\boldsymbol{x}} \int \mathsf{C}(\boldsymbol{y}) d\Omega\left(\boldsymbol{y}\right) + \bar{R}_{\boldsymbol{X}}\right)^{-1} \bar{R}_{\boldsymbol{x}} \bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}$. Together with (30a), we get (30b). □

Economic forces drive theorem 1. In a competitive equilibrium, agents care about the distribution Ω only because it helps them predict aggregate prices and income. That means that effects $\partial \bar{\boldsymbol{x}}$ on individual variables from a perturbation of distribution Ω can be factored into effects $\partial \bar{\boldsymbol{X}}$ on aggregate variables and a known loading matrix $\mathsf{C}(\boldsymbol{z})$, which captures how individual variables respond to changes in the aggregates. Equation (30a) captures this.

Feasibility and market clearing impose a tight relationship between how individual policy functions respond to aggregate shocks in the current period, $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$, and how aggregates are expected to change next period, $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}'$. This relationship sets up a fixed point problem presented in equation (33). Together with (30a), equation (33) allows us to express the Fréchet derivative $\partial \bar{\boldsymbol{x}} \cdot \bar{\Omega}_{\boldsymbol{\mathcal{E}}}$ as a linear function of $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}\left(\boldsymbol{z}\right)$.

The preceding analysis puts us in a position to compute the coefficients $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}$ and $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}$. Setting the first square brackets in (28) and (29) to zero and using the definition of $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}'$ from (33), we obtain the following system of linear equations in the unknowns $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}, \bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$ for all $\boldsymbol{z}$:

$$\left(\bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\right)\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\mathsf{C}(\boldsymbol{z})\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}' + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}} + \bar{F}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z}) = \boldsymbol{0} \tag{34a}$$

$$\bar{R}_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y}) d\Omega\left(\boldsymbol{y}\right) + \bar{R}_{\boldsymbol{X}} \bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}} + \bar{R}_{\boldsymbol{\mathcal{E}}} = \boldsymbol{0}. \tag{34b}$$

This linear system allows us to split one large problem of simultaneously finding $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$ for all $\boldsymbol{z}$ into a large number of small problems that independently characterize $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$ for each $\boldsymbol{z}$. Thus, we use equation (34a) to calculate matrices $\mathsf{D}_0(\boldsymbol{z})$ and $\mathsf{D}_1(\boldsymbol{z})$, which define the affine function

$$\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z}) = \mathsf{D}_0(\boldsymbol{z}) + \mathsf{D}_1(\boldsymbol{z}) \cdot \left[\begin{array}{cc} \bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}} & \bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}' \end{array}\right]^{\mathsf{T}}.$$

We can substitute this function into equations (33) and (34b) to compute $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}$ and $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}'$. Values of $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$ can be found either from the previous equation or from (30a). This completes calculations comprising first-order terms.

3.1.5 Higher-order expansions

Because a generalization of theorem 1 applies to higher-order expansions, our approach preserves linear and parallelizable structures when used to construct second- and higher-order expansions. The independence property, $\partial\bar{z}(\boldsymbol{z},\Omega) = \mathbf{0}$, allows a counterpart to equation (30a) to hold for all higher-order Fréchet derivatives. This enables us to compute higher order analogues of $\partial\bar{\boldsymbol{x}}\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}}$ explicitly as weighted sums of higher-order coefficients $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}}$, $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}\sigma}$, $\bar{\boldsymbol{x}}_{\sigma\sigma}$ ..., with weights known from lower-order expansions. We then can form higher-order analogues of equations (34). The structure of these equations allows us again to split one large system of equations into a large number of low-dimensional linear problems that can be solved fast and simultaneously. Formal proofs and constructions involve much additional notation, but the steps mirror those in section 3.1.4. An online appendix provides details.

3.2 Approximations more generally

To use our small-noise expansion method to approximate a Ramsey plan for the section 2 economy, we modify two features of the section 3.1 computations. First, now the optimality condition (16) typically binds and cannot be omitted. We add this constraint to our Lagrangian formulation (21), so its multiplier Λ now becomes an aggregate state variable for the continuation Ramsey problem. Second, since shocks are persistent, policy functions also depend on previous period values of aggregate shocks $\boldsymbol{\Theta} = (\Theta, \Phi)$ as well as idiosyncratic shocks θ. Thus, now $\boldsymbol{z} = (m, \mu, s, \theta, \vartheta)$ is the individual state, Ω is a measure over $\boldsymbol{z}$, and the aggregate and individual policy functions are functions $\tilde{\boldsymbol{X}}(\Omega, \Lambda, \boldsymbol{\Theta}, \boldsymbol{\mathcal{E}})$ and $\tilde{\boldsymbol{x}}(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta}, \boldsymbol{\mathcal{E}}, \boldsymbol{\varepsilon})$, respectively. Zeroth-order terms have non-trivial (deterministic) transition paths that can be computed with a shooting algorithm.

With persistent shocks, two ways to perturb policy functions yield approximation errors of the same orders of magnitude. One is to scale $\{\sigma\boldsymbol{\mathcal{E}}, \sigma\boldsymbol{\varepsilon}\}$ and expand with respect to σ around current values of $(\boldsymbol{\Theta}, \theta)$ and Ω. Since to the zeroth-order $\bar{\theta}(\boldsymbol{z}, \Omega) \neq \theta$, it is no longer true that $\bar{\Omega}(\Omega) = \Omega$, so lemma 1 does not apply.[15] However, functional derivative techniques used to prove theorem 1 still apply, and we can construct required Fréchet derivatives along the transition path. Tractability is preserved because policy functions still satisfy the independence property; that is, $\partial\bar{z}(\boldsymbol{z},\Omega) = \mathbf{0}$ for all $\boldsymbol{z}, \Omega$. The law of motion for exogenous variables does not depend on the distribution Ω, so adding those variables to vector $\boldsymbol{z}$ leaves the independence property intact.

An alternative approach is to scale as $\{\sigma\boldsymbol{\mathcal{E}}, \sigma\boldsymbol{\varepsilon}, \sigma\boldsymbol{\Theta}, \sigma\theta\}$ and then to expand around $\sigma = 0$. Since θ is a component of the vector $\boldsymbol{z}$ of individual characteristics, $\boldsymbol{z}$ and therefore Ω are now also functions of the scaling parameter σ. A zeroth-order approximation satisfies lemma 1, but expansions of policy functions involve additional Fréchet derivatives including $\partial\bar{\boldsymbol{X}}\cdot\bar{\Omega}_\sigma$ and $\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_\sigma$. These derivatives are easy to compute using techniques deployed in theorem 1 .

[15] When persistence of the idiosyncratic shocks ρ_θ is close to one, we can recover lemma 1 if we approximate ρ_θ by $\rho_\theta(\sigma) = 1 - \sigma\rho$ for some $\rho \geq 0$ and expand $\rho_\theta(\sigma)$ with respect to σ.

Although the two approaches imply errors of the same orders of approximation, one approach can be better than the other depending on circumstances. We use the first approach in our application, but in some cases, the second approach maybe can be faster to implement as it does not require computing a transition path. The online appendix provides explicit formulas and extensions of theorem 1 for both approaches.

3.3 Accuracy and comparisons

Our method builds on perturbation techniques widely used in computational economics (see, for example, Judd and Guu, 1993, Judd and Guu, 1997, and Schmitt-Grohe and Uribe, 2004b). We perturb "around the current state" in a way closely related to practices of Fleming (1971), Fleming and Souganidis (1986), Anderson et al. (2012), Bhandari et al. (2017), and Phillips (2017). In all of those applications, the state vector has low dimension, and approximations do not require computing high-dimensional Fréchet derivatives. Our approach is designed to apply even when the underlying state is a complicated, high-dimensional object, as typically occurs in HA economies.

To our knowledge, ours is the first method that incorporates effects of the complete current state on continuation Ramsey plans in HA economies. Our approach applies to HA economies for which equilibrium dynamics can be written in the form given by equations (22)-(24), a large class of economies.[16]

To assess accuracy, we approximate a competitive equilibrium for a given monetary-fiscal policy within an environment that we have simplified enough to allow us to compute an equilibrium analytically. We then compare that analytical solution to our approximation by varying key parameters likely to affect approximation quality.

In particular, we follow Acharya and Dogra (2018) and assume that labor is supplied inelastically at $n_{i,t} = 1$; preferences are given by $U(c_t, n_t) = -\exp(-\gamma c_t)$; equity holdings are uniform across consumers; there are no aggregate shocks; idiosyncratic shocks $\epsilon_{i,t}$ are i.i.d. normal; government spending and all tax rates equal zero; and interest rates are set according to a Taylor rule $Q_t^{-1} - 1 = a_0 (1 + \Pi_t)^{a_1}$ for coefficients a_0 and a_1 chosen to make steady state inflation equal zero.

Under these assumptions, household income $W_t \epsilon_{i,t} + T_t + D_t$ is normally distributed. Together with the CARA utility function, that means that the consumption-saving problem can be solved analytically. One can then derive explicit expressions both for steady-state aggregate quantities and for deterministic transition paths from given initial conditions.

[16]HA economies with inequality constraints, such as ones with additional ad-hoc debt limits, can be written in this form by including appropriate complementary slackness conditions. Inequality constraints often imply policy functions with kinks. Although such kinks violate the smoothness assumption that we imposed on equations (25) and (26), we foresee no impediments to extending our method to such structures in our future work.

Acharya and Dogra (2018) called this a Pseudo Representative Agent New Keynesian economy (abbreviated as PRANK) and drew from it useful insights about how more complicated HANK models work. They showed that their PRANK economy has a unique steady state in which all aggregate variables, including output, inflation, and real interest rates, are constant. In a steady state, individual assets follow a random walk, so the dispersion of asset holdings across agents grows without limit. The availability of explicit expressions for policy functions along the transition path means that there are also expressions for how this PRANK economy is affected by an unanticipated aggregate shock.

We list equilibrium conditions and calibrated parameter values in the online appendix. We start at the steady state and study equilibrium responses to a one-time, unanticipated 1.23% shock to aggregate productivity in period t that then decays deterministically.[17] We compare our second-order approximation to the exact solution. We report two comparisons: one in which the shock occurs in period $t = 1$ and another one in which it occurs in period $t = 250$. In both cases, shocks arrive when all aggregate variables are at the same steady state values; the two cases differ only in spreads of asset distributions at the time of the shock.

Blue and black solid lines in figure I show exact and approximate impulse responses of output, inflation, and asset inequality measured by the standard deviation of individual wealth. They are almost identical in both experiments. The PRANK economy is engineered to make impulse responses of output and inflation be independent of the asset distribution, so dynamics of output and inflation are the same in the top and bottom rows of figure I. This is not the case for other items of potential interest such as dynamics of asset inequality, as can be seen in the rightmost panels in figure I.

By comparing the two experiments, we can evaluate the precision of our approximation and how it deteriorates with the time horizon. In the PRANK economy, we can calculate distribution Ω_t exactly for all t and the corresponding impulse responses. Our approach approximates policy functions and distributions for $t = 1, 2, ...$, so approximation errors can accumulate over time as we compute responses far into the future. Individual wealth follows a random walk process so approximation errors accrued by our method endure forever, making this environment a worst-case test bed for our approximation method. Despite that, we find that our approximate distribution is very close to the exact distribution even at $t = 250$ (see online appendix) and that we capture responses of asset inequality to an aggregate shock in that period very well.

To document the accuracy, we compute several types of approximation errors. For the individual policy functions, we compute: (i) % gap in the approximated and true policies (ii) % gap in the approximated and policy rules implied by the Euler equation (or the Euler

[17]This corresponds to a one standard deviation shock to productivity in our baseline economy.

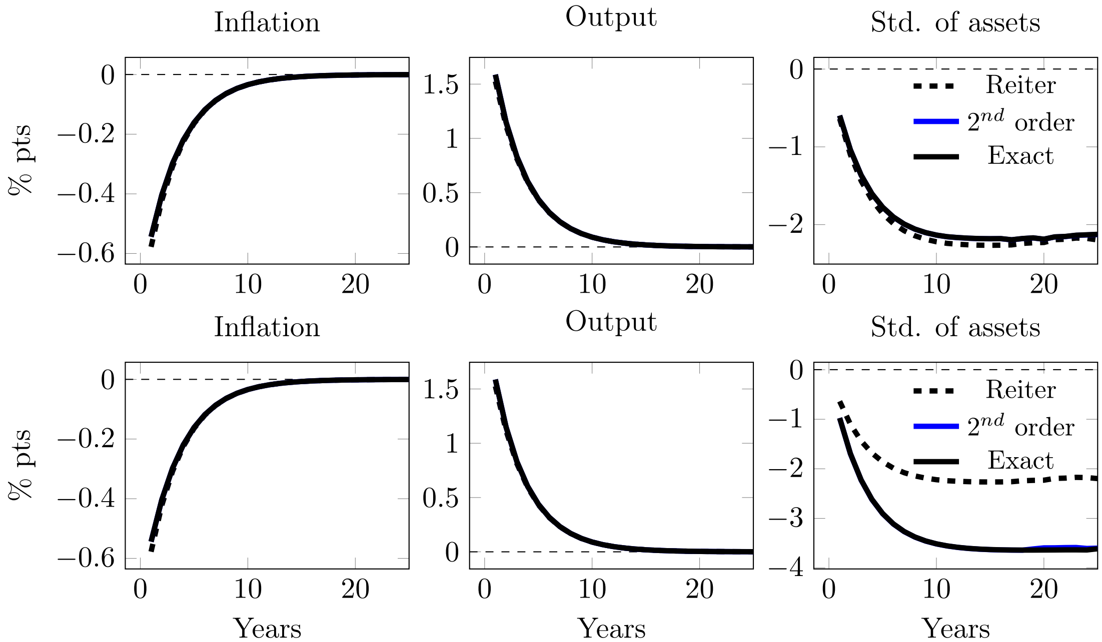

Figure I: Comparisons of impulse responses to a 1% TFP shock at $t = 1$ in the top panel and $t = 250$ in the bottom panel across approximation methods. The bold lines are are the exact solution (black) and our method applied to second-order (blue). The dashed black line are responses under the Reiter method.

equation errors) and (iii) dynamic Euler Equation errors from Den Haan (2010). The first two errors are acquired from the formulas[18]

$$E^{pol}_{c,t}(b,\epsilon) \equiv 100 \times \frac{|c^{True}_t(b,\epsilon) - c^{apprx}_t(b,\epsilon)|}{C_t} \tag{35}$$

$$E^{EE}_{c,t}(b,\epsilon) \equiv 100 \times \frac{|c^{imp}_t(b,\epsilon) - c^{apprx}_t(b,\epsilon)|}{C_t}, \tag{36}$$

where $c^{True}_t(b,\epsilon)$ are constructed using the exact solution and $c^{imp}_t(b,\epsilon)$ is defined as

$$c^{imp}_t(b,\epsilon) = -\frac{1}{\alpha}\log\left(\frac{\beta}{Q_t}\mathbb{E}_t\left[\frac{1}{1+\pi_{t+1}}\exp(-\alpha c^{approx}_{t+1}(b',\epsilon'))\right]\right).$$

The dynamic Euler equation errors are constructed by simulating a panel $\left\{\tilde{b}_{i,t}, \tilde{c}_{i,t}\right\}_{i,t}$ in the following recursive fashion. For some agent i, fix an initial value for assets $\tilde{b}_{i,-1}$ and a history of shocks $\{\epsilon_{i,t}\}$. At any period t apply the function $c^{imp}_t(\cdot)$ to the pair $\left(\tilde{b}_{i,t-1}, \epsilon_{i,t}\right)$ to construct $\tilde{c_{i,t}}$, and then use the budget constraint to construct $\tilde{b}_{i,t}$. The dynamic Euler

[18] As CARA preferences feature an aversion to absolute risk and, possibly, negative consumption, we report the absolute errors scaled by average consumption.

Maximum Errors (%)	Ind. Consumption	Agg. Output	Inflation	Interest Rate
2^{nd} Order				
$\gamma = 1, \sigma_\epsilon = 0.50$	0.0039	4.2e-6	3.1e-5	4.3e-5
$\gamma = 1, \sigma_\epsilon = 0.75$	0.0134	2.6e-5	1.5e-4	2.2e-4
$\gamma = 1, \sigma_\epsilon = 1.00$	0.0328	8.2e-5	4.9e-4	6.9e-4
$\gamma = 3, \sigma_\epsilon = 0.5$	0.0453	0.0011	0.0024	0.0034
Reiter-based				
$\gamma = 1, \sigma_\epsilon = 0.50$	0.0374	0.0616	0.0337	0.0505
$\gamma = 1, \sigma_\epsilon = 0.75$	0.0466	0.0610	0.0335	0.0501
$\gamma = 1, \sigma_\epsilon = 1.00$	0.0492	0.0602	0.0329	0.0493
$\gamma = 3, \sigma_\epsilon = 0.5$	0.0896	0.2252	0.1327	0.1991

TABLE I: Percentage errors in policy functions in response to an one standard deviation unanticipated shock to aggregate TFP at date $t = 1$. The values reported are the maximum errors across states (b, ϵ) and time t relative to the true solution.

equation errors are computed using the analogue of expression (36) in which we compare $\tilde{c}_t$ to c_t^{apprx}, where the later is computed for the same sequence of shocks and initial assets. These errors have the advantage of allowing for the possibility of small errors accumulating into large errors over time. For aggregate variable X_t we simply report

$$E_{X,t} = 100 \times \frac{|X_t^{True} - X_t^{apprx}|}{X_t^{True}}.$$

We will often report maximum errors where the max is taken over state space (b, ϵ) as well as t.

For brevity, we summarize the % gap in the approximated and true policies errors and details of computational speed of our approach in Table I. (See the online appendix for Euler Equation and Dynamic Euler Equation errors.) The errors reported are all for a quadratic approximation. Percent errors relative to the true solution for the individual consumption policies are less than 0.05% and vary between 0.004%-0.033% as we double volatility of idiosyncratic shocks, while percent errors for aggregate output, inflation, and the interest rate range from 4.3×10^{-5}% to 0.0007%. In terms of percent errors, increasing the coefficient of relative risk aversion to 3 increases approximation errors by roughly the same amount as does doubling the volatility of the idiosyncratic shocks. Since our approach is easily parallelizable and sidesteps the computationally intensive step of computing Fréchet derivatives, it works quickly and allows us to simulate transition dynamics of a path of 100 periods in 1.5 seconds on a dual AMD EPYC 7351 processor with 32 cores.

3.4 Comparison to Reiter's method

We can also compare our method to a widely-used approach of Reiter (2009). Our method with order n yields approximation errors that scale with the size of both idiosyncratic and aggregate shocks, that is, $\mathcal{O}\left(\sigma_{agg}^{n+1}, \sigma_{idiosync}^{n+1}\right)$. An approximation like Reiter's delivers policy functions that are linear in aggregate shocks and globally accurate with respect to idiosyncratic shocks around a fixed distribution $\Omega_t = \bar{\Omega}$. Therefore, errors in such an approximation scale with both the size of the aggregate shock and the distance of the current distribution from the point of approximation $\mathcal{O}\left(\sigma_{agg}^2, \|\Omega - \bar{\Omega}\|^2\right)$. We use the PRANK setting to show the trade-offs involved in these two types of errors. All of our comparisons assume $n = 2$.

The bottom four rows of Table I present the percentage errors of the Reiter method[19] relative to the analytic solution, allowing us to compare our method directly to Reiter's approach. Both approaches yield accurate approximations of the analytic solution with small percentage errors range from 0.04% to 0.9%. Despite that, we observe that our method yields errors for the individual consumption policy rules that are consistently smaller than those of the Reiter approach while errors for the aggregate variables are *two orders of magnitude smaller* than the Reiter approach. Our approach is less accurate with respect to the idiosyncratic risk, but those errors partially average out for the aggregates. Meanwhile, second-order errors in aggregates variables under the Reiter approach propagate down to the individual policy rules.

To extract long run consequences of these approximation errors, we augment the PRANK economy with a non-degenerate stochastic process for aggregate TFP. In line with standard calibrations, aggregate shocks are much less volatile than idiosyncratic shocks (standard deviation of innovations about 1% for aggregate TFP vs. 50% for individual productivities). We then simulate a long sequence of aggregate shocks and compare distributions of assets at $t = 250$ associated with our method and Reiter's. We find that the obtained asset distributions are visibly different under the two methods, with the standard deviation of assets being more than 1 percent larger under Reiter's approach. To understand the source of these differences, we drop the second-order terms with respect to aggregate risk from our expansions and re-compute the long run distribution associated with using this inferior approximation. This distribution, which has error of the order $\mathcal{O}\left(\sigma_{agg}^2, \sigma_{idiosync}^3\right)$, is almost identical to the distribution generated by Reiter's method. That finding prompts us to conclude that ignoring higher-order effects of aggregate shocks can inject long-run drifts into approximation errors. We report details about this and some related experiments in

[19] Conventional applications of Reiter's method requires expanding policy functions around the invariant distribution in the economy without aggregate shocks. Since individual assets follow a random walk, no such distribution exists in the PRANK economy. In our application of Reiter's method, we used initial distribution Ω_0 as the point of expansion.

the online appendix.

Consequences of ignoring movements of Ω_t away from $\bar{\Omega}$ can be gleaned from responses of the standard deviation of assets in figure I. While the PRANK economy is constructed so that responses of output and inflation do not depend on the asset distribution, the responses of other moments, including those that describe dispersion of individuals' asset holdings, do depend on it. This implies that Reiter's approximation of these impulse responses deteriorates progressively as the distribution of assets drifts away from the point of approximation. In this example, movement of the distribution away from the point of approximation is due to idiosyncratic income risk.

4 Calibration

To isolate key trade-offs that a continuation Ramsey planner faces, we start from a baseline economy that is close to specifications commonly used in the New Keynesian literature. We start with an initial calibration that ignores important features, including the ample heterogeneities in marginal propensities to consume and effects of recessions on labor earnings that have been documented to prevail in U.S. data. We incorporate these as extensions in section 6.

Preferences and technology parameters

We assume $u\left(c,n\right)=\frac{c^{1-\nu}}{1-\nu}-\frac{n^{1+\gamma}}{1+\gamma}$ and set $\nu=3$, $\gamma=2$. This yields a Frisch labor supply elasticity of 0.5. We calibrate to annual data and set the discount factor $\beta=0.96$. We set $\bar{\Theta}=1$ and $\bar{\Phi}=6$ to attain average markups of 20%. We abstract from the use of intermediate goods in production and set $\alpha=1$. We choose the cost of nominal price changes ψ to match the slope of the aggregate Phillips curve. Sbordone (2002) estimated the slope of the U.S. Phillips curve in quarterly data to be about 6%. We convert that to an annual frequency by multiplying by 4. To a first-order approximation the slope of the Phillips curve in our model is $\left(\bar{\Phi}-1\right)/\psi$, which implies $\psi=21$.

Idiosyncratic and aggregate uncertainty

We assume that all shocks are Gaussian and set the standard deviations of $\varepsilon_{\epsilon,i,t}$ and $\varepsilon_{\theta,i,t}$ to 8.7% and 10.3% and the autocorrelation ρ_θ =0.992 to match evidence on individual wage dynamics from Low et al. (2010).

We calibrate the stochastic process for the markup shocks so that movements in the labor share of output are consistent with movements in the U.S. corporate sector's labor share (Table 1.14, NIPA) over the period 1947-2016. Calibrated values for (ρ_Φ,σ_Φ) are

$(0.85, 4.6\%)$.[20] We calibrate the stochastic process for aggregate labor productivity $\log \Theta_t$ so that output per hour is consistent with detrended U.S. non-farm real output per hour (BLS) over the period 1947-2016. Calibrated values for $(\rho_\Theta, \sigma_\Theta)$ are $(0.73, 1.23\%)$.

Initial conditions

A common approach in the heterogeneous agent macro-labor literature is to specify government policy $\left(\bar{G}, \Upsilon_t, Q_t\right)$ and study long-run allocations in an associated competitive equilibrium. A deficiency of some current workhorse models is that their invariant distributions understate wealth inequality.[21] As we shall see, asset inequality has important implications for optimal policy responses. We calibrate initial conditions $\{\theta_{i,-1}, b_{i,-1}, s_i\}_i$ to be consistent with empirical distributions of wages, nominal claims, and claims to real firm profits. In section 6.1, we show that drift from this distribution is slow and that its presence matters very little for optimal policy responses. We use the 2007 wave of the Survey of Consumer Finances (SCF) as our benchmark for earnings and asset inequality. We adopt a procedure proposed by Doepke and Schneider (2006) to map household-level direct and indirect holdings of financial assets to the joint distribution of claims to nominal debt and claims to equity.[22] Table II reports summary statistics for our sample. Of particular relevance to results presented in section 5 is the fact that earnings and assets are positively correlated and that inequality in asset holdings is much larger than earnings inequality.

We calibrate government expenditures $\bar{G}$ to be consistent with the ratio of non-transfer government expenditures to tax revenues. To obtain tax revenues, we model a stylized U.S. tax system. To be consistent with estimates of consolidated federal and state-level average marginal tax rates calculated in Bhandari and McGrattan (2019), we assume that tax rates on labor income, dividends, and interest income are time invariant and set to $\left(\bar{\Upsilon}^n, \bar{\Upsilon}^d, \bar{\Upsilon}^b\right) = \bar{\Upsilon}^{US} = (0.38, 0.34, 0)$.[23] We set $\bar{G}$ so that on average the ratio of non-transfer

[20] There is substantial variety in how macro and financial economists have modeled and calibrated markup shocks. In the DSGE literature, for instance, Smets and Wouters (2007), Justiniano et al. (2010), and Galí et al. (2007) use ARMA(1,1) processes and estimate the quarterly persistence to be in the range of 0.90–0.95. In the finance literature, for instance, Greenwald et al. (2014) estimate factor share shocks with a monthly persistence of 0.995. Our calibrated value for $\rho_\Phi = 0.85$ lies within the range of these estimates.

[21] Incorporating one or more of the popular "fixes" to obtain a sufficiently skewed invariant distribution of wealth–for instance, allowing persistent shocks to discount factors (Krusell and Smith, 1998), bequests (De Nardi, 2004), entrepreneurial choice (Cagetti and De Nardi, 2006), or persistent idiosyncratic differences in returns to financial assets (Benhabib et al., 2019), and then computing a Ramsey allocation for such an economy would be interesting but is not something that we do in this paper.

[22] The online appendix contains details the sample restriction as well as how we apply the Doepke and Schneider (2006) procedure.

[23] To arrive at the estimate of the marginal tax rate on capital income, we combine the Bhandari and McGrattan (2019) estimates of the effective marginal tax rate on corporate business income, on distributed dividends, and the schedule of marginal tax rates on non-corporate business income into a single number by using the Barro and Redlick (2011) procedure. We use the same steps to combine the schedule of marginal tax rates on wage income into the flat tax rate used above. We set the tax rate on bond income to zero in

TABLE II: FIT OF THE INITIAL DISTRIBUTION

Moments	
Fraction of pop. with zero equities	30%
Std. share of equities	2.63
Std. bond	6.03
Gini of financial wealth	0.82
Std. ln wages	0.80
Corr(share of equities, ln wages)	0.40
Corr(share of equities, bond holdings)	0.62
Corr(bond, ln wages)	0.33

Notes: Moments correspond to SCF 2007 wave after scaling wages, equity holdings, and debt holdings by the average yearly wage in our sample. The share of equities refers to the ratio of individual equity holdings to the total in our sample; the weighted sum of shares equals one. Financial wealth is defined as the sum of nominal and real claims.

government expenditures to total tax receipts equals 50%, also estimated by Bhandari and McGrattan (2019).

Continuation Ramsey responses depend on the joint distribution of assets and after-tax incomes. For our baseline simulations, we choose Pareto weights so that average optimal levels of taxes are similar to U.S. data. In particular, we assume that Pareto weights are described by

$$\vartheta_i \propto \exp\left(\delta_1 \theta_{i,-1}\right) + \exp\left(\delta_2 s_{i,-1}\right) + \exp\left(\delta_3 b_{i,-1}\right), \tag{37}$$

where $(\theta_{i,-1}, s_{i,-1}, b_{i,-1})$ are the three dimensions of initial heterogeneity and $\delta = (\delta_1, \delta_2, \delta_3)$ are parameters that we set so that in the non-stochastic economy setting $\bar{\Upsilon} = \bar{\Upsilon}^{US}$ is optimal. In section 6.1, we discuss how outcomes vary with alternative choices of Pareto weights.

5 Optimal monetary and monetary-fiscal policies

The Ramsey planner sets a common lump-sum transfer for all agents as well as the nominal interest rate and tax rates on labor income, dividends, and bond income. The government acquires revenues directly through taxes and indirectly through nominal interest rates and inflation. It spends these revenues on servicing government debt, paying for exogenous expenditures $\bar{G}$, and financing transfers. Ricardian equivalence prevails, disarming effects

order to represent the observation that most of government bonds are held through tax-deferred accounts.

TABLE III: RAMSEY ALLOCATION: MOMENTS

	RANK					HANK				
	Std.	Correlations				Std.	Correlations			
	Dev(%)	i_t	Π_t	W_t	$\ln Y_t$	Dev(%)	i_t	Π_t	W_t	$\ln Y_t$
Nominal Rate i_t	0.87	1				1.82	1			
Inflation Π_t	0.03	-0.01	1			0.46	-0.94	1		
Labor Share W_t	1.18	-0.09	-0.32	1		2.13	-0.78	0.78	1	
Log Output $\ln Y_t$	0.92	-0.98	-0.09	0.24	1	0.88	-0.31	0.10	0.12	1

Notes: Moments are computed using allocations under RANK (left) and HANK (right) optimal monetary policies.

that might otherwise be produced by altering the timing of transfers.[24] Explicit taxes and an implicit tax via inflation generate dead-weight losses. Average levels of taxes and transfers depend on inequalities in incomes from labor and assets and on Pareto weights.

We compare continuation Ramsey plans in our baseline HANK setting to continuation Ramsey plans in a representative agent version of our model (abbreviated as RANK). In the RANK economy, we keep all the parameters the same, except we assume that agents are identical with initial conditions being equal to the average levels in our baseline HANK calibration, and there are no idiosyncratic shocks. When we compute optimal monetary policy in the RANK economy, we set $\left(\bar{\Upsilon}^n, \bar{\Upsilon}^d, \bar{\Upsilon}^b\right) = (-1/\bar{\Phi}, 0, 0)$, which corresponds to the non-stochastic optimum.

Table III reports several statistics that summarize stochastic properties of optimal monetary policies in the HANK and the RANK models (stochastic properties of optimal monetary-fiscal policies are similar and are omitted). We refer to these statistics as cyclical properties of optimal policies. To compute these moments, we run 1000 simulated paths, each 50 periods long, and for each simulated path we compute the covariance matrix for output, inflation, wages per effective hour, and nominal rates. We then average the covariance matrix across simulations.[25] As one can see from table III, inflation and nominal interest rates are much more volatile in HANK, and the co-movement of labor share with output is lower. Covariances of inflation with output and labor share have different signs in HANK and RANK settings.

[24]Extending the model as we do in 6.2 to introduce liquidity-constrained households rearms the timing of transfers as a government instrument. In that section, we shall explore optimal timing of transfers.

[25]We also experimented with running longer simulations, such as 100 periods, and did not find substantial differences in the results.

5.1 Sources of welfare gains

We first want to understand what drives differences in optimal policies between representative and heterogeneous agent economies. In the representative agent economy, an optimal policy is determined by a trade-off between maintaining price stability and reducing the deviation in output from its efficient level, the so-called "output gap". Two additional considerations affect optimal policy in heterogeneous agents settings. Government policies can redistribute resources across *ex-ante* different agents and increase welfare that the planner measures using a Pareto-weighed sum of agents life-time utilities. The planner can also increase welfare by providing better insurance *ex-post.* To understand whether differences in policy responses between HANK and RANK economies are driven by redistribution or insurance concerns, we run diagnostics.

The first diagnostic uses a method proposed by Bhandari et al. (2021) to separate welfare changes resulting from switching from some policy A to another policy B into three components: *aggregate efficiency*, that measures welfare effects of changes in the level of aggregate resources induced by the policy switch; *redistribution*, that measures welfare effects from changes in expected shares of resources received by ex-ante different agents; and *insurance*, that measures effects of changes in the ex-post volatility of consumption. The Bhandari et al. (2021) decomposition is similar in spirit to the approaches developed by Benabou (2002) and Floden (2001) but, unlike those papers, can be applied to a much larger class of heterogeneous agent environments that includes the one studied here.

In order to construct our decomposition, we need counterparts of optimal RANK policies for our heterogeneous agents setting. Average levels of taxes and transfers are very different in optimal policies for HANK and RANK economies. In the RANK economy, labor tax rates are negative and are financed by lump-sum taxes in order to offset markups. In the HANK economy, labor tax rates are positive and finance transfers. Since our focus is on cyclical properties of optimal policies, we construct a RANK-equivalent policy as follows. We set the average level of tax rates to be the same as under the HANK optimum, but choose stochastic processes for *deviations* of policy variables from their means to be the same as in a RANK optimum. The allocation in the heterogeneous agent economy with RANK-equivalent policies closely mimics the cyclical properties of the RANK economy reported in table III. We use it as "policy A" in our Bhandari et al. (2021) decomposition. "Policy B" is the Ramsey optimal allocation in a HANK economy.

By construction, policy B has higher welfare than policy A. Table IV decomposes this welfare gain into aggregate efficiency, insurance, and redistribution components and reports decompositions for both optimal monetary and monetary-fiscal policies, as well as for several extensions that we consider in section 6. This table shows several insights that carry through all of our extensions: the insurance component is positive and greater than

100%; the redistribution component is small; and the aggregate efficiency component is negative. This means that essentially all the welfare gains from optimal HANK policies arise from the additional insurance that they provide. Provision of insurance comes at the cost of sacrificing price stability, which creates deadweight losses and lowers total aggregate resources available for consumption. This explains why the aggregate efficiency component is negative. Finally, cyclical variations in monetary and fiscal policies contribute little to redistribution. Most of the redistribution is done by setting average tax rates appropriately. Deviations of tax rates from those average levels mostly provide insurance, not additional redistribution.

A second diagnostic test helps us to distinguish between insurance against aggregate and idiosyncratic shocks. In the online appendix, we study how optimal policies are affected when we switch off idiosyncratic shocks, and also when we allow agents to trade a full set of Arrow securities. We find that the latter experiment accounts for nearly all the differences in cyclical properties of optimal policies reported in table III. While switching off idiosyncratic shocks has little effect optimal policies, adding Arrow securities contingent on aggregate shocks brings optimal policies close to those in the RANK economy. From these findings we conclude that the planner's desire to replace the missing insurance markets for aggregate risk explains the differences in the optimal HANK and RANK policies.

5.2 Policy responses to aggregate shocks

To gather further insights into how the Ramsey planner sets optimal policies in HANK settings, we focus on studying policy responses to specific shocks. We summarize optimal policies with implied nonlinear impulse response functions. We define an impulse response of variable X_t to unexpected shock $\mathcal{E}_k$ of size Δ in a particular period $k \geq t+1$ (often $k = t+1$) as

$$\mathbb{E}[X_t|\Omega_{-1}, \Theta_{-1}, \Phi_{-1}, \mathcal{E}_k = \Delta] - \mathbb{E}[X_t|\Omega_{-1}, \Theta_{-1}, \Phi_{-1}, \mathcal{E}_k = 0]$$

where $\Omega_{-1}, \Theta_{-1}, \Phi_{-1}$ are time 0 states for the Ramsey planner and conditional mathematical expectations are taken over the ensemble of paths generated by iterations on the optimal policy functions. We approximate conditional expectations by taking averages over $N = 1000$ simulations of sample paths. Below, we typically set Δ to be one standard deviation of $\mathcal{E}$. Impulse response functions are state-dependent and non-linear in sizes of shocks. We report impulse responses for several values of the shock arrival date k.

TABLE IV: WELFARE DECOMPOSITION

	Efficiency	Redistribution	Insurance
Baseline			
(a) Optimal monetary policy	-122	9	213
(b) Optimal monetary and fiscal policy	-16	1	115
Extensions			
(c) Liquidity Frictions	-78	-4	182
(d) Mutual Fund	-154	-12	266
(e) Heterogeneous labor income exposures	-327	-7	334
Alternative Pareto Weights			
(f) High Labor Tax	-180	-125	405
(g) High Bond Tax	-115	52	163
(h) High Dividend Tax	-165	-1	266

Notes: We decompose welfare differences between optimal HANK and optimal RANK policies using the Bhandari et al. (2021) procedure. For all cases, the point of comparison (optimal RANK policy) is set so that expected levels of policy variables equal their optimal HANK counterparts in the absence of aggregate risk and stochastic processes for deviations of the policy variables from their means are optimal in the representative agent version. Lines (a) and (b) report results for our baseline calibration applied to both monetary and monetary-fiscal policies. Lines (c), (d), (e) report our decomposition of the optimal monetary policy for extensions that we describe in sections 6.2, 6.3, 6.4, respectively. Lines (f), (g), (h) consider alternative specifications of Pareto weights discussed in section 6.1.

5.2.1 Monetary policy responses to a markup shock

We describe an optimal monetary policy response to a negative innovation in $\mathcal{E}_{\Phi,t}$. Because this shock increases the desired markup $1/(\Phi_t - 1)$, we call it a positive markup shock. Figure II plots optimal responses of the nominal interest rate, inflation, the real pre-tax wage per unit of effective labor, and real output to a positive markup in period one.

Figure II shows that optimal responses in the HANK and RANK economies differ significantly. While the RANK Ramsey planner slightly *increases* nominal interest rates in response to a markup shock, the HANK planner aggressively *cuts* them. The response of inflation in HANK is an order of magnitude larger, and paths of real wages and output are temporarily above their RANK counterparts.

To illustrate how the insurance considerations drive the shapes of the impulse responses, we construct responses in intermediate economies located between HANK and RANK. We (i) start with our calibrated HANK economy (plotted as a solid blue line labeled as "HANK"), (ii) shut down idiosyncratic shocks (plotted as a dashed line with square markers labeled as "HANK No Idio. risk"), (iii) allow agents to trade Arrow securities contingent on aggregate shocks (plotted as a dashed line with circle markers labeled as "HANK CM"), and finally (iv) shut down heterogeneity in initial productivities and assets to obtain our RANK economy (plotted as a solid red line labeled as "RANK"). This procedure allows us to isolate contributions from providing insurance against idiosyncratic shocks by comparing responses of economy (ii) with those of economy (i); contributions from providing insurance against aggregate shocks by comparing the responses of economy (iii) with those of economy (ii); and contributions from redistribution by comparing responses in economy (iv) with those of economy (iii).

Before interpreting figure II, note that when lump-sum transfers are available, complete and incomplete market versions of RANK are identical. Not so with HANK: the presence of heterogeneity means that the absence of complete markets puts concerns about insurance into the mind of a HANK Ramsey planner. Monetary policy can't provide redistribution or insurance against idiosyncratic shocks beyond what taxes $\bar{\Upsilon}$ can do. However, monetary policy *can* provide insurance against aggregate shocks and imperfectly substitute for missing markets in Arrow securities.

The decomposition in figure II shows that nearly all differences in policy responses between HANK and RANK are driven by the planner's wish to provide insurance against aggregate shocks. To understand how motives to provide insurance account for key differences between RANK and HANK economies, it is helpful to study a one-time, fully transient positive markup shock. The textbook effect of this shock is an increase in inflationary pressure that monetary policy can offset by depressing marginal costs. Since marginal costs are proportional to aggregate demand, a contractionary increase in nominal rates is optimal in the RANK economy. Galí (2015) calls this "leaning against the wind."

Such a one-time markup shock also changes the mix of factor payments by increasing dividends and lowering wages. When households are homogeneous as they are in the RANK economy, or when they can trade Arrow securities *ex-ante*, a change in the composition of firms' payments does not affect welfare: agents who receive mainly wage income will hold a portfolio of Arrow securities that payoff when their wage income is low. When Arrow securities are missing, monetary policy can improve welfare by providing insurance against aggregate shocks. To offset a drop in labor income, the Ramsey planner sets interest rates to *increase* real wages. Since real wages are firms' marginal costs, a monetary policy action that provides insurance is opposite to one that promotes price stability.

The net effect of a markup shock on optimal monetary policy depends on relative strengths of the Ramsey planner's motives to provide price stability and insurance. The cost of inflation is set by the price adjustment cost parameter ψ, while insurance provision motives depend on inequalities in stock ownership relative to inequalities in wage income. If stock holdings are perfectly aligned with labor earnings in the sense that the share of stocks that each person owns equals his or her share of aggregate labor compensation, then the insurance provision motive vanishes; a positive effect on dividends exactly offsets the negative effect on labor earnings. In that case, optimal responses are similar to those in RANK.[26] HANK responses in figure II differ so much from responses that would support price stability because stock holdings are much more skewed than labor earnings in U.S. data, making insurance concerns the principal motive in our calibration.

Figure II shows that the most efficient way to provide insurance is to *front-load* it: virtually all differences in HANK and RANK optimal outcomes occur in policy decisions in period $t = 1$. Households can borrow and lend freely so that their utilities depend only on present values of factor payments, not their timing. But price-setting firms care about the path of factor prices. A first-order approximation of a firm's optimality condition (16) is written as

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + const \cdot \hat{w}_t + const \cdot (-\ln \Phi_t)\,, \tag{38}$$

where π_t is log inflation, $\hat{w}_t$ is the log deviation of the wage per unit of effective labor from its average level. Solving difference equation (38) forward shows that inflation in period t is proportional to the discounted present value of future wages. As a consequence, an increase in $\hat{w}$ in some period k increases inflation in all $t \leq k$. Therefore, to minimize costs of inflation it is optimal to deliver all of the adjustment in the present discounted value of unit labor costs at the moment the shock arrives.

[26]The RANK economy is a special case in which both shares are equal to one, but the same result holds in an economy with heterogeneity so long as shares of dividends and earnings are aligned. See the online appendix for an illustration.

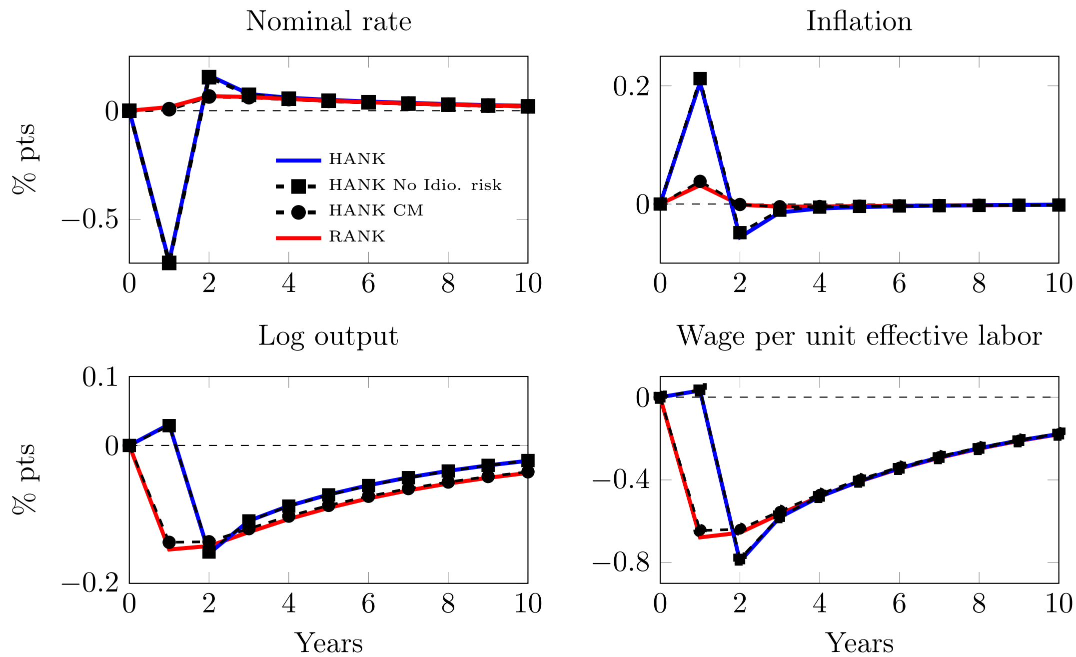

Figure II: Optimal monetary response to a markup shock. The bold blue and red lines are the calibrated HANK and RANK responses respectively. The dashed black lines with squares and circles are responses under HANK with idiosyncratic shocks shut down and with complete markets, respectively.

An optimal response to a negative markup shock is virtually a mirror image of an optimal response to a positive markup shock. Averaged over time, the expected net flow of resources to each agent generated by a monetary policy response is approximately zero; an outcome consistent with optimal responses being driven mainly by the planner's insurance motive and not redistribution.

5.2.2 Monetary-fiscal responses to a markup shock

We now study a Ramsey planner who chooses monetary and fiscal policies. Figure III shows optimal responses to the markup shock. The planner offsets the shock by combining a labor subsidy with a dividend tax while holding the nominal interest rate unchanged.

In the RANK economy, lump-sum taxes are non-distortionary while taxes on dividend and interest income are redundant. Under the optimal monetary-fiscal policy, the planner achieves a first best by setting $\Upsilon_t^n = -1/\Phi_t$ to offset monopoly distortions, setting a path of nominal rates that delivers a constant price level and setting lump-sum taxes to satisfy the government's budget constraint.

In the HANK economy, the burden of lump-sum taxes falls disproportionately on poor households. The RANK prescription of a proportional labor subsidy financed by reducing

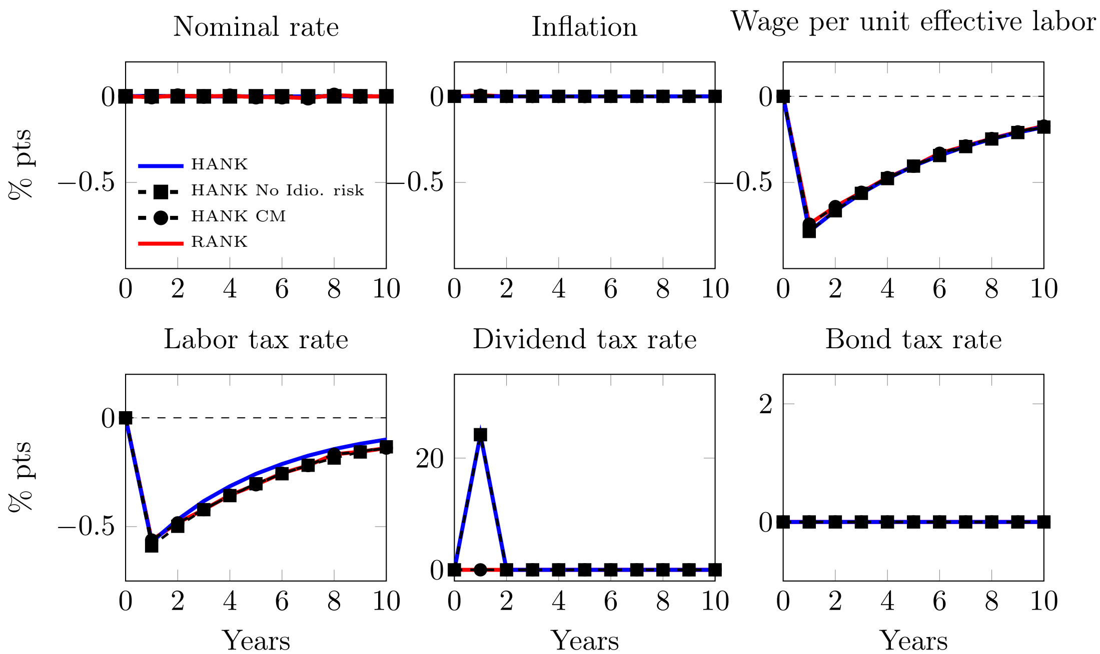

Figure III: Optimal monetary-fiscal response to a markup shock. The bold blue and red lines are the calibrated HANK and RANK responses, respectively. The dashed black lines with squares and circles are responses under HANK with idiosyncratic shocks shut down and with complete markets, respectively.

lump-sum transfers is therefore not optimal. Instead, a HANK Ramsey planner finances the labor income subsidy by levying a one-time tax on dividends. This tax exactly offsets the gains that stock owners receive from the higher markups and thus the policy response provides complete insurance against the markup shock.

5.2.3 Optimal responses to productivity shocks

Figure IV shows optimal monetary response to a negative TFP shock. HANK responses differ significantly from RANK; the decomposition reveals that differences are once again driven by the absence in HANK of a market for insuring against aggregate shocks.

To understand what motivates the Ramsey planner to provide insurance, first observe that an adverse TFP shock reduces both profits and labor earnings, so firm owners and workers both suffer. A TFP shock affects agents differently when they hold different quantities of bonds: agents who own substantial quantities of bonds suffer less than do otherwise identical agents who are debtors. Market incompleteness prevents borrowers and savers from directly hedging TFP shocks. Monetary policy fills this gap by lowering returns on debt, thereby transferring resources from savers to borrowers and smoothing relative consumption shares.

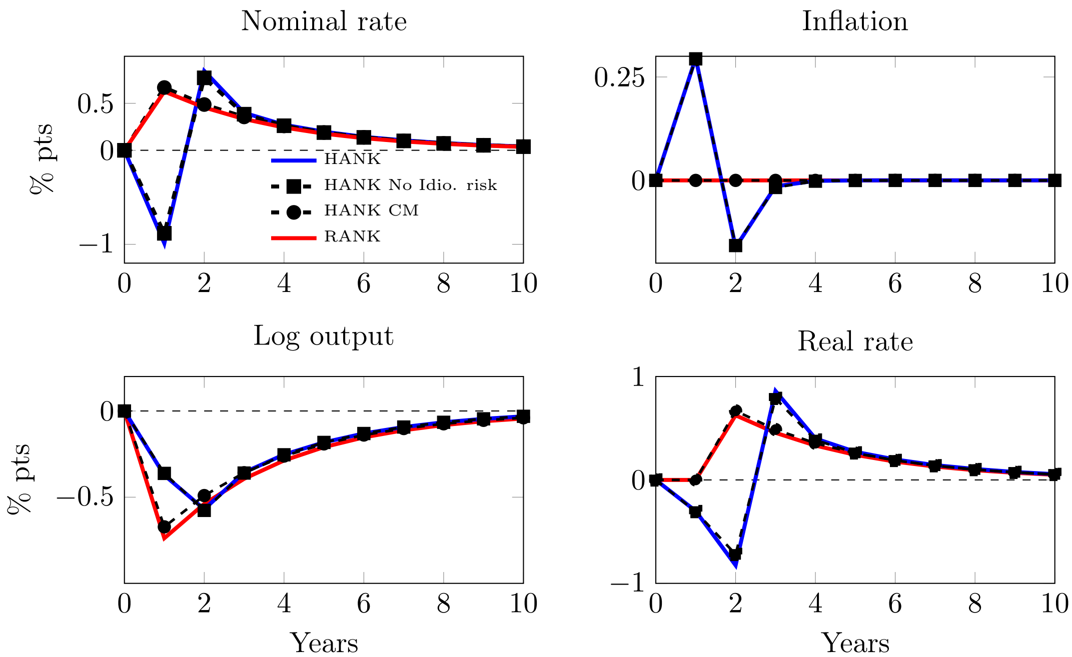

Figure IV: Optimal monetary response to a TFP shock. The bold blue and red lines are the calibrated HANK and RANK responses, respectively. The dashed black lines with squares and circles are responses under HANK with idiosyncratic shocks shut down and with complete markets, respectively.

The optimal policy response evidently departs from RANK and, more generally, from the prescription that monetary policy should aim to minimize fluctuations in inflation and an "output gap." In response to TFP shocks, by setting the interest rate to a "natural rate" that would prevail if prices were flexible, a planner can eliminate fluctuations in both inflation and the output gap. Optimal responses in the RANK economy and in the complete market version of a HANK economy both follow this prescription.[27] That prescription is not optimal in the incomplete markets economy because it does not provide insurance to borrowers and savers. To transfer resources from savers to borrowers, the planner lowers *ex post* real returns on debt by engineering surprise inflation and pushing the expected real rate of interest below the natural rate. A lower real rate requires temporarily higher output. A higher output triggers more inflation, as indicated by equation (38). To offset extra inflationary pressure, the Ramsey plan sets the stage for deflation at $t = 2$.

The strength of planner's insurance motive depends on heterogeneity in holdings of nominal bonds. U.S. data indicate considerable heterogeneity in nominal asset holdings, which explains a big difference between an optimal policy in the HANK economy relative to the one in RANK.

[27] Blanchard and Galí (2007) refer to this property of New Keynesian models as a "divine coincidence."

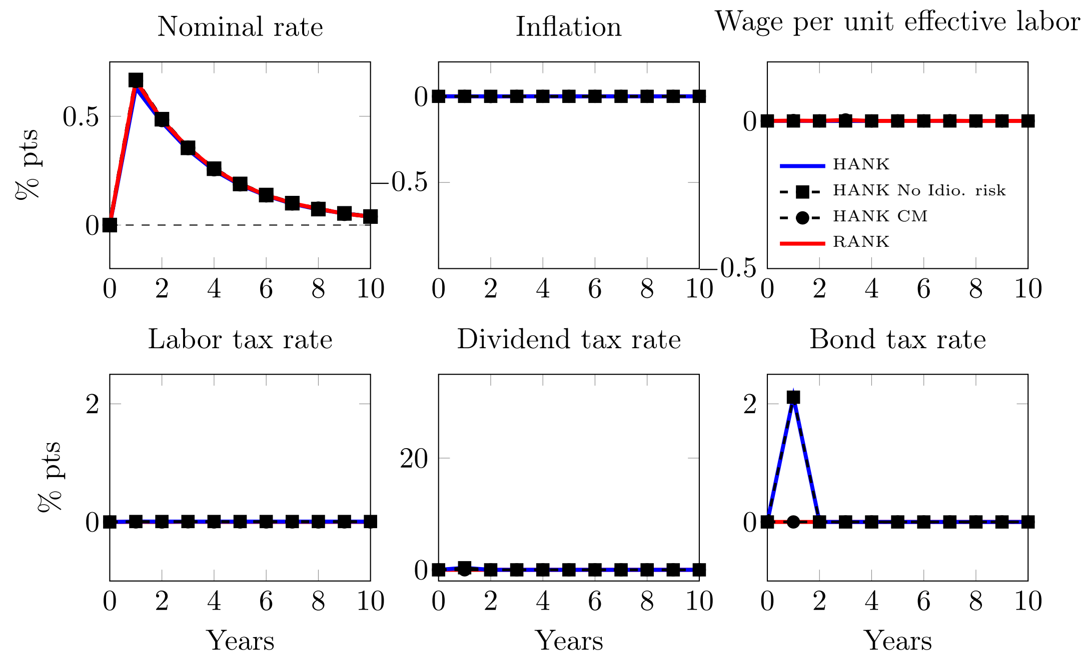

Figure V: Optimal monetary-fiscal response to a TFP shock. The bold blue and red lines are calibrated HANK and RANK responses, respectively. The dashed black lines with squares and circles are responses under HANK with idiosyncratic shocks shut down and with complete markets, respectively.

If fiscal policy can be adjusted in response to the TFP shock, then the Ramsey plan provides optimal insurance to borrowers and savers directly via a state-contingent tax on interest income. That tax effectively completes markets and brings together optimal policy responses for HANK and RANK economies. (See figure V).

6 Extensions and robustness

We describe several extensions to the baseline calibration and environment.

6.1 Roles of key parameters

Optimal responses vary with parameters that shape a trade-off between price stabilization and insurance motives. The strength of a price stabilization motive is driven by the price adjustment cost parameter ψ. The strength of the planner's insurance motive depends on relative sizes of post-tax inequalities in dividends and labor earnings. These inequalities, in turn, depend on the joint distribution of assets and labor productivities and on tax rates $\bar{\Upsilon}$ that are pinned down by initial Pareto weights. In this section, we study how optimal monetary policy responses vary with these objects. We report main findings here and details in the online appendix.

We set the value of ψ according to estimates of Sbordone (2002). In a staggered price adjustment model, her numbers indicate that firms change prices on average every nine months. Studies using micro evidence on price changes (see Nakamura and Steinsson 2013) recover estimates that range from 6.8 to 12 months. In the online appendix, we study the sensitivity of our results to variations in ψ between half to twice of our baseline value. We find on-impact changes in the nominal rates that are fairly similar to the baseline, while the peak response of inflation varies roughly linearly in ψ over this range. In addition, we set $\psi \approx 0$ to study a benchmark with fully flexible prices.

Next we study sensitivity of outcomes to Pareto weights. In section 4, we assigned Pareto weights using the exponential specification presented in equation (37). This specification maps a three dimensional vector δ (loadings on the three dimensions of initial heterogeneity) to optimal tax rates $\bar{\Upsilon}(\delta)$. Our baseline calibration set $\bar{\Upsilon}(\delta) = \bar{\Upsilon}^{US}$. We explore the dependence of policy on Pareto weights by raising each component of $\bar{\Upsilon}$ to 50%, one at a time, and computing the optimal responses for the corresponding δ. In table IV, lines (f), (g) and (h) report the welfare decompositions for the optimal policy with higher labor, bond, and dividend taxes respectively. In line with the benchmark model, the vast majority of welfare gains are generated by insurance, which comes at the cost of aggregate efficiency. Int the online appendix, we corroborate this fact by showing that the impulse responses are quantitatively similar to the baseline responses in figures II and IV for each of these cases as well as alternative experiments where we lower the tax rates from their U.S. values.

Since inequality drifts over time in our economy, optimal responses depend on time t. In our calibrated economy, the drift is slow and therefore responses in $t = 25$ and $t = 50$ appear to be very similar to those at $t = 1$. When we initialize a Ramsey plan at an approximation to a long-run distribution of assets and productivities gleaned from simulating a competitive equilibrium for 100 periods, we find responses that are approximately the same as those in the baseline model. This outcome reflects a balance of two forces. On the one hand, the passage of time decreases the correlation between stock holdings and labor earnings, which renders inequality more misaligned. That increases the planner's gains from providing insurance. On the other hand, the correlation between shares of equities and bond holdings decreases, which diminishes gains from insuring using unanticipated inflation.

6.2 Liquidity frictions

Thus far, we have assumed that households have unrestricted access to risk-free bond markets. Empirical work documenting large marginal propensities to consume points to presence of liquidity constrained households (see, for instance Jappelli and Pistaferri 2014 and Johnson et al., 2006). In this section, we investigate how liquidity frictions affect optimal monetary policy. We augment our model with "hand-to-mouth" agents who own equities and fixed amounts of nominal assets. They can consume dividends, interest on their nominal bond holdings, and their labor income; but they cannot trade financial assets. Therefore, the budget constraint of a hand-to-mouth agent satisfies equation (2) subject to the restriction that the market value of nominal debt holdings $P_t Q_t b_{i,t}$ must be constant over time. Thus, we modify our baseline model to include an additional dimension of permanent heterogeneity–namely an indicator variable $h_i \in \{0, 1\}$, where $h_i = 1$ if the agent is a hand-to-mouth type and $h_i = 0$ if the agent is not a hand-to-mouth type. All other aspects of the baseline model remain the same, including how we set initial conditions.

To calibrate a distribution of hand-to-mouth agents, we need data on individual marginal propensities to consume (MPCs) broken down by observable characteristics that we can map to our model. Such data for the U.S. are not readily available. However, using Italian data, Jappelli and Pistaferri (2014) measure average MPCs by cash-in-hand (defined as financial wealth plus current period wage minus taxes). Their findings are broadly consistent with Kaplan et al. (2014) and Kaplan et al. (2018), who, based on U.S. data, incorporate substantial heterogeneity among liquidity constrained agents. In view of such evidence, we incorporate liquidity-constrained agents by binning households into cash-in-hand quantile groups and, for each quantile group randomly assign a hand-to-mouth status to calibrate the model-generated MPC gradient with respect to cash-in-hand. This approach also preserves the distribution of real and nominal claims of the baseline model, which presents the same insurance motives to a Ramsey planner and, therefore, allows us to isolate effects on optimal policies that are attributable to the 'liquidity frictions' that constrain hand-to-mouth agents.

Figures VI and VII show optimal monetary responses to markup and aggregate TFP shocks for the calibration with hand-to-mouth agents and compare them to optimal responses in the baseline calibration. Evidently, the trading frictions that give rise to heterogeneities in MPCs make paths for nominal rates, real wages, and inflation smoother than they are in the baseline model. Since Ricardian equivalence no longer holds, an optimal path of transfers is now uniquely pinned down.

We show next that deviations of optimal policy responses from those in our baseline model are driven by the inability of hand-to-mouth agents to borrow and save in order to smooth consumption over time, as well as the substantial heterogeneity within the set of liquidity constrained households. We start with optimal monetary policy responses to a markup shock in figure VI. In the baseline calibration (solid blue line), the planner provides insurance against the shock by front-loading higher wages and avoiding additional future inflation from firms' rationally anticipating higher future marginal costs. Such a front-loading policy would be costly for hand-to-mouth agents because they would have too much income (equal to consumption) in the short run relative to the future. Thus, in addition to providing insurance, the Ramsey planner wants to smooth over time the consumption paths of hand-to-mouth agents.

A natural way to achieve insurance and also to smooth consumption would be to adjust the timing of lump-sum transfers. But hand-to-mouth agents are not homogeneous. A path of transfers that would smooth the consumption of poor hand-to-mouth agents, who rely mainly on wage income, would exacerbate the volatility of consumption of rich hand-to-mouth agents, who rely mostly on their dividend income. Thus, heterogeneity among liquidity constrained agents makes transfers a less effective tool.[28]

In addition to timing transfers, the planner distorts allocations to induce a smoother path for the real wage (dashed black line). A smoother path means a real wage above a "natural wage" that would prevail with flexible prices when markups are high and below that "natural wage" when the economy recovers. Implementing such a path for real wages requires expansionary monetary policy and associated inflation upon arrival of the shock to be followed by a contractionary monetary policy that brings persistent deflation.

Figure VII displays optimal responses to a productivity shock. Overall, the policy actions help approximate within-agent transfers that would occur if all agents, including those who are hand-to-mouth, were free to trade assets. The TFP shock widens disparities in total income between hand-to-mouth agents who are borrowers and those who are lenders. As in the baseline, the planner inflates away debt in the short run, thereby transferring resources from lenders to borrowers during a recession. In addition, the planner engineers a persistent but small deflation. A higher price level acts like a tax on wealth by lowering obligations of hand-to-mouth borrowers and reducing assets of hand-to-mouth lenders. Since the shock is transitory, agents would want to reverse that reshuffling of resources after the shock wears off. Since they cannot trade, the planner uses a post-shock deflation to generate a smooth path of repayments from borrowers to asset holders as TFP reverts to its steady-state value. Like responses to markup shocks, these paths of prices and real interest rates require lower nominal interest rates for a few periods, followed by high nominal interest rates, later making the path of nominal rates smooth relative to outcomes in the baseline model.[29]

The preceding discussion suggests that the persistence of individual MPC, as well as their dispersion, has important quantitative implications for optimal policy. The calibration

[28] To highlight this point and isolate the role of wealthy hand-to-mouth consumers, in the online appendix, we study optimal policy in an alternative calibration in which the bottom 15% of the cash-in-hand distribution is set to be hand-to-mouth. This example has the flavor of typical calibrations of one-asset-Aiyagari models because constrained hand-to-mouth agents are more homogeneous and depend almost entirely on their labor incomes. We show that the dynamics of interest rates, inflation, output and wages are almost identical to those of the baseline economy with no liquidity frictions. When liquidity-constrained agents are homogeneous, the planner can effectively borrow on their behalf and use transfers to smooth their consumption.

[29] In the online appendix, we report results for an extension that allows marginal propensity to consume to vary across sources of income. The general principles guiding the optimal policy are unchanged.

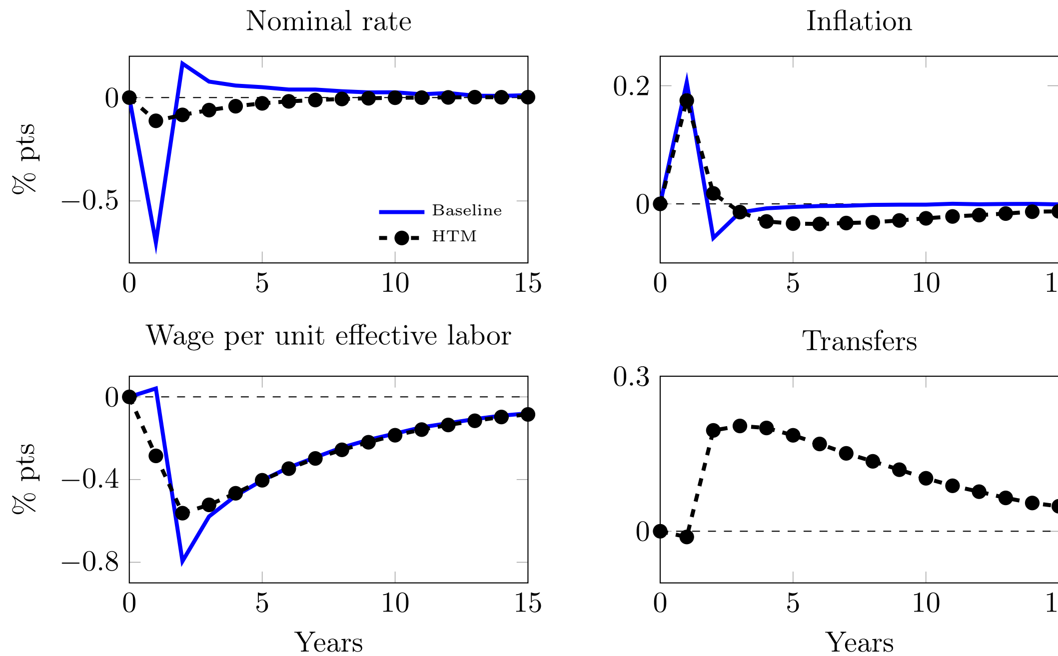

Figure VI: Optimal monetary response to a markup shock with liquidity frictions. The bold blue lines are responses under the baseline without hand-to-mouth agents; the dashed black lines with circles are responses with hand-to-mouth agents. Transfers are not plotted for the baseline because Ricardian equivalence holds and timing of transfers is indeterminate.

in this section implicitly assumes that individual MPCs are permanent. The more frequently the identities of liquidity constrained agents switch, the smaller the persistence will be. Less persistence would reduce motives for the planner to smooth insurance benefits over time and bring impulse responses closer to the baseline model. Empirical work on persistence in MPCs is still in early stages, although Auclert et al. (2018) and Kaplan et al. (2018) provide valuable insights into MPC dynamics.

6.3 Mutual Fund

In the baseline model, we calibrated households' portfolios to counterparts in the SCF but imposed that households could not trade claims to dividends. In this section, we follow Gornemann et al. (2016) by having agents trade shares in a mutual fund. A competitive mutual fund sector invests in corporate equity and government bonds and remits after-tax earnings to households in proportion to their holdings of the mutual fund. Shares in the mutual fund are indirect claims to returns from a common financial portfolio and are traded by all households in a competitive market.

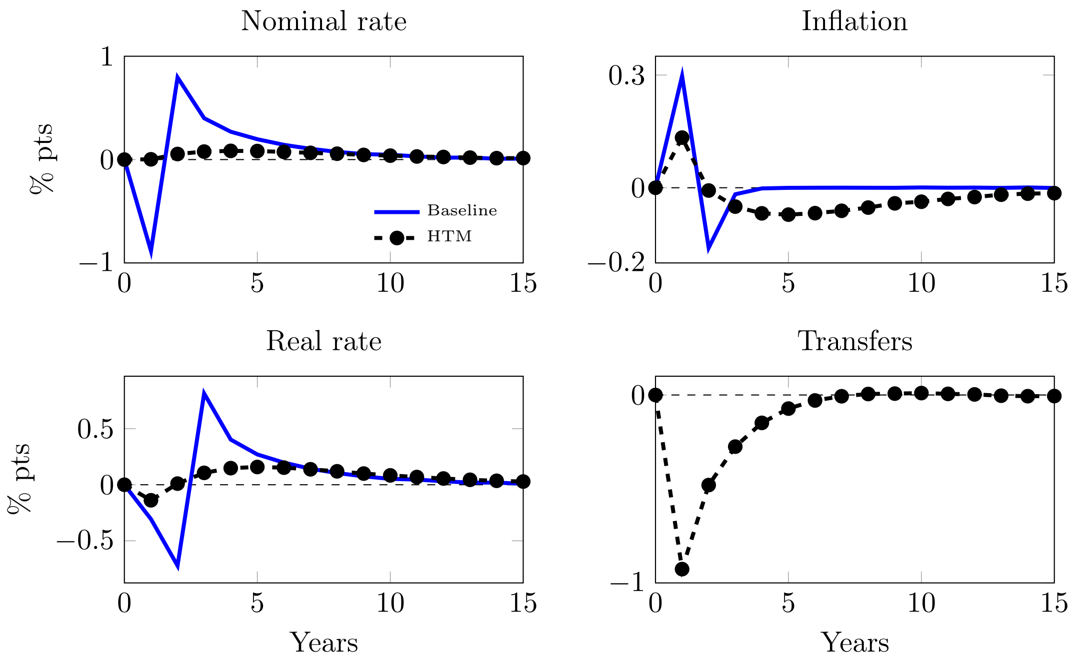

Figure VII: Optimal monetary response to a TFP shock in an economy with liquidity frictions. The bold blue lines are responses under the baseline without hand-to-mouth agents; the dashed black lines with circles are responses when hand-to-mouth agents are present. Transfers are not plotted for the baseline because Ricardian equivalence holds and the timing of transfers is indeterminate.

The mutual fund solves

$$\max_{B_t} \mathbb{E}_0 \sum_t S_t^{mf} D_{a,t}$$

$$Q_t B_t + D_{a,t} = \frac{\left(1 - \Upsilon_t^b\right) B_{t-1}}{1 + \Pi_t} + (1 - \Upsilon_t^d) D_t,$$

where we follow Gornemann et al. (2016) and set $\frac{S_{t+1}^{mf}}{S_t^{mf}}$ to be an asset-weighted average of intertemporal marginal rates of substitutions across households. Households' budget constraint (2) becomes

$$c_{i,t} + P_{a,t} a_{i,t} = (1 - \Upsilon_t^n) W_t \epsilon_{i,t} n_{i,t} + T_t + \left(D_{a,t} + P_{a,t}\right) a_{i,t-1}, \tag{39}$$

where $a_{i,t}$ are household i's holdings of the mutual fund and $\int a_{i,t} di = 1$. Households freely trade $a_{i,t}$. The production side of the model is unchanged.

We study optimal responses to markup and TFP shocks and contrast them with our baseline model. Our calibration in this mutual fund setting closely follows section 4. We initialize the distribution of mutual fund holdings by using the distribution of financial wealth from the SCF 2007 formed by summing claims to all bonds and stocks; we again set

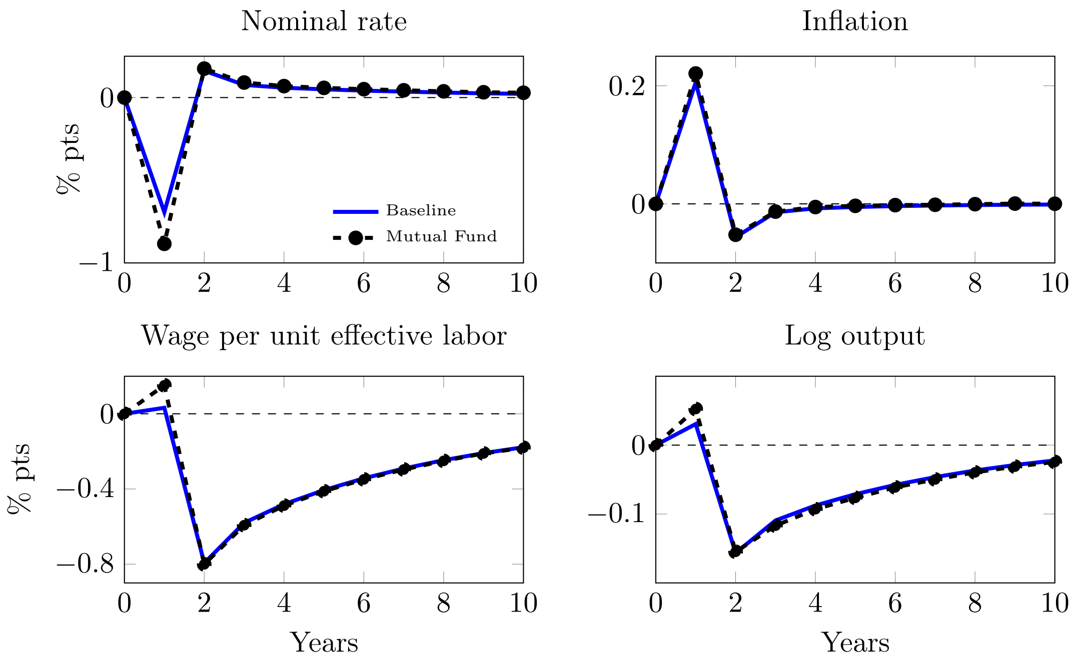

Figure VIII: Optimal monetary responses to a markup shock with mutual fund. The bold blue lines are responses under the baseline and the dashed black lines with circles are responses under the mutual fund setting.

Pareto weights to rationalize observed U.S. average tax rates. Other parameters are those in the baseline model. When we study optimal monetary policy, we impose $B_t = B_0$.[30]

Figure VIII plots optimal monetary policy responses to a markup shock. Responses in the baseline model and those for the model with a mutual fund are very close because optimal policy is driven largely by insurance motives largely drive optimal policy, as discussed extensively in section 5. Magnitudes of optimal policy responses are determined by cross-sectional heterogeneity in exposures of labor and non-labor incomes to aggregate shocks. Even after we sum bond and stock claims, total financial wealth remains quite skewed relative to labor earnings. As was the case before, the planner provides insurance in response to markup shocks by boosting the present discounted value of wages. In the online appendix, we show that responses to the TFP shock are also similar to those in the baseline. In the row labeled "Mutual fund" in table IV, we show that insurance considerations drive most of the welfare gains.

[30] This is mainly done to assure comparability of results. Different from the baseline, in which we had Ricardian equivalence, the planner in the mutual fund economy is motivated to vary the level of debt and change the riskiness of returns on the mutual fund. Setting the debt level to a constant imposes parity between the baseline monetary planner and the mutual fund monetary planner in their abilities to affect returns. We relax this restriction when we study optimal monetary-fiscal policies in a mutual fund economy in which the planner can use taxes on bonds income or dividends to directly affect returns. See the online appendix.

6.4 Heterogeneous labor income exposures

In our baseline calibration, percentage falls in labor income during recessions are the same across workers. Using administrative data on W2 forms over the period 1978-2010, Guvenen et al. (2014) document that relative to a typical worker, individuals who have either low past incomes or very high past incomes face larger drops in earnings in recessions. In this section, we compute optimal monetary and fiscal responses under a richer stochastic process for idiosyncratic risk that captures the Guvenen et al. (2014) patterns.

We modify equation (12) to

$$\ln \epsilon_{i,t} = (1 + f(\theta_{i,t-1})) \ln \Theta_t + \ln \theta_{i,t} + \varepsilon_{\epsilon,i,t}, \tag{40}$$

and set $f(\theta)$ so that the aggregate productivity shock has different loadings for agents with different earning histories. We assume a quadratic function $f(\theta) = f_0 + f_1\theta + f_2\theta^2$ and normalize f_0 so that an agent with median productivity faces a drop similar to the drop in aggregate TFP. We then simulate a competitive equilibrium for 30 periods and extract "recessions" as consecutive periods in which the growth rate of output falls one standard deviation below zero. Following the empirical procedure in Guvenen et al. (2014), we rank workers by percentiles of their average log labor earnings 5 years prior to the shock and compute the percent earnings loss for each percentile relative to the median. We set parameters f_1 and f_2 to match earnings losses of the 5th and 95th percentiles.

In figure IX, we report the optimal monetary policy response to a TFP shock with heterogeneous exposures. Amplified inequality induced by a recession increases gains that the planner earns from providing insurance. As compared to our baseline monetary response, the planner further lowers the nominal rate and thereby induces higher inflation in the short run and a lower ex-ante real rate.[31]

7 Concluding Remarks

We forged a method to approximate Ramsey plans in economies with heterogeneous agents and used it to reassess quantitative lessons for monetary and fiscal policy brought by earlier contributions to New Keynesian economics. Heterogeneity adds an insurance motive that quantitatively dominates the motives to stabilize nominal prices that have typically driven New Keynesian policy prescriptions. For our laboratory, we combined basic versions

[31] In the online appendix, we plot the monetary-fiscal responses. With optimal monetary as well as fiscal policy, the planner responds by increasing labor income taxes in addition to the tax on bond income. For reasons similar to Werning (2007), the marginal cost of extracting resources from high-income agents is lower in times of higher inequality. Therefore, optimal labor tax rates are higher on impact and then revert as effects of inequality shocks decay to a permanently higher level.

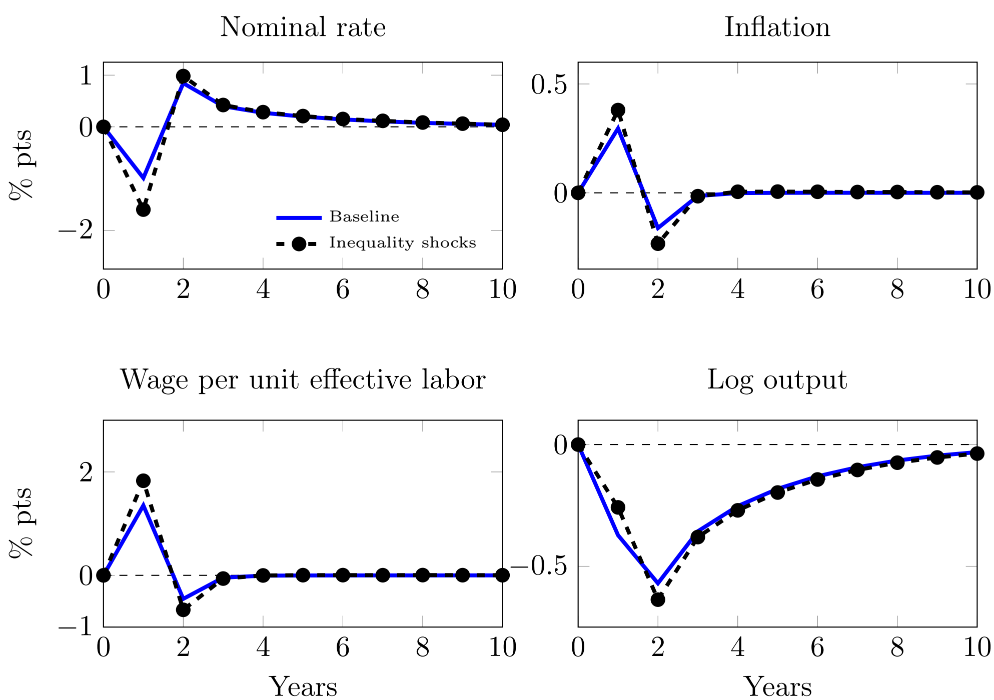

Figure IX: Optimal monetary responses to a TFP shock with heterogeneity in labor income exposures. The bold blue lines are responses under the baseline, and the dashed black lines with circles are responses when the idiosyncratic productivity process has heterogeneous exposures to aggregate TFP.

of the New Keynesian model and the incomplete market models. We are convinced that our method will be useful for computing optimal policies in environments that have more detailed household balance sheets, richer labor market dynamics that include realistic wage-setting frictions, and asset markets formulations capable of fitting observed returns. Adding these features is likely to affect how a Ramsey planner would deliver insurance in order to supplement missing markets and financial frictions. Nevertheless, we suspect that insurance concerns with respect to aggregate shocks and other determinants of heterogeneous income exposures will remain decisive determinants of optimal policies.

References

Acharya, Sushant, and Keshav Dogra. 2018. "Understanding HANK: Insights from a PRANK." *FRB of New York Staff Report*(835): .

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *The Quarterly Journal of Economics*, 109(3): 659–684.

Aiyagari, S. Rao, Albert Marcet, Thomas J. Sargent, and Juha Seppala. 2002. "Optimal Taxation without State-Contingent Debt." *Journal of Political Economy*, 110(6): 1220–1254.

Anderson, Evan W, Lars Peter Hansen, and Thomas J Sargent. 2012. "Small noise methods for risk-sensitive/robust economies." *Journal of Economic Dynamics and Control*, 36(4): 468–500.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2018. "The Intertemporal Keynesian Cross." Working Paper 25020, National Bureau of Economic Research.

Barro, Robert J. 1979. "On the Determination of the Public Debt." *Journal of Political Economy*, 87(5): 940–971.

Barro, Robert J., and Charles J. Redlick. 2011. "Macroeconomic Effects From Government Purchases and Taxes." *Quarterly Journal of Economics*, 126(1): 51–102.

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

Benhabib, Jess, Alberto Bisin, and Mi Luo. 2019. "Wealth Distribution and Social Mobility in the US: A Quantitative Approach." *American Economic Review*, 109(5): 1623–47.

Bewley, Truman. 1980. "The Optimum Quantity of Money." In *Models of Monetary Economies.* eds. by John H. Kareken, and Neil Wallace, Minneapolis, Minnesota: Federal Reserve Bank of Minneapolis, 169–210.

Bewley, Truman F. 1977. "The Permanent Income Hypothesis: A Theoretical Formulation"." *Journal of Economic Theory*, 16(2): , p. 252?92.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas Sargent. 2017. "Fiscal Policy and Debt Management with Incomplete Markets." *Quarterly Journal of Economics*, 132(2): 617–663.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas Sargent. 2021. "Efficiency, Insurance, and Redistribution Effects of Government Policies." working paper.

Bhandari, Anmol, and Ellen McGrattan. 2019. "Sweat Equity in U.S. Private Business."

Bilbiie, Florin O. 2019. "Monetary Policy and Heterogeneity: An Analytical Framework." *Unpublished manuscript.*

Bilbiie, Florin Ovidiu, and Xavier Ragot. 2017. "Optimal Monetary Policy and Liquidity with Heterogeneous Households." CEPR Discussion Papers 11814, C.E.P.R. Discussion Papers.

Blanchard, Olivier, and Jordi Galí. 2007. "Real Wage Rigidities and the New Keynesian Model." *Journal of Money, Credit and Banking*, 39 35–65.

Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy*, 114(5): 835–870.

Challe, Edouard. 2017. "Uninsured Unemployment Risk and Optimal Monetary Policy." Working Paper 2017-54, Center for Research in Economics and Statistics.

Chari, V. V., and Patrick J. Kehoe. 1999. "Chapter 26 Optimal fiscal and monetary policy." In *Handbook of macroeconomics.* eds. by John B Taylor of Macroeconomics, and Michael Woodford B T Handbook, Volume 1,: Elsevier, 1671–1745.

Childers, David. 2018. "On the Solution and Application of Rational Expectations Models with Function-Valued States." *Unpublished manuscript.*

De Nardi, Mariacristina. 2004. "Wealth Inequality and Intergenerational Links." *The Review of Economic Studies*, 71(3): 743–768.

Debortoli, Davide, and Jordi Galí. 2017. "Monetary Policy with Heterogeneous Agents: Insights from TANK models." working papers, Department of Economics and Business, Universitat Pompeu Fabra.

Den Haan, Wouter J. 2010. "Comparison of solutions to the incomplete markets model with aggregate uncertainty." *Journal of Economic Dynamics and Control*, 34(1): 4 – 27, Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty.

Doepke, Matthias, and Martin Schneider. 2006. "Inflation and the Redistribution of Nominal Wealth." *Journal of Political Economy*, 114(6): 1069–1097.

Evans, David. 2015. "Perturbation Theory with Heterogeneous Agents: Theory and Applications." Ph.D. dissertation, New York University.

Farhi, Emmanuel. 2010. "Capital Taxation and Ownership When Markets Are Incomplete." *Journal of Political Economy*, 118(5): 908–948.

Fleming, Wendell H. 1971. "Stochastic Control for Small Noise Intensities." *SIAM Journal on Control*, 9(3): 473–517.

Fleming, Wendell H, and PE Souganidis. 1986. "Asymptotic series and the method of vanishing viscosity." *Indiana University Mathematics Journal*, 35(2): 425–447.

Floden, Martin. 2001. "The effectiveness of government debt and transfers as insurance." *Journal of Monetary Economics*, 48(1): 81–108.

Galí, Jordi. 2015. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications.*: Princeton University Press.

Galí, Jordi, Mark Gertler, and J. David Lopez-Salido. 2007. "Markups, Gaps, and the Welfare Costs of Business Fluctuations." *The Review of Economics and Statistics*, 89(1): 44–59.

Gornemann, Nils, Keith Kuester, and Makoto Nakajima. 2016. "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy." International Finance Discussion Papers 1167, Board of Governors of the Federal Reserve System (U.S.).

Greenwald, Daniel L, Martin Lettau, and Sydney C Ludvigson. 2014. "Origins of Stock Market Fluctuations." Working Paper 19818, National Bureau of Economic Research.

Guvenen, Fatih, Serdar Ozkan, and Jae Song. 2014. "The Nature of Countercyclical Income Risk." *Journal of Political Economy*, 122(3): 621–660.

Huggett, Mark. 1993. "The risk-free rate in heterogeneous-agent incomplete-insurance economies." *Journal of Economic Dynamics and Control*, 17(5): 953–969.

Jappelli, Tullio, and Luigi Pistaferri. 2014. "Fiscal Policy and MPC Heterogeneity." *American Economic Journal: Macroeconomics*, 6(4): 107–36.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review*, 96(5): 1589–1610.

Judd, Kenneth L, and Sy-Ming Guu. 1993. "Perturbation solution methods for economic growth models." In *Economic and Financial Modeling with Mathematica®.*: Springer, 80–103.

Judd, Kenneth L, and Sy-Ming Guu. 1997. "Asymptotic methods for aggregate growth models." *Journal of Economic Dynamics and Control*, 21(6): 1025–1042.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti. 2010. "Investment shocks and business cycles." *Journal of Monetary Economics*, 57(2): 132 – 145.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

Kaplan, Greg, Giovanni L Violante, and Justin Weidner. 2014. "The Wealthy Hand-to-Mouth." Working Paper 20073, National Bureau of Economic Research.

Krusell, Per, and Anthony A Smith, Jr. 1998. "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy*, 106(5): 867–896.

LeGrand, Francois, Alais Martin-Baillon, and Xavier Ragot. 2020. "Should monetary policy care about redistribution? Optimal fiscal and monetary policy with heterogeneous agents." working paper.

LeGrand, Francois, and Xavier Ragot. 2017. "Optimal policy with heterogeneous agents and aggregate shocks: An application to optimal public debt dynamics." working paper.

Low, Hamish, Costas Meghir, and Luigi Pistaferri. 2010. "Wage risk and employment risk over the life cycle." *American Economic Review*, 100(4): 1432–1467.

Luenberger, David G. 1997. *Optimization by Vector Space Methods.* New York, NY, USA: John Wiley & Sons.

Marcet, Albert, and Ramon Marimon. 2019. "Recursive Contracts." *Econometrica*, 87(5): 1589–1631.

Nakamura, Emi, and Jon Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." *Annual Review of Economics*, 5(1): 133–163.

Nuno, Galo, and Carlos Thomas. 2016. "Optimal monetary policy with heterogeneous agents." Working Paper 1624.

Phillips, Kerk L. 2017. "Solving and simulating unbalanced growth models using linearization about the current state." *Economics Letters*, 151 35–38.

Reiter, Michael. 2009. "Solving heterogeneous-agent models by projection and perturbation." *Journal of Economic Dynamics and Control*, 33(3): 649–665.

Rotemberg, Julio J. 1982. "Monopolistic Price Adjustment and Aggregate Output." *Review of Economic Studies*, 49(4): 517–531.

Sbordone, Argia M. 2002. "Prices and unit labor costs: a new test of price stickiness." *Journal of Monetary Economics*, 49(2): 265–292.

Schmitt-Grohe, Stephanie, and Martin Uribe. 2004a. "Optimal fiscal and monetary policy under sticky prices." *Journal of Economic Theory*, 114(2): 198–230.

Schmitt-Grohe, Stephanie, and Martin Uribe. 2004b. "Solving dynamic general equilibrium models using a second-order approximation to the policy function." *Journal of Economic Dynamics and Control*, 28(4): 755–775.

Siu, Henry E. 2004. "Optimal Fiscal and Monetary Policy with Sticky Prices." *Journal of Monetary Economics*, 51(3): 575–607.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97(3): 586–606.

Werning, Ivan. 2007. "Optimal Fiscal Policy with Redistribution,." *Quarterly Journal of Economics*, 122(August): 925–967.

Woodford, Michael. 2003. *Interest and prices.*: Princeton University Press.

Online Appendix

A Additional details for section 3

We fill in the missing steps for section 3. First, in section A.1, we show how to formulate the Ramsey problem recursively, then in the context of the section 3.1 economy, how our method extends to higher-order approximations. Second, we show how to generalize the expansions so that we can deal with persistent aggregate and idiosyncratic shocks as well as additional state variables, as discussed in section 3.2.

A.1 Recursive Formulation of Ramsey problem

Here we show that the Lagrangian in equation (21) in section 3 of the main text admits a recursive solution from $t \geq 1$. We will also describe the F and R mappings that appear in equation (22) and (23) in this case. For completeness, we repeat the maximization problem here and list all implementability constraints. Given $\{b_{i,-1}\}_i$ and $\mu_{i,-1} = 0$, the planning problem is

$$\begin{aligned}\inf\sup\ \mathbb{E}_0 \textstyle\sum_{t=0}^{\infty} \beta^t \int \Big[u\left(c_{i,t}, n_{i,t}\right) + \left(u_{c,i,t}c_{i,t} + u_{n,i,t}n_{i,t} - u_{c,i,t}\left\{T_t + \left(1-\Upsilon_t^d\right)D_t\right\}\right)\mu_{i,t} \\ + \left(\tfrac{1-\Upsilon_t^b}{1+\Pi_t}\right) b_{i,t-1}u_{c,i,t}\left(\mu_{i,t-1} - \mu_{i,t}\right)\Big]\, di\end{aligned}$$

subject to

$$Q_{t-1}M_{t-1} = \beta m_{i,t-1}^{-1}\mathbb{E}_{t-1}\left[u_{c,i,t}\left(1-\Upsilon_t^b\right)(1+\Pi_t)^{-1}\right] \tag{41a}$$

$$u_{c,i,t}W_t(1-\Upsilon_t^n)\mathcal{E}_t\varepsilon_{,i,t} = -u_{n,i,t} \tag{41b}$$

$$M_t = m_{i,t}^{-1}u_{c,i,t} \tag{41c}$$

$$\int u_{c,i,t}di = M_t \tag{41d}$$

$$\int c_{i,t}di = C_t \tag{41e}$$

$$C_t + \bar{G} = \int \mathcal{E}_t\varepsilon_{,i,t}n_{i,t}di - \frac{\psi}{2}\Pi_t^2 \tag{41f}$$

$$D_t = (1-W_t)\int \mathcal{E}_t\varepsilon_{,i,t}n_{i,t}di - \frac{\psi}{2}\Pi_t^2 \tag{41g}$$

First-order conditions Let $\beta^{t-1}\rho_{i,t-1}, \beta^t\phi_{i,t}, \beta^t\varphi_{i,t}$ be Lagrange multipliers on household-level constraints (41a)–(41c); let $\beta^t\lambda_t, \beta^t\chi_t, \beta^t\Xi_t, \beta^t\zeta_t$ be Lagrange multipliers on aggregate constraints (41d)–(41g). First-order conditions with respect to household-level variables: $b_{i,t-1}, c_{i,t}, n_{i,t}, m_{i,t}^{-1}, \mu_{i,t}$ are

$$
\begin{aligned}
0 =& \mathbb{E}_{t-1}\left(\frac{1-\Upsilon_t^b}{1+\Pi_t}\right) u_{c,i,t}\left(\mu_{i,t-1}-\mu_{i,t}\right), && \text{(42a)}\\
0 =& \mu_{i,t}\left(u_{cc,i,t}\left[c_{i,t}-(T_t+\left(1-\Upsilon_t^d\right)D_t\right]+u_{c,i,t}\right)+\left(\frac{1-\Upsilon_t^b}{1+\Pi_t}\right) b_{i,t-1}u_{cc,i,t}\left(\mu_{i,t-1}-\mu_{i,t}\right)\\
&-\phi_{i,t}W_t(1-\Upsilon_t^n)\epsilon_{i,t}u_{cc,i,t}+\rho_{i,t-1}m_{i,t-1}^{-1}\left(1-\Upsilon_t^b\right)(1+\Pi_t)^{-1}u_{cc,i,t}\\
&+\varphi_{i,t}u_{cc,it}m_{i,t}^{-1}-\chi_t-\lambda_t+u_{c,i,t}, && \text{(42b)}\\
0 =& u_{n,i,t}+\mu_{i,t}\left(u_{nn,i,t}n_{i,t}+u_{n,i,t}\right)-\phi_{i,t}u_{nn,i,t}+\left[\Xi_t+\zeta_t(1-W_t)\right]\mathcal{E}_t\varepsilon_{,i,t}, && \text{(42c)}\\
0 =& \beta\mathbb{E}_t\left[\rho_{i,t}u_{c,i,t+1}\left(1-\Upsilon_{t+1}^b\right)(1+\Pi_{t+1})^{-1}\right]+\varphi_{i,t}u_{c,it}, && \text{(42d)}\\
0 =& \left(u_{c,i,t}c_{i,t}+u_{n,i,t}n_{i,t}-u_{c,i,t}\left\{T_t+\left(1-\Upsilon_t^d\right)D_t\right\}\right)\\
&-\left(\frac{1-\Upsilon_t^b}{1+\Pi_t}\right)b_{i,t-1}u_{c,i,t}+\beta\mathbb{E}_t\left(\frac{1-\Upsilon_{t+1}^b}{1+\Pi_{t+1}}\right)b_{i,t}u_{c,i,t+1}. && \text{(42e)}
\end{aligned}
$$

First-order conditions with respect to aggregate variables: $C_t, D_t, M_t, Q_t, W_t, (1+\Pi_t)^{-1}$,$T_t$, Υ_t^d, Υ_t^b, Υ_t^n are

$$
\begin{aligned}
0 =& \chi_t - \Xi_t, && \text{(43a)}\\
0 =& -\left(1-\Upsilon_t^d\right)\int u_{c,i,t}\mu_{i,t}di - \zeta_t, && \text{(43b)}\\
0 =& -\int \rho_{i,t}Q_t - \int \varphi_{i,t}di + \lambda_t, && \text{(43c)}\\
0 =& -\int \rho_{i,t-1}di, && \text{(43d)}\\
0 =& -\int \phi_{i,t}u_{c,i,t}(1-\Upsilon_t^n)\mathcal{E}_t\varepsilon_{,i,t}di - \zeta_t\int \mathcal{E}_t\varepsilon_{,i,t}n_{i,t}di, && \text{(43e)}\\
0 =& \left(1-\Upsilon_t^b\right)\int b_{i,t-1}u_{c,i,t}\left(\mu_{i,t-1}-\mu_{i,t}\right)di + \psi\Pi_t(1+\Pi_t)^2\left(\Xi_t+\zeta_t\right)\\
& + \beta\left(1-\Upsilon_t^b\right)\int \rho_{i,t-1}m_{i,t-1}^{-1}u_{c,i,t}di, && \text{(43f)}\\
0 =& \int u_{c,i,t}\mu_{i,t}di, && \text{(43g)}\\
0 =& \int u_{c,i,t}\mu_{i,t}di, && \text{(43h)}\\
0 =& -\int b_{i,t-1}u_{c,i,t}\left(\mu_{i,t-1}-\mu_{i,t}\right)di - \beta\int \rho_{i,t-1}m_{i,t-1}^{-1}u_{c,i,t}di, && \text{(43i)}\\
0 =& W_t\int \phi_{i,t}u_{c,i,t}\mathcal{E}_t\varepsilon_{,i,t}. && \text{(43j)}
\end{aligned}
$$

We can simplify some equations. We can set $\Pi_t = \zeta_t = 0$ and define $\hat{T}_t \equiv T_t + \left(1-\Upsilon_t^d\right)D_t$ and ignore (43f). Solving equations (42a)–(43j) is then the same as solving the following equations

$$
\begin{aligned}
0 =& \mathbb{E}_{t-1} \left(1 - \Upsilon_t^b\right) u_{c,i,t} \left(\mu_{i,t-1} - \mu_{i,t}\right), && \text{(44a)}\\
0 =& \mu_{i,t} \left(u_{cc,i,t} \left[c_{i,t} - \hat{T}_t\right] + u_{c,i,t}\right) + \left(1 - \Upsilon_t^b\right) b_{i,t-1} u_{cc,i,t} \left(\mu_{i,t-1} - \mu_{i,t}\right) \\
& - \phi_{i,t} W_t (1 - \Upsilon_t^n) \epsilon_{i,t} u_{cc,i,t} + \rho_{i,t-1} m_{i,t-1}^{-1} \left(1 - \Upsilon_t^b\right) u_{cc,i,t} \\
& + \varphi_{i,t} u_{cc,it} m_{i,t}^{-1} - \Xi_t - \lambda_t + u_{c,i,t}, && \text{(44b)}\\
0 =& u_{n,i,t} + \mu_{i,t} \left(u_{nn,i,t} n_{i,t} + u_{n,i,t}\right) - \phi_{i,t} u_{nn,i,t} + \Xi_t \epsilon_{i,t}, && \text{(44c)}\\
0 =& \beta \mathbb{E}_t \left[\rho_{i,t} u_{c,i,t+1} \left(1 - \Upsilon_{t+1}^b\right)\right] + \varphi_{i,t} u_{c,it}, && \text{(44d)}\\
0 =& u_{c,i,t} c_{i,t} + u_{n,i,t} n_{i,t} - u_{c,i,t} \hat{T}_t, \\
& - \left(1 - \Upsilon_t^b\right) b_{i,t-1} u_{c,i,t} + \beta \mathbb{E}_t \left(1 - \Upsilon_t^b\right) b_{i,t} u_{c,i,t+1}, && \text{(44e)}\\
0 =& - \int \varphi_{i,t} di + \lambda_t, && \text{(44f)}\\
0 =& \int \rho_{i,t-1} di, && \text{(44g)}\\
0 =& \int \phi_{i,t} u_{c,i,t} \epsilon_{i,t} di, && \text{(44h)}\\
0 =& \int b_{i,t-1} u_{c,i,t} \left(\mu_{i,t-1} - \mu_{i,t}\right) di, && \text{(44i)}\\
0 =& \int u_{c,i,t} \mu_{i,t} di. && \text{(44j)}
\end{aligned}
$$

Recursive Ramsey problems For $t \geq 1$, define individual-level states

$$
\boldsymbol{z}_{i,t-1} \equiv \left(m_{i,t-1}, \mu_{i,t-1}\right),
$$

and the aggregate state as a joint distribution over $\boldsymbol{z}_{i,t-1}$ to be denoted Ω_{t-1}; the individual-level choice variables as

$$
\tilde{\boldsymbol{x}}_{i,t} \equiv \left(c_{i,t}, n_{i,t}, b_{i,t-1}, \rho_{i,t-1}, \phi_{i,t}, \varphi_{i,t}, \mu_{i,t}, m_{i,t}\right),
$$

and the aggregate-level choice variables as

$$
\tilde{\boldsymbol{X}}_t \equiv \left(C_t, D_t, Q_t, W_t, M_t, \hat{T}_t, \Upsilon_t^b, \Upsilon_t^n, \lambda_t, \Xi_t\right).
$$

For $t \geq 1$, given Ω_{t-1} and shocks $\left(\mathcal{E}_t, \{\varepsilon_{i,t}\}_i\right)$, functions $\tilde{\boldsymbol{X}}\left(\Omega, \boldsymbol{\mathcal{E}}\right)$, $\tilde{\boldsymbol{x}}\left(\boldsymbol{z}, \Omega, \boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}\right)$, in the main text are defined as solutions to 17 equations (41a)–(41g) and (44a)–(44j) to be solved for 17 unknowns $\tilde{\boldsymbol{x}}_{i,t}$ and $\boldsymbol{X}_t$. The collection of equations (41a)–(41g) constitutes the F mapping in the text, and the collection of equations (44a)–(44j) constitutes the R mapping in the text.

For $t = 0$, define vectors $\tilde{\boldsymbol{x}}_{i,0}$ and $\tilde{\boldsymbol{X}}_0$ as

$$\tilde{\boldsymbol{x}}_{i,0} \equiv (c_{i,0}, n_{i,0}, \phi_{i,0}, \varphi_{i,0}, \mu_{i,0}, m_{i,0})$$

$$\tilde{\boldsymbol{X}}_0 \equiv \left(C_0, D_0, Q_0, W_0, M_0, \hat{T}_0, \Upsilon_0^b, \Upsilon_0^n, \lambda_0, \Xi_0 \right).$$

Given an initial condition $\Omega_{-1}^b \equiv \{b_{i,-1}\}_i$ and shocks $\left(\mathcal{E}_0, \{\varepsilon_{,i,0}\}_i\right)$ the time-0 policy functions $\tilde{\boldsymbol{X}}_0\left(\Omega_{-1}^b, \boldsymbol{\mathcal{E}}\right), \tilde{\boldsymbol{x}}_0\left(b, \Omega_{-1}^b, \boldsymbol{\varepsilon}, \mathcal{E}\right)$ solve 15 equations (41b)–(41g), and (44b)–(44j) for 15 unknowns $\tilde{\boldsymbol{x}}_{i,0}$ and $\tilde{\boldsymbol{X}}_0$ given $\tilde{\boldsymbol{x}}_{i,1}$ and $\tilde{\boldsymbol{X}}_1$.

A.2 Higher order approximations for section 3.1

We start with a second-order approximation to the model presented in section 3.1. These are given by

$$\begin{aligned}\tilde{\boldsymbol{X}}(\Omega, \sigma\boldsymbol{\mathcal{E}}; \sigma) = &\bar{\boldsymbol{X}} + \sigma\left(\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}\boldsymbol{\mathcal{E}} + \bar{\boldsymbol{X}}_\sigma\right) \\ &+ \frac{1}{2}\sigma^2\left(\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}}\boldsymbol{.}(\boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{E}}) + 2\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}\boldsymbol{\sigma}}\boldsymbol{\mathcal{E}} + \bar{\boldsymbol{X}}_{\sigma\sigma}\right) \\ &+ \mathcal{O}(\sigma^3),\end{aligned}$$

where the symbol $\mathsf{a} \cdot (\mathsf{b}, \mathsf{c})$ denotes a bilinear map.[32] A similar expansion can be written for $\tilde{\boldsymbol{x}}(\boldsymbol{z}, \Omega, \sigma\boldsymbol{\varepsilon}, \sigma\boldsymbol{\mathcal{E}}; \sigma)$.

To obtain the necessary terms, we proceed in two steps: section A.2.1 computes intermediate terms including higher-order Fréchet derivatives for individual and aggregate policy functions, and section A.2.2 uses these terms to compute the second-order expansion. Although the second-order expansion requires additional notation, the steps below highlight how the the same fundamental insights presented in section 3 maintain the tractability of the problem.

[32] Specifically, if a is a $n_1 \times n_2 \times n_3$ tensor, b is a $n_2 \times n_4$ matrix and c is a $n_3 \times n_5$ matrix then $\mathsf{d} = \mathsf{a} \cdot (\mathsf{b}, \mathsf{c})$ is $n_1 \times n_4 \times n_5$ tensor defined by

$$\mathsf{d}_{ilm} = \sum_{j,k} \mathsf{a}_{ijk}\mathsf{b}_{jl}\mathsf{c}_{km}.$$

This definition generalizes to when a, b, or c is infinite dimensional, such as with $\partial\bar{\boldsymbol{x}}_{\mathsf{z}}$.

A.2.1 Intermediate terms for second-order expansions

Differentiating equation (22) twice with respect to $\boldsymbol{z}$ we find

$$\begin{aligned}0 =&\bar{F}_{\boldsymbol{x}^-}\bar{\boldsymbol{x}}_{\boldsymbol{zz}} + \bar{F}_{\boldsymbol{x}}\bar{\boldsymbol{x}}_{\boldsymbol{zz}} + \bar{F}_{\boldsymbol{x}^+}\left(\bar{\boldsymbol{x}}_{\boldsymbol{zz}} + \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{zz}}\right)\\ &+ \bar{F}_{\boldsymbol{zz}} + \bar{F}_{\boldsymbol{zx}^-}\cdot(I,\bar{\boldsymbol{x}}_z) + \bar{F}_{\boldsymbol{zx}}\cdot(I,\bar{\boldsymbol{x}}_z) + \bar{F}_{\boldsymbol{zx}^+}\cdot(I,\boldsymbol{x}_{\boldsymbol{z}})\\ &+ \bar{F}_{\boldsymbol{x}^-\boldsymbol{z}}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},I) + \bar{F}_{\boldsymbol{x}^-\boldsymbol{x}^-}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\bar{\boldsymbol{x}}_{\boldsymbol{z}}) + \bar{F}_{\boldsymbol{x}^-\boldsymbol{x}}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\bar{\boldsymbol{x}}_{\boldsymbol{z}}) + \bar{F}_{\boldsymbol{x}^-\boldsymbol{x}^+}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\boldsymbol{x}_{\boldsymbol{z}})\\ &+ \bar{F}_{\boldsymbol{xz}}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},I) + \bar{F}_{\boldsymbol{xx}^-}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\bar{\boldsymbol{x}}_{\boldsymbol{z}}) + \bar{F}_{\boldsymbol{xx}}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\bar{\boldsymbol{x}}_{\boldsymbol{z}}) + \bar{F}_{\boldsymbol{xx}^+}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\boldsymbol{x}_{\boldsymbol{z}})\\ &+ \bar{F}_{\boldsymbol{x}^+\boldsymbol{z}}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},I) + \bar{F}_{\boldsymbol{x}^+\boldsymbol{x}^-}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\bar{\boldsymbol{x}}_{\boldsymbol{z}}) + \bar{F}_{\boldsymbol{x}^+\boldsymbol{x}}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\bar{\boldsymbol{x}}_{\boldsymbol{z}}) + \bar{F}_{\boldsymbol{x}^+\boldsymbol{x}^+}\cdot(\bar{\boldsymbol{x}}_{\boldsymbol{z}},\boldsymbol{x}_{\boldsymbol{z}}),\end{aligned}$$

where I represents the identity matrix and we use $\mathsf{a}\cdot(\mathsf{b},\mathsf{c})$ to denote a bilinear map. Lines 2-5 appear complicated but are actually simply combining all of the already known derivatives of $\boldsymbol{x}$ with cross derivatives of F. It will prove convenient to combine all of these terms into a single term: $\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\boldsymbol{z},\boldsymbol{x}^-,\boldsymbol{x},\boldsymbol{x}^+\}}\bar{F}_{\boldsymbol{\alpha\beta}}\cdot(\bar{\boldsymbol{\alpha}}_{\boldsymbol{z}},\bar{\boldsymbol{\beta}}_{\boldsymbol{z}})$ with the knowledge that $\bar{\boldsymbol{z}}_{\boldsymbol{z}}\equiv I, \bar{\boldsymbol{x}}^-_{\boldsymbol{z}}\equiv\bar{\boldsymbol{x}}_{\boldsymbol{z}}$, and $\bar{\boldsymbol{x}}^+_{\boldsymbol{z}}\equiv\bar{\boldsymbol{x}}_{\boldsymbol{z}}$. In doing this $\bar{\boldsymbol{x}}_{\boldsymbol{zz}}$ can be represented by a simple linear equation

$$\bar{\boldsymbol{x}}_{\boldsymbol{zz}} = -\left[\bar{F}_{\boldsymbol{x}^-} + \bar{F}_{\boldsymbol{x}} + \bar{F}_{\boldsymbol{x}^+}\left(I + \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\right)\right]^{-1}\left(\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\boldsymbol{z},\boldsymbol{x}^-,\boldsymbol{x},\boldsymbol{x}^+\}}\bar{F}_{\boldsymbol{\alpha\beta}}\cdot(\bar{\boldsymbol{\alpha}}_{\boldsymbol{z}},\bar{\boldsymbol{\beta}}_{\boldsymbol{z}})\right).$$

In a similar manner one can show that

$$\partial\boldsymbol{x}_{\boldsymbol{z}}\cdot\Delta = -\left[\bar{F}_{\boldsymbol{x}^-} + \bar{F}_{\boldsymbol{x}} + \bar{F}_{\boldsymbol{x}^+}\left(I + \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\right)\right]^{-1}\left(\sum_{\substack{\boldsymbol{\alpha}\in\{\boldsymbol{z},\boldsymbol{x}^-,\boldsymbol{x},\boldsymbol{x}^+\}\\ \boldsymbol{\beta}\in\{\boldsymbol{x}^-,\boldsymbol{x},\boldsymbol{x}^+,\boldsymbol{X}\}}}\bar{F}_{\boldsymbol{\alpha\beta}}\cdot\left(\bar{\boldsymbol{\alpha}}_{\boldsymbol{z}},\partial\bar{\boldsymbol{\beta}}\cdot\Delta\right)\right),$$

where we use $\partial\bar{\boldsymbol{x}}^-\cdot\Delta\equiv\partial\bar{\boldsymbol{x}}^+\cdot\Delta\equiv\partial\bar{\boldsymbol{x}}\cdot\Delta$.

The last of the derivatives with respect to the state variables that are required for the second order expansion is $\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2)$. We will use the pre computed expressions for $\partial\bar{\boldsymbol{x}}$ and $\partial\bar{\boldsymbol{X}}$ evaluating them in the directions Δ_1 and Δ_2. Differentiating (22) we find

$$\begin{aligned}0 =&\bar{F}_{\boldsymbol{x}^-}\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2) + \bar{F}_{\boldsymbol{x}}\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2) + \bar{F}_{\boldsymbol{x}^+}\left(\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2) + \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2)\right) + \bar{F}_{\boldsymbol{X}}\partial^2\bar{\boldsymbol{X}}\cdot(\Delta_1,\Delta_2)\\ &+ \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\boldsymbol{x}^-,\boldsymbol{x},\boldsymbol{x}^+,\boldsymbol{X}\}}\bar{F}_{\boldsymbol{\alpha\beta}}\cdot\left(\partial\bar{\boldsymbol{\alpha}}\cdot\Delta_1,\partial\bar{\boldsymbol{\beta}}\cdot\Delta_2\right).\end{aligned}$$

In solving this equation for $\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2)$ we find

$$\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2) = \mathsf{A}(\boldsymbol{z},\Delta_1,\Delta_2) + \mathsf{C}(\boldsymbol{z})\partial^2\bar{\boldsymbol{X}}\cdot(\Delta_1,\Delta_2)$$

where

$$\mathsf{A}(\boldsymbol{z},\Delta_1,\Delta_2) = -\left[\bar{F}_{\boldsymbol{x}^-} + \bar{F}_{\boldsymbol{x}} + \bar{F}_{\boldsymbol{x}^+}\left(I + \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\right)\right]^{-1}\left(\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\boldsymbol{x}^-,\boldsymbol{x},\boldsymbol{x}^+,\boldsymbol{X}\}} \bar{F}_{\boldsymbol{\alpha}\boldsymbol{\beta}} \cdot \left(\partial\bar{\boldsymbol{\alpha}}\cdot\Delta_1, \partial\bar{\boldsymbol{\beta}}\cdot\Delta_2\right)\right)$$

from terms already known and $\mathsf{C}(\boldsymbol{z})$ is the same term computed in section 3.1. To find $\partial^2\bar{\boldsymbol{X}}\cdot(\Delta_1,\Delta_2)$ we differentiate (23) to find

$$\begin{aligned}0 =&\bar{R}_{\boldsymbol{x}}\int \partial^2\bar{\boldsymbol{x}}(\boldsymbol{y})\cdot(\Delta_1,\Delta_2)d\Omega(\boldsymbol{y}) + \bar{R}_{\boldsymbol{X}}\partial^2\bar{\boldsymbol{X}}\cdot(\Delta_1,\Delta_2) + \int\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\boldsymbol{x}(\boldsymbol{y}),\boldsymbol{X}\}}\bar{R}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\cdot\left(\partial\boldsymbol{\alpha}\cdot\Delta_1,\partial\boldsymbol{\beta}\cdot\Delta_2\right)d\Omega(\boldsymbol{y})\\ &+\bar{R}_{\boldsymbol{x}}\int\partial\bar{\boldsymbol{x}}(\boldsymbol{y})\cdot\Delta_1 d\Delta_2(\boldsymbol{y}) + \bar{R}_{\boldsymbol{x}}\int\partial\bar{\boldsymbol{x}}(\boldsymbol{y})\cdot\Delta_2 d\Delta_1(\boldsymbol{y}).\end{aligned}$$

Plugging in for $\partial^2\bar{\boldsymbol{x}}\cdot(\Delta_1,\Delta_2)$ yields a linear equation which can be easily solved for $\partial^2\bar{\boldsymbol{X}}\cdot(\Delta_1,\Delta_2)$.

A.2.2 Second-order expansions

We can use these derivatives to compute the second-order terms. To find $\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}$, differentiate F twice with respect to $\boldsymbol{\varepsilon}$ to get the linear equation[33]

$$0 = \bar{F}_{\boldsymbol{x}}\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} + \bar{F}_{\boldsymbol{x}^+}\bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} + \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\boldsymbol{x},\boldsymbol{x}^+,\boldsymbol{\varepsilon}\}}\bar{F}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\cdot\left(\bar{\boldsymbol{\alpha}}_{\boldsymbol{\varepsilon}},\bar{\boldsymbol{\beta}}_{\boldsymbol{\varepsilon}}\right),$$

where $\bar{\boldsymbol{x}}^+_{\boldsymbol{\varepsilon}} \equiv \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}_{\boldsymbol{\varepsilon}} \equiv I$. Similarly, $\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\mathcal{E}}}$ solves the following linear equation

$$0 = \bar{F}_{\boldsymbol{x}}\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\mathcal{E}}} + \bar{F}_{\boldsymbol{x}^+}\bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\mathcal{E}}} + \sum_{\substack{\boldsymbol{\alpha}\in\{\boldsymbol{x},\boldsymbol{x}^+,\boldsymbol{\varepsilon}\}\\ \boldsymbol{\beta}\in\{\boldsymbol{x},\boldsymbol{x}^+,\boldsymbol{X},\boldsymbol{\mathcal{E}}\}}}\bar{F}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\cdot\left(\bar{\boldsymbol{\alpha}}_{\boldsymbol{\varepsilon}},\bar{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{E}}}\right),$$

with the understanding that $\bar{\boldsymbol{\mathcal{E}}}_{\boldsymbol{\mathcal{E}}} \equiv I$ and $\bar{\boldsymbol{x}}^+_{\boldsymbol{\mathcal{E}}} \equiv \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}} + \partial\bar{\boldsymbol{x}}\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}}$.

Differentiating twice with respect to $\boldsymbol{\mathcal{E}}$ yields

$$\begin{aligned}0 =&\bar{F}_{\boldsymbol{x}}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}} + \bar{F}_{\boldsymbol{x}^+}\left(\bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}} + \partial\bar{\boldsymbol{x}}\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}}\right) + \bar{F}_{\boldsymbol{X}}\bar{\boldsymbol{X}}_{\mathcal{E}\mathcal{E}}\\ &+\bar{F}_{\boldsymbol{x}^+}\left(\bar{\boldsymbol{x}}_{\boldsymbol{z}\boldsymbol{z}}\cdot(\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}},\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}) + \partial\bar{\boldsymbol{x}}_{\boldsymbol{z}}\cdot(\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}},\bar{\Omega}_{\boldsymbol{\mathcal{E}}}) + \partial\bar{\boldsymbol{x}}_{\boldsymbol{z}}\cdot(\bar{\Omega}_{\boldsymbol{\mathcal{E}}},\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}}) + \partial^2\bar{\boldsymbol{x}}\cdot(\bar{\Omega}_{\boldsymbol{\mathcal{E}}},\bar{\Omega}_{\boldsymbol{\mathcal{E}}})\right)\\ &+\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\boldsymbol{x},\boldsymbol{x}^+,\boldsymbol{X},\boldsymbol{\mathcal{E}}\}}\bar{F}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\cdot(\bar{\boldsymbol{\alpha}}_{\boldsymbol{\mathcal{E}}},\bar{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{E}}})\end{aligned} \tag{45}$$

[33] For parsimony we have dropped the dependence on $\boldsymbol{z}$ when not necessary.

and

$$\int \bar{R}_{\boldsymbol{x}}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{y}) + \bar{R}_{\boldsymbol{X}}\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}\mathcal{E}}} + \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\{\bar{\boldsymbol{x}}(\boldsymbol{y}),\bar{\boldsymbol{X}}\}} \bar{R}_{\boldsymbol{\alpha\beta}}\cdot(\bar{\boldsymbol{\alpha}}_{\boldsymbol{\mathcal{E}}},\bar{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{E}}})d\Omega(\boldsymbol{y}). \tag{46}$$

All the terms in the second line can be computed from our analysis of the previous section and all the terms in the third line are known. What remains is to find $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}$ and $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}$. This requires us extend the steps we used in the proof of theorem 1.

Differentiating (24) twice with respect to $\boldsymbol{\mathcal{E}}$, evaluated at $\sigma = 0$, yields

$$\begin{aligned}\bar{\Omega}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{y}) = &-\int\sum_i \delta(\boldsymbol{z}^i-\boldsymbol{y}^i)\prod_{j\neq i}\iota(\boldsymbol{z}^j-\boldsymbol{y}^j)\bar{\boldsymbol{z}}^i_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{z})d\Omega(\boldsymbol{z})\\ &-\int\sum_i \delta'(\boldsymbol{z}^i-\boldsymbol{y}^i)\prod_{j\neq i}\iota(\boldsymbol{z}^j-\boldsymbol{y}^j)\left[\bar{\boldsymbol{z}}^i_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})\right]^2 d\Omega(\boldsymbol{z})\\ &+\int\sum_i \delta(\boldsymbol{z}^i-\boldsymbol{y}^i)\sum_{j\neq i}\delta(\boldsymbol{z}^j-\boldsymbol{y}^j)\prod_{k\neq i,j}\iota(\boldsymbol{z}^k-\boldsymbol{y}^k)\bar{\boldsymbol{z}}^j_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})\bar{\boldsymbol{z}}^i_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})d\Omega(\boldsymbol{z}).\end{aligned}$$

The density is then

$$\bar{\omega}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{y}) = \frac{\partial^{n_{\boldsymbol{z}}}}{\partial\boldsymbol{y}^1\partial\boldsymbol{y}^2\cdots\partial\boldsymbol{y}^{n_{\boldsymbol{z}}}}\bar{\Omega}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{y}) = -\sum_i\frac{\partial}{\partial\boldsymbol{y}^i}\left(\bar{\boldsymbol{z}}^i_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{y})\omega_{\mathcal{E}}(\boldsymbol{y})\right) + \sum_i\sum_j\frac{\partial^2}{\partial\boldsymbol{y}^i\partial\boldsymbol{y}^j}\Big(\bar{\boldsymbol{z}}^i_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y})\bar{\boldsymbol{z}}^j_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y})\omega_{\mathcal{E}}(\boldsymbol{y})\Big).$$

The identical steps to (1) then show that

$$\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}\mathcal{E}}} = \mathsf{C}(\boldsymbol{z})\partial\bar{\boldsymbol{X}}\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}\mathcal{E}}} \equiv \mathsf{C}(\boldsymbol{z})\bar{\boldsymbol{X}}'_{\boldsymbol{\mathcal{E}\mathcal{E}}}$$

with

$$\bar{\boldsymbol{X}}'_{\boldsymbol{\mathcal{E}\mathcal{E}}} = -\left(\bar{R}_{\boldsymbol{x}}\int\mathsf{C}(\boldsymbol{y})d\Omega(\boldsymbol{y})+\bar{R}_{\boldsymbol{X}}\right)^{-1}\bar{R}_{\boldsymbol{x}}\left(\int\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{y})\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{y})+\bar{\boldsymbol{x}}_{\boldsymbol{zz}}(\boldsymbol{y})\cdot(\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}},\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}})\,d\Omega(\boldsymbol{y})\right). \tag{47}$$

As with $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}}}$, rather than solving for $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{z})$ and $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}$ jointly, we substitute for $\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\bar{\Omega}_{\boldsymbol{\mathcal{E}\mathcal{E}}}$ in (45) and solve for $\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{z})$ yielding the linear relationship

$$\bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}(\boldsymbol{z}) = \mathsf{D}_1(\boldsymbol{z})\cdot\begin{bmatrix}\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}\mathcal{E}}} & \bar{\boldsymbol{X}}'_{\boldsymbol{\mathcal{E}\mathcal{E}}}\end{bmatrix}^{\mathsf{T}} + \mathsf{D}_2(\boldsymbol{z})$$

where $\mathsf{D}_1(\boldsymbol{z})$ is identical to the D_1 in section 3.1. We then use this relationship to substitute into equations (46) and (47) to find $\bar{\boldsymbol{X}}_{\boldsymbol{\mathcal{E}\mathcal{E}}}$ and $\bar{\boldsymbol{X}}'_{\boldsymbol{\mathcal{E}\mathcal{E}}}$.

A key part of the second-order approximations is capturing the effect of risk via the terms $\bar{\boldsymbol{x}}_{\sigma\sigma}(\boldsymbol{z})$ and $\bar{\boldsymbol{X}}_{\sigma\sigma}(\boldsymbol{z})$.[34] Let $\mathbb{C}_{\boldsymbol{\varepsilon}} \equiv \mathbb{E}\boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon}$ and $\mathbb{C}_{\mathcal{E}} \equiv \mathbb{E}\boldsymbol{\mathcal{E}}^{\mathsf{T}}\boldsymbol{\mathcal{E}}$ be the variance-covariance matrix of the idiosyncratic and aggregate shocks respectively. Differentiating (22) and (23)

[34] It is easy to verify that the cross derivatives with shocks and σ are zero.

yields[35]

$$0 = \bar{F}_{\boldsymbol{x}-}\left(\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} \cdot \mathbb{C}_{\varepsilon} + \bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}} \cdot \mathbb{C}_{\mathcal{E}}\right) + \bar{F}_{\boldsymbol{x}}\bar{\boldsymbol{x}}_{\sigma\sigma} + \bar{F}_{\boldsymbol{X}}\bar{\boldsymbol{X}}_{\sigma\sigma} + \mathsf{F}_{\boldsymbol{x}+}\left(\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} \cdot \mathbb{C}_{\varepsilon} + \bar{\boldsymbol{x}}_{\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}} \cdot \mathbb{C}_{\mathcal{E}} + \bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p}\bar{\boldsymbol{x}}_{\sigma\sigma} + \partial\bar{\boldsymbol{x}} \cdot \bar{\Omega}_{\sigma\sigma}\right) \tag{48}$$

and

$$0 = \bar{R}_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}_{\sigma\sigma}(\boldsymbol{y}) + \bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}(\boldsymbol{y}) \cdot \mathbb{C}_{\boldsymbol{\varepsilon}} d\Omega(\boldsymbol{y}) + \bar{R}_{\boldsymbol{X}}\bar{\boldsymbol{X}}_{\sigma\sigma}. \tag{49}$$

Before this set of equations can be solved for $\bar{\boldsymbol{x}}_{\sigma\sigma}$, we must evaluate $\bar{\Omega}_{\sigma\sigma}$. Differentiating (24) and evaluating at $\sigma = 0$ yields

$$\begin{aligned}
\bar{\Omega}_{\sigma\sigma}(\boldsymbol{y}) = &- \int \sum_i \delta(\boldsymbol{z}^i - \boldsymbol{y}^i) \prod_{j \neq i} \iota(\boldsymbol{z}^j - \boldsymbol{y}^j) \left(\bar{\boldsymbol{z}}^i_{\sigma\sigma}(\boldsymbol{z}) + \bar{\boldsymbol{z}}^i_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} \cdot \mathbb{C}_{\boldsymbol{\varepsilon}}\right) d\Omega(\boldsymbol{z}) \\
&- \int \sum_i \delta'(\boldsymbol{z}^i - \boldsymbol{y}^i) \prod_{j \neq i} \iota(\boldsymbol{z}^j - \boldsymbol{y}^j) \left[\bar{\boldsymbol{z}}^i_{\boldsymbol{\varepsilon}}(\boldsymbol{z})\right]^2 \cdot \mathbb{C}_{\boldsymbol{\varepsilon}} d\Omega(\boldsymbol{z}) \\
&+ \int \sum_i \delta(\boldsymbol{z}^i - \boldsymbol{y}^i) \sum_{j \neq i} \delta(\boldsymbol{z}^j - \boldsymbol{y}^j) \prod_{k \neq i,j} \iota(\boldsymbol{z}^k - \boldsymbol{y}^k) \left(\bar{\boldsymbol{z}}^j_{\boldsymbol{\varepsilon}}(\boldsymbol{z})\bar{\boldsymbol{z}}^i_{\boldsymbol{\varepsilon}}(\boldsymbol{z})\right) \cdot \mathbb{C}_{\boldsymbol{\varepsilon}} d\Omega(\boldsymbol{z})
\end{aligned}$$

which gives

$$\begin{aligned}
\bar{\omega}_{\sigma\sigma}(\boldsymbol{y}) = &- \sum_i \frac{\partial}{\partial \boldsymbol{y}^i} \left(\left(\bar{\boldsymbol{z}}^i_{\sigma\sigma}(\boldsymbol{y}) + \bar{\boldsymbol{z}}^i_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}(\boldsymbol{y}) \cdot \mathbb{C}_{\boldsymbol{\varepsilon}}\right) \omega(\boldsymbol{y})\right) \\
&+ \sum_i \sum_j \frac{\partial^2}{\partial \boldsymbol{y}^i \partial \boldsymbol{y}^j} \left(\left(\bar{\boldsymbol{z}}^i_{\boldsymbol{\varepsilon}}(\boldsymbol{y})\bar{\boldsymbol{z}}^j_{\boldsymbol{\varepsilon}}(\boldsymbol{y})\right) \cdot \mathbb{C}_{\boldsymbol{\varepsilon}} \omega(\boldsymbol{y})\right).
\end{aligned}$$

Following the identical steps as theorem (1) to show that show that

$$\partial\bar{\boldsymbol{x}}(\boldsymbol{z}) \cdot \bar{\Omega}_{\sigma\sigma} = \mathsf{C}(\boldsymbol{z})\partial\bar{\boldsymbol{X}} \cdot \bar{\Omega}_{\sigma\sigma} \equiv \mathsf{C}(\boldsymbol{z})\bar{\boldsymbol{X}}'_{\sigma\sigma}$$

with

$$\begin{aligned}
\bar{\boldsymbol{X}}'_{\sigma\sigma} = - \left(\bar{R}_{\boldsymbol{x}} \int \mathsf{C}(\boldsymbol{y}) d\Omega(\boldsymbol{y}) + \bar{R}_{\boldsymbol{X}}\right)^{-1} \bar{R}_{\boldsymbol{x}} \int \Bigg(&\bar{\boldsymbol{x}}_{\boldsymbol{z}}(\boldsymbol{y})\mathsf{p}\left(\bar{\boldsymbol{x}}_{\sigma\sigma}(\boldsymbol{y}) + \bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}(\boldsymbol{y}) \cdot \mathbb{C}_{\boldsymbol{\varepsilon}}\right) \\
&+ \bar{\boldsymbol{x}}_{\boldsymbol{z}\boldsymbol{z}}(\boldsymbol{y}) \cdot \left(\mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}}, \mathsf{p}\bar{\boldsymbol{x}}_{\boldsymbol{\varepsilon}}\right) \cdot \mathbb{C}_{\boldsymbol{\varepsilon}} \Bigg) d\Omega(\boldsymbol{y}).
\end{aligned} \tag{50}$$

We then substitute for $\partial\bar{\boldsymbol{x}} \cdot \bar{\Omega}_{\sigma\sigma} = \mathsf{C}(\boldsymbol{z})\bar{\boldsymbol{X}}'_{\sigma\sigma}$ in (48) to and solve for $\bar{\boldsymbol{x}}_{\sigma\sigma}(\boldsymbol{z})$ to find the

[35]If a is a $n_1 \times n_2 \times n_2$ tensor and $\mathbb{C}$ is a $n_2 \times n_2$ matrix then $\mathsf{d} = \mathsf{a} \cdot \mathbb{C}$ is length n_1 vector defined by

$$\mathsf{d}_i = \sum_{j,k} \mathsf{a}_{ijk}\mathbb{C}_{jk}.$$

linear relationship

$$\bar{\boldsymbol{x}}_{\sigma\sigma}(\boldsymbol{z}) = \mathsf{E}_0(\boldsymbol{z}) + \mathsf{E}_1(\boldsymbol{z}) \begin{bmatrix} \bar{\boldsymbol{X}}_{\sigma\sigma} & \bar{\boldsymbol{X}}'_{\sigma\sigma} \end{bmatrix}^{\mathsf{T}}.$$

This relationship can then be plugged into (49) and (50) to yield a linear equation for $\bar{\boldsymbol{X}}_{\sigma\sigma}$ and $\bar{\boldsymbol{X}'}_{\sigma\sigma}$.

A.3 Expansions in the general case of section 3.2

We extend our method to handle persistent shocks and other endogenous persistent state variables besides the distributional state Ω. To do so, we extend the equilibrium conditions in the following manner

$$F\left(\mathbb{E}_{-}\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}, \mathbb{E}_{+}\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{X}}, \Lambda, \boldsymbol{\Theta}, \varepsilon, \boldsymbol{\mathcal{E}}, \boldsymbol{z}\right) = \boldsymbol{0}, \tag{51}$$

which must hold for all $\boldsymbol{z}$ in the support of Ω,

$$R\left(\int \tilde{\boldsymbol{x}} d\Omega d\Pr(\boldsymbol{\varepsilon}), \tilde{\boldsymbol{X}}, \mathbb{E}_{+}\tilde{\boldsymbol{X}}, \Lambda, \boldsymbol{\Theta}, \boldsymbol{\mathcal{E}}\right) = 0, \tag{52}$$

and a first-order vector autoregression model $\boldsymbol{\Theta}' = \rho_\Theta \boldsymbol{\Theta} + (1-\rho_\Theta)\bar{\boldsymbol{\Theta}} + \boldsymbol{\mathcal{E}}$ for the exogenous shocks. The law of motion of the distribution is given by

$$\widetilde{\Omega}\left(\Omega, \Lambda, \boldsymbol{\Theta}, \boldsymbol{\mathcal{E}}\right)(\boldsymbol{z}) = \int \iota\left(\tilde{\boldsymbol{z}}\left(\boldsymbol{y}, \Omega, \Lambda, \boldsymbol{\Theta}, \varepsilon, \boldsymbol{\mathcal{E}}\right) \le \boldsymbol{z}\right) d\Pr\left(\varepsilon\right) d\Omega\left(\boldsymbol{y}\right) \quad \forall \boldsymbol{z}. \tag{53}$$

We consider a family of perturbations indexed by a positive scalar σ that scales all shocks $\boldsymbol{\varepsilon}, \boldsymbol{\mathcal{E}}$ so that the policy functions are $\tilde{\boldsymbol{X}}(\Omega, \Lambda, \boldsymbol{\Theta}, \sigma\boldsymbol{\mathcal{E}}; \sigma)$ and $\tilde{\boldsymbol{x}}(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta}, \sigma\boldsymbol{\varepsilon}, \sigma\boldsymbol{\mathcal{E}}; \sigma)$. We will use $\bar{\cdot}$ to denote these functions evaluated at $\sigma = 0$.

Unlike section 3.1, we cannot assume that $\bar{\Omega}(\Omega, \Lambda, \boldsymbol{\Theta})$ is stationary but we recover the independence property

Lemma 2. *For any $\Omega, \Lambda, \boldsymbol{\Theta}$, the policy functions $\bar{z}(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta})$ satisfy $\partial \bar{z}(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta}) = \boldsymbol{0}$ for all $\boldsymbol{z}$ and $\bar{z}_{\boldsymbol{z}}(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta})$ independent of $\boldsymbol{z}$.*

Proof. We proceed similar to the proof of Lemma 1 in the main text. The first order condition with respect to $b_{i,t-1}$ yields

$$\mathbb{E}\left[\frac{\left[\tilde{c}\left(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta}, \cdot, \cdot\right)\right]^{-\nu}}{1 + \tilde{\Pi}\left(\Omega, \Lambda, \boldsymbol{\Theta}, \cdot, \right)}\left(\mu - \tilde{\mu}\left(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta}, \cdot, \cdot\right)\right)\right] \quad = \quad 0.$$

When $\sigma = 0$, this yields $\bar{\mu}(\boldsymbol{z}, \Omega, \Lambda, \boldsymbol{\Theta}) = \mu$ for all $\boldsymbol{z}$. While equation (20) to the zeroth order is

$$\bar{Q}(\Omega,\Lambda,\mathbf{\Theta})\bar{M}(\Omega,\Lambda,\mathbf{\Theta})m=\bar{m}\left(\boldsymbol{z},\Omega,\Lambda,\mathbf{\Theta}\right)\bar{M}\left(\bar{\Omega}(\Omega,\Lambda,\mathbf{\Theta})\right)\left(1+\bar{\Pi}\left(\bar{\Omega}(\Omega,\Lambda,\mathbf{\Theta})\right)\right)^{-1}.$$

By construction, the Pareto weights integrate to one which implies $\bar{m}\left(\boldsymbol{z},\Omega,\Lambda,\mathbf{\Theta}\right)=m$ for all $\boldsymbol{z}$. Finally, the law of motion for θ implies

$$\bar{\theta}\left(\boldsymbol{z},\Omega,\Lambda,\Theta\right)=\rho_\theta\theta.$$

Together they imply $\partial\bar{\boldsymbol{z}}(\boldsymbol{z},\Omega,\Lambda,\mathbf{\Theta})=\mathbf{0}$ for all $\boldsymbol{z}$ and $\bar{\boldsymbol{z}}_{\boldsymbol{z}}(\boldsymbol{z},\Omega,\Lambda,\mathbf{\Theta})$ independent of $\boldsymbol{z}$. $\square$

A by product of 2 is that $\bar{\boldsymbol{z}}_{\boldsymbol{z}}$ is diagonal. Although we exploit this property in the next section it is not essential.

We start by showing how our expansion extends to the transition path. We assume for a given $\Omega,\Lambda,\mathbf{\Theta}$ we have solved for the $\sigma=0$ transition dynamics $\{\bar{\Omega}^n,\bar{\Lambda}^n,\bar{\mathbf{\Theta}}^n\}_{n=0}^N$ with $(\bar{\Omega}^0,\bar{\Lambda}^0,\bar{\mathbf{\Theta}}^0)=\Omega,\Lambda,\Theta$ and $(\bar{\Omega}^N,\bar{\Lambda}^N,\bar{\mathbf{\Theta}}^N)=(\bar{\mathbf{\Omega}},\bar{\Lambda},\bar{\mathbf{\Theta}})$ at a non-stochastic steady state. Solving the the transition dynamics is eased by the fact that we know, a priori, the transition dynamics of Ω. For the remainder of this appendix we use $\bar{\cdot}^n$ to denote derivatives evaluated at $(\bar{\Omega}^n,\bar{\Lambda}^n,\bar{\mathbf{\Theta}}^n)$ and, to save on notation, and use $\bar{\cdot}$ to denote derivatives evaluated at the steady state $(\bar{\Omega}^N,\bar{\Lambda}^N,\bar{\mathbf{\Theta}}^N)$. We'll start by showing how to compute derivatives at the steady state and then show how to evaluate derivatives along the path.

The policy rules for $\boldsymbol{X}$ and $\boldsymbol{x}$ can then be approximated via Taylor expansion. The first order expansions for these variables are given by

$$\tilde{\boldsymbol{X}}(\Omega,\Lambda,\mathbf{\Theta},\sigma\boldsymbol{\mathcal{E}};\sigma)=\bar{\boldsymbol{X}}^0+\sigma(\bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}}\boldsymbol{\mathcal{E}}+\bar{\boldsymbol{X}}^0_\sigma)+\mathcal{O}(\sigma^2)$$

and

$$\tilde{\boldsymbol{x}}(\boldsymbol{z},\Omega,\Lambda,\mathbf{\Theta},\sigma\boldsymbol{\varepsilon},\sigma\boldsymbol{\mathcal{E}};\sigma)=\bar{\boldsymbol{x}}^0(\boldsymbol{z})+\sigma\left(\bar{\boldsymbol{x}}^0_{\boldsymbol{\varepsilon}}(\boldsymbol{z})\boldsymbol{\varepsilon}+\bar{\boldsymbol{x}}^0_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})\boldsymbol{\mathcal{E}}+\bar{\boldsymbol{x}}^0_\sigma(\boldsymbol{z})\right)+\mathcal{O}(\sigma^2).$$

For brevity, we present the necessary derivatives for the first order expansions. Higher order terms extend analogously to section A.2 .

A.3.1 Derivatives at the steady state

The derivatives of the policy functions with respect to Λ and $\mathbf{\Theta}$ as well as the Fréchet derivative with respect to the distribution Ω are used repeatedly in what follows.

Differentiating (51) with respect to Λ yields (lemma 2 implies that $\bar{\Omega}_\Lambda=0$).

$$\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z})\bar{\boldsymbol{x}}_\Lambda(\boldsymbol{z})+\bar{F}_{\boldsymbol{x}}(\boldsymbol{z})\bar{\boldsymbol{x}}_\Lambda(\boldsymbol{z})+\bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\left(\bar{\boldsymbol{x}}_\Lambda(\boldsymbol{z})\bar{\Lambda}_\Lambda\right)+\bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\bar{\boldsymbol{X}}_\Lambda=0$$

and

$$\bar{R}_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}_{\Lambda}(\boldsymbol{z}) d\Omega(\boldsymbol{z}) + \bar{R}_{\boldsymbol{X}} \bar{\boldsymbol{X}}_{\Lambda} + \bar{R}_{\boldsymbol{X}+} \bar{\boldsymbol{X}}_{\Lambda} \bar{\Lambda}_{\Lambda} + \bar{R}_{\Lambda} = 0.$$

The object $\bar{\Lambda}_{\Lambda}$ is unknown. It requires solving a nonlinear equation which we show below can be expressed using operations that involve matrices of small dimension. First note that

$$\bar{\boldsymbol{x}}_{\Lambda}(\boldsymbol{z}) = -(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{\Lambda}_{\Lambda} \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}))^{-1} \bar{F}_{\boldsymbol{X}}(\boldsymbol{z}) \bar{\boldsymbol{X}}_{\Lambda}$$

Let $\mathsf{A}(\boldsymbol{z}) = -(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{\Lambda}_{\Lambda} \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}))^{-1} \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})$, then

$$\bar{\boldsymbol{X}}_{\Lambda} = -\left(\bar{R}_{\boldsymbol{x}} \int \mathsf{A}(\boldsymbol{z}) d\Omega(\boldsymbol{z}) + \bar{R}_{\boldsymbol{X}} + \bar{\Lambda}_{\Lambda} \bar{R}_{\boldsymbol{X}+} \right)^{-1} \bar{R}_{\Lambda}.$$

Let P be such that $\Lambda = \mathsf{P}\boldsymbol{X}$. Therefore, $\bar{\Lambda}_{\Lambda}$ must solve

$$\bar{\Lambda}_{\Lambda} = -\mathsf{P} \left(\bar{R}_{\boldsymbol{x}} \int \mathsf{A}(\boldsymbol{z}) d\Omega(\boldsymbol{z}) + \bar{R}_{\boldsymbol{X}} + \bar{\Lambda}_{\Lambda} \bar{R}_{\boldsymbol{X}+} \right)^{-1} \bar{R}_{\Lambda}.$$

This can be found easily with a 1-dimensional root solver as all the matrices that need to be inverted are small dimensional.

Next differentiating (51) with respect to $\boldsymbol{\Theta}$ yields (lemma 2 implies that $\bar{\Omega}_{\boldsymbol{\Theta}} = 0$).

$$\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) \bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) \bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}) \left(\bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}(\boldsymbol{z}) \rho_{\boldsymbol{\Theta}} + \bar{\boldsymbol{x}}_{\Lambda}(\boldsymbol{z}) \mathsf{P} \bar{\boldsymbol{X}}_{\boldsymbol{\Theta}} \right) + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z}) \bar{\boldsymbol{X}}_{\boldsymbol{\Theta}} + \bar{F}_{\boldsymbol{\Theta}}(\boldsymbol{z}) = 0.$$

This yields a linear equation in $\bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}$ and $\bar{\boldsymbol{X}}_{\boldsymbol{\Theta}}$ which we can solve for $\bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}$.[36] Plugging in for the linear relationship between $\bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}$ and $\bar{\boldsymbol{X}}_{\boldsymbol{\Theta}}$ in

$$\bar{R}_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}(\boldsymbol{z}) d\Omega(\boldsymbol{z}) + \bar{R}_{\boldsymbol{X}} \bar{\boldsymbol{X}}_{\boldsymbol{\Theta}} + \bar{R}_{\boldsymbol{X}+} \bar{\boldsymbol{X}}_{\boldsymbol{\Theta}} \rho_{\boldsymbol{\Theta}} + \bar{R}_{\boldsymbol{X}+} \bar{\boldsymbol{X}}_{\Lambda} \mathsf{P} \bar{\boldsymbol{X}}_{\boldsymbol{\Theta}} + \bar{R}_{\boldsymbol{\Theta}} = 0.$$

yields a linear equation for $\bar{\boldsymbol{X}}_{\boldsymbol{\Theta}}$.

Finally to determine the Fréchet derivative, we differentiate (51) along the direction Δ. Doing so yields

$$\left(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) \right) \partial \bar{\boldsymbol{x}}(\boldsymbol{z}) \cdot \Delta + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}) \partial \bar{\boldsymbol{x}}(\boldsymbol{z}) \cdot \partial \bar{\Omega} \cdot \Delta + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}) \bar{\boldsymbol{x}}_{\Lambda}(\boldsymbol{z}) \mathsf{P} \partial \bar{\boldsymbol{X}} \cdot \Delta + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z}) \partial \bar{\boldsymbol{X}} \cdot \Delta = 0.$$

We first derive an analogue of the property $\partial \bar{\Omega} \cdot \Delta = \Delta$. This holds in the simple section 3.1 economy but fails in the more general case. We proceed by showing that we can evaluate $\partial \bar{\Omega}$ along a direction Δ^j that satisfies the property that there exists a function $a(\cdot)$ such

[36] Easiest to exploit $\rho_{\boldsymbol{\Theta}} = \begin{pmatrix} \rho_{\Theta} & 0 \\ 0 & \rho_{\Phi} \end{pmatrix}$ and solve for each column of $\bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}$ separately.

that the density of Δ^j takes the form

$$\frac{\partial}{\partial \boldsymbol{y}^j}\left(a(\boldsymbol{y})\bar{\omega}(\boldsymbol{y})\right)$$

Begin by differentiating the law of motion for $\tilde{\Omega}$ at $\sigma = 0$. Since $\partial \bar{\boldsymbol{z}} = 0$, we get

$$\begin{aligned}
(\partial\bar{\Omega}\cdot\Delta^j)(\boldsymbol{y}) &= \int \prod_i \iota(\bar{\boldsymbol{z}}^i(\boldsymbol{z}) \le \boldsymbol{y}^i)\frac{\partial}{\partial \boldsymbol{z}^j}\left(a(\boldsymbol{z})\bar{\omega}(\boldsymbol{z})\right) d\boldsymbol{z} \\
&= \int \sum_i \delta(\bar{\boldsymbol{z}}^i(\boldsymbol{z}) - \boldsymbol{y}^i)\prod_{k\neq i}\iota(\bar{\boldsymbol{z}}^k(\boldsymbol{z}) \le \boldsymbol{y}^k)\frac{\partial \bar{\boldsymbol{z}}^i}{\partial \boldsymbol{z}^j}(\boldsymbol{z})a(\boldsymbol{z})\bar{\omega}(\boldsymbol{z})d\boldsymbol{z}. \\
&= \bar{\boldsymbol{z}}^j_{\boldsymbol{z}}\int \delta(\bar{\boldsymbol{z}}^j - \boldsymbol{y}^j)\prod_{k\neq j}\iota(\bar{\boldsymbol{z}}^k \le \boldsymbol{y}^k)a(\boldsymbol{z})\bar{\omega}(\boldsymbol{z})d\boldsymbol{z}
\end{aligned}$$

where the second line was achieved through integration by parts. The third line was achieved by noting that $\bar{\omega}$ is the density of the steady state so $\bar{\boldsymbol{z}}(\boldsymbol{z}) = \boldsymbol{z}$ for all $\boldsymbol{z}$ in its support and exploiting that $\frac{\partial \bar{\boldsymbol{z}}^i}{\partial \boldsymbol{z}^j}(\boldsymbol{z})$ is both independent of $\boldsymbol{z}$ and diagonal. We can also compute the density of $(\partial\bar{\Omega}\cdot\Delta^j)(\boldsymbol{y})$ by applying the derivative $\frac{\partial^{n_z}}{\partial \boldsymbol{y}^1 \partial \boldsymbol{y}^2 \cdots \partial \boldsymbol{y}^{n_z}}$ which gives

$$\bar{\boldsymbol{z}}^j_{\boldsymbol{z}}\frac{\partial}{\partial \boldsymbol{y}^j}\int \prod_k \boldsymbol{\delta}(\boldsymbol{z}^k - \boldsymbol{y}^k)a(\boldsymbol{z})\bar{\omega}(\boldsymbol{z})d\boldsymbol{z} = \bar{\boldsymbol{z}}^j_{\boldsymbol{z}}\frac{\partial}{\partial \boldsymbol{y}^j}\left(a(\boldsymbol{y})\bar{\omega}(\boldsymbol{y})\right).$$

We conclude that $\partial\bar{\Omega}\cdot\Delta^j = \bar{\boldsymbol{z}}^j_{\boldsymbol{z}}\Delta^j$.

Evaluating the Fréchet derivative of (51) in this particular direction Δ^j we find

$$\left(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p} + \boldsymbol{z}^j_{\boldsymbol{z}}\bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\right)\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\Delta^j + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}_{\Lambda}(\boldsymbol{z})\mathsf{P}\partial\bar{\boldsymbol{X}}\cdot\Delta^j + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\partial\bar{\boldsymbol{X}}\cdot\Delta^j = 0.$$

Solving for $\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\Delta^j$ we conclude that[37]

$$\begin{aligned}
\partial\bar{\boldsymbol{x}}(\boldsymbol{z})\cdot\Delta^j &= -\left(\bar{F}_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}_{\boldsymbol{z}}\mathsf{p} + \boldsymbol{z}^j_{\boldsymbol{z}}\bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\right)^{-1}\left(\bar{F}_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}_{\Lambda}(\boldsymbol{z})\mathsf{P} + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\right)\partial\bar{\boldsymbol{X}}\cdot\Delta^j \\
&\equiv \mathsf{C}^j(\boldsymbol{z})\partial\bar{\boldsymbol{X}}\cdot\Delta^j.
\end{aligned}$$

Taking the derivative of R along this direction we get

$$\begin{aligned}
\partial\bar{\boldsymbol{X}}\cdot\Delta^j &= -\left(\bar{R}_{\boldsymbol{x}}\int \mathsf{C}^j(\boldsymbol{z})d\Omega(\boldsymbol{z}) + \bar{R}_{\boldsymbol{X}+}(\boldsymbol{z}^j_{\boldsymbol{z}}I + \bar{\boldsymbol{X}}_{\Lambda}\mathsf{P}) + \bar{R}_{\boldsymbol{X}}\right)^{-1}\bar{R}_{\boldsymbol{x}}\int \bar{\boldsymbol{x}}(\boldsymbol{z})d\Delta^j(\boldsymbol{z}) \\
&\equiv \left(\mathsf{D}^j\right)^{-1}\bar{R}_{\boldsymbol{x}}\int \bar{\boldsymbol{x}}(\boldsymbol{z})d\Delta^j(\boldsymbol{z}).
\end{aligned}$$

[37] For generality we have written this as one value of C for each individual state, i.e. C^j. In fact, one only needs one value for each level of $\bar{\boldsymbol{z}}^j_{\boldsymbol{z}}$ which in our case is two: 1 and ρ_θ.

From the definition of Δ^j, we can use integration by parts to find that

$$\partial \bar{\boldsymbol{X}} \cdot \Delta^j = \left(\mathsf{D}^j\right)^{-1} \bar{R}_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}_{\boldsymbol{z}^j}(\boldsymbol{z}) a(\boldsymbol{z}) d\Omega(\boldsymbol{z}).$$

A.3.2 Expansion along the path

We will use the derivatives of the state variables at the end of the transition path to evaluate our expansion along the path using backward induction. This approach is recursive, so we'll compute the derivatives at $(\bar{\Omega}^n, \bar{\Lambda}^n, \bar{\boldsymbol{\Theta}}^n)$ assuming derivatives at period $n+1$ of the transition are known.

Differentiating (51) and (52) with respect to Λ we obtain

$$\bar{F}^n_{\boldsymbol{x}-}(\boldsymbol{z})\bar{\boldsymbol{x}}^n_\Lambda(\boldsymbol{z}) + \bar{F}^n_{\boldsymbol{x}}(\boldsymbol{z})\bar{\boldsymbol{x}}^n_\Lambda(\boldsymbol{z}) + \bar{F}^n_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}^{n+1}_\Lambda(\boldsymbol{z})\mathsf{P}\bar{\boldsymbol{X}}^n_\Lambda + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\bar{\boldsymbol{X}}^n_\Lambda = 0$$

and

$$\bar{R}^n_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}^n_\Lambda(\boldsymbol{z}) d\Omega(\boldsymbol{z}) + \bar{R}^n_{\boldsymbol{X}}\bar{\boldsymbol{X}}^n_\Lambda + \bar{R}^n_{\boldsymbol{X}+}\bar{\boldsymbol{X}}^{n+1}_\Lambda \mathsf{P}\bar{\boldsymbol{X}}^n_\Lambda + \bar{R}^n_\Lambda = 0.$$

As both $\bar{\boldsymbol{x}}^{n+1}_\Lambda(\boldsymbol{z})$ and $\bar{\boldsymbol{X}}^{n+1}_\Lambda$ are already known we can solve for $\bar{\boldsymbol{x}}^n_\Lambda(\boldsymbol{z})$ to find

$$\bar{\boldsymbol{x}}^n_\Lambda(\boldsymbol{z}) = -\left(\bar{\mathsf{F}}^n_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{\mathsf{F}}^n_{\boldsymbol{x}}(\boldsymbol{z})\right)^{-1}\left(\bar{\mathsf{F}}^n_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}^{n+1}_\Lambda(\boldsymbol{z})\mathsf{P} + \bar{\mathsf{F}}^n_{\boldsymbol{X}}(\boldsymbol{z})\right)\bar{\boldsymbol{X}}^n_\Lambda,$$

and, therefore, $\bar{\boldsymbol{X}}^n_\Lambda$ equals

$$-\left(-\bar{\mathsf{R}}^n_{\boldsymbol{x}} \int \left(\bar{\mathsf{F}}^n_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{\mathsf{F}}^n_{\boldsymbol{x}}(\boldsymbol{z})\right)^{-1}\left(\bar{\mathsf{F}}^n_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}^{n+1}_\Lambda(\boldsymbol{z})\mathsf{P} + \bar{\mathsf{F}}^n_{\boldsymbol{X}}(\boldsymbol{z})\right) d\Omega(\boldsymbol{z}) + \bar{\mathsf{R}}^n_{\boldsymbol{X}} + \bar{\mathsf{R}}^n_{\boldsymbol{X}+}\bar{\boldsymbol{X}}^{n+1}_\Lambda\mathsf{P}\right)^{-1}\bar{\mathsf{R}}^n_\Lambda$$

Differentiating with respect to $\boldsymbol{\Theta}$ we find

$$\bar{F}^n_{\boldsymbol{x}-}(\boldsymbol{z})\bar{\boldsymbol{x}}^n_{\boldsymbol{\Theta}}(\boldsymbol{z}) + \bar{F}^n_{\boldsymbol{x}}(\boldsymbol{z})\bar{\boldsymbol{x}}^n_{\boldsymbol{\Theta}}(\boldsymbol{z}) + \bar{F}^n_{\boldsymbol{x}+}(\boldsymbol{z})\left(\bar{\boldsymbol{x}}^{n+1}_{\boldsymbol{\Theta}}(\boldsymbol{z})\rho_{\boldsymbol{\Theta}} + \bar{\boldsymbol{x}}^{n+1}_\Lambda(\boldsymbol{z})\mathsf{P}\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}}\right) + \bar{F}^n_{\boldsymbol{X}}(\boldsymbol{z})\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}} + \bar{F}^n_{\boldsymbol{\Theta}}(\boldsymbol{z}) = 0.$$

This yields a linear equation in $\bar{\boldsymbol{x}}^n_{\boldsymbol{\Theta}}$ and $\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}}$ which we can solve for $\bar{\boldsymbol{x}}^n_{\boldsymbol{\Theta}}$ as a linear function of $\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}}$. Plugging in for the linear relationship between $\bar{\boldsymbol{x}}^n_{\boldsymbol{\Theta}}$ and $\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}}$ in

$$\bar{R}^n_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}^n_{\boldsymbol{\Theta}}(\boldsymbol{z}) d\Omega(\boldsymbol{z}) + \bar{R}^n_{\boldsymbol{X}}\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}} + \bar{R}^n_{\boldsymbol{X}+}\bar{\boldsymbol{X}}^{n+1}_{\boldsymbol{\Theta}}\rho_{\boldsymbol{\Theta}} + \bar{R}_{\boldsymbol{X}+}\bar{\boldsymbol{X}}^{n+1}_\Lambda\mathsf{P}\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}} + \bar{R}^n_{\boldsymbol{\Theta}} = 0$$

yields a linear equation that can be solved for $\bar{\boldsymbol{X}}^n_{\boldsymbol{\Theta}}$.

To compute the Fréchet derivative, we'll evaluate the derivative in direction $\Delta^{j,n}$ with density of the form

$$\frac{\partial}{\partial \boldsymbol{y}^j}\left(a^n(\boldsymbol{y})\bar{\omega}^n(\boldsymbol{y})\right)$$

where $\bar{\omega}^n$ is the density of $\bar{\Omega}^n$ and $a^n(\boldsymbol{y})$ is some arbitrary function. For this derivative we find

$$
\begin{aligned}
(\partial\bar{\Omega}^n \cdot \Delta^{j,n})(\boldsymbol{y}) &= \int \prod_i \iota(\bar{\boldsymbol{z}}^i(\boldsymbol{z}) \leq \boldsymbol{y}^i) \frac{\partial}{\partial \boldsymbol{z}^j} \left(a^n(\boldsymbol{z})\bar{\omega}^n(\boldsymbol{z})\right) d\boldsymbol{z} \\
&= \int \sum_i \delta(\bar{\boldsymbol{z}}^i(\boldsymbol{z}) - \boldsymbol{y}^i) \prod_{k\neq i} \iota(\bar{\boldsymbol{z}}^k(\boldsymbol{z}) \leq \boldsymbol{y}^k) \frac{\partial \bar{\boldsymbol{z}}^{i,n}}{\partial \boldsymbol{z}^j}(\boldsymbol{z}) a^n(\boldsymbol{z})\bar{\omega}^n(\boldsymbol{z}) d\boldsymbol{z}. \\
&= \bar{\boldsymbol{z}}^j_{\boldsymbol{z}} \int \delta(\bar{\boldsymbol{z}}^i(\boldsymbol{z}) - \boldsymbol{y}^j) \prod_{k\neq j} \iota(\bar{\boldsymbol{z}}^i(\boldsymbol{z}) \leq \boldsymbol{y}^k) a^n(\boldsymbol{z})\bar{\omega}^n(\boldsymbol{z}) d\boldsymbol{z}.
\end{aligned}
$$

The density of $(\partial\bar{\Omega}^n \cdot \Delta^{j,n})(\boldsymbol{y})$ is found by applying the derivative $\frac{\partial^{n_z}}{\partial \boldsymbol{y}^1 \partial \boldsymbol{y}^2 \cdots \partial \boldsymbol{y}^{n_z}}$ to get

$$
\bar{\boldsymbol{z}}^j_{\boldsymbol{z}} \frac{\partial}{\partial \boldsymbol{y}^j} \int \prod_k \boldsymbol{\delta}(\bar{\boldsymbol{z}}^k - \boldsymbol{y}^k) a^n(\boldsymbol{z})\bar{\omega}^n(\boldsymbol{z}) d\boldsymbol{z} = \bar{\boldsymbol{z}}^j_{\boldsymbol{z}} \frac{\partial}{\partial \boldsymbol{y}^j} \left(a^{n+1}(\boldsymbol{y})\bar{\omega}^{n+1}(\boldsymbol{y})\right),
$$

where $a^{n+1}(\boldsymbol{y}) = a^n(\bar{\boldsymbol{z}}^{-1}(\boldsymbol{y}))$. We conclude therefore that $\partial\bar{\Omega}^n \cdot \Delta^{j,n} = \bar{\boldsymbol{z}}^j_{\boldsymbol{z}} \Delta^{j,n+1}$ were we acknowledge the implicit relationship between $\Delta^{j,n}$ and $\Delta^{j,n+1}$ through $a^{n+1}(\boldsymbol{y}) = a^n(\bar{\boldsymbol{z}}^{-1}(\boldsymbol{y}))$.

The Fréchet derivative of F then is

$$
\begin{aligned}
\left(\bar{F}^n_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}^n_{\boldsymbol{x}}(\boldsymbol{z})\right) \partial\bar{\boldsymbol{x}}^n(\boldsymbol{z}) \cdot \Delta^{j,n} + \boldsymbol{z}^j_{\boldsymbol{z}} \bar{F}_{\boldsymbol{x}+}(\boldsymbol{z}) \partial\bar{\boldsymbol{x}}^{n+1}(\boldsymbol{z}) \cdot \Delta^{j,n+1} \\
+ \bar{F}^n_{\boldsymbol{x}+}(\boldsymbol{z}) \bar{\boldsymbol{x}}^{n+1}_{\Lambda}(\boldsymbol{z}) \mathsf{P} \partial\bar{\boldsymbol{X}}^n \cdot \Delta^{j,n} + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z}) \partial\bar{\boldsymbol{X}} \cdot \Delta^{j,n} = 0.
\end{aligned}
$$

Since the previous equation is recursive in $\partial\bar{\boldsymbol{x}}^n(\boldsymbol{z}) \cdot \Delta^{j,n}$, we can solve it forward to obtain

$$
\partial\bar{\boldsymbol{x}}^n(\boldsymbol{z}) \cdot \Delta^{j,n} = \sum_{k=0}^{N-n} \mathsf{C}^{j,n}_k(\boldsymbol{z}) \partial\bar{\boldsymbol{X}}^{n+k} \cdot \Delta^{j,n+k}
$$

with $\mathsf{C}^{j,N}_0$ defined from C^j in the previous section and

$$
\begin{aligned}
\mathsf{C}^{j,n}_0(\boldsymbol{z}) &= -\left(\bar{F}^n_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}^n_{\boldsymbol{x}}(\boldsymbol{z})\right)^{-1} \left(\bar{F}^n_{\boldsymbol{x}+}(\boldsymbol{z}) \bar{\boldsymbol{x}}^{n+1}_{\Lambda}(\boldsymbol{z}) \mathsf{P} + \bar{F}_{\boldsymbol{X}}(\boldsymbol{z})\right) \\
\mathsf{C}^{j,n}_k(\boldsymbol{z}) &= -\boldsymbol{z}^j_{\boldsymbol{z}} \left(\bar{F}^n_{\boldsymbol{x}-}(\boldsymbol{z}) + \bar{F}^n_{\boldsymbol{x}}(\boldsymbol{z})\right)^{-1} \mathsf{C}^{j,n+1}_{k-1}(\boldsymbol{z}).
\end{aligned}
$$

Similarly, differentiating R generates

$$
\begin{aligned}
\bar{R}^n_{\boldsymbol{x}} \int \partial\bar{\boldsymbol{x}}^n(\boldsymbol{z}) \cdot \Delta^{j,n} d\Omega^n(\boldsymbol{z}) + \boldsymbol{z}^j_{\boldsymbol{z}} \bar{R}^n_{\boldsymbol{X}+} \partial\bar{\boldsymbol{X}}^{n+1} \cdot \Delta^{j,n+1} + \bar{\boldsymbol{X}}^{n+1}_{\Lambda} \mathsf{P} \partial\bar{\boldsymbol{X}}^n \cdot \Delta^{j,n}) \\
+ \bar{R}^n_{\boldsymbol{X}}(\boldsymbol{z}) \partial\bar{\boldsymbol{X}}^n \cdot \Delta^{j,n} + \bar{R}^n_{\boldsymbol{x}} \int \bar{\boldsymbol{x}}^n(\boldsymbol{z}) d\Delta^{j,n}(\boldsymbol{z}) = 0
\end{aligned}
$$

Substituting for $\partial\bar{\boldsymbol{x}}^n(\boldsymbol{z})\cdot\Delta^{j,n}$ yields a recursive equation with solution

$$\partial\bar{\boldsymbol{X}}^n\cdot\Delta^{j,n}=-\left(\mathsf{D}^{j,n}\right)^{-1}\left(\bar{R}^n_{\boldsymbol{x}}\int\bar{\boldsymbol{x}}^n(\boldsymbol{z})d\Delta^{j,n}+\sum_{k=1}^{N-n}\mathsf{E}^{j,n}_k\partial\bar{\boldsymbol{X}}^{n+k}\cdot\Delta^{j,n+k}\right)$$

with $\mathsf{D}^{j,N}$ defined by D^j in the previous section and

$$D^{j,n}=\bar{R}^n_{\boldsymbol{x}}(\boldsymbol{z})\int\mathsf{C}^{j,n}_0(\boldsymbol{z})d\Omega^n(\boldsymbol{z})+\bar{R}^n_{\boldsymbol{X}+}\bar{\boldsymbol{X}}^{n+1}_\Lambda\mathsf{P}+\bar{R}^n_{\boldsymbol{X}}$$

and

$$\mathsf{E}^{j,n}_k=\bar{R}^n_{\boldsymbol{x}}\int\mathsf{C}^{j,n}_k(\boldsymbol{z})d\Omega^n(\boldsymbol{z})+\mathbf{1}_{k=1}\boldsymbol{z}^j_{\boldsymbol{z}}\bar{R}^n_{\boldsymbol{X}+}.$$

Finally we can use this knowledge to solve for $\bar{\boldsymbol{X}}_{\mathcal{E}}$. We'll give expressions for $\bar{\boldsymbol{X}}^0_{\mathcal{E}}$ all others are analogous. Differentiating with respect to $\mathcal{E}$ yields

$$\bar{\mathsf{F}}^0_{\boldsymbol{x}}(\boldsymbol{z})\bar{\boldsymbol{x}}^0_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})+\bar{\mathsf{F}}^0_{\boldsymbol{x}+}(\boldsymbol{z})\left(\bar{\boldsymbol{x}}^1_{\boldsymbol{\Theta}}(\boldsymbol{z})+\bar{\boldsymbol{x}}^1_{\boldsymbol{z}}(\boldsymbol{z})\mathsf{p}\bar{\boldsymbol{x}}^0_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})+\partial\bar{\boldsymbol{x}}^1(\boldsymbol{z})\cdot\bar{\Omega}^0_{\boldsymbol{\mathcal{E}}}+\bar{\boldsymbol{x}}^1_\Lambda(\boldsymbol{z})\mathsf{P}\bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}}\right)+\bar{\mathsf{F}}^0_{\boldsymbol{X}}(\boldsymbol{z})\bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}}+\bar{\mathsf{F}}^0_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})=0. \tag{54}$$

In order to proceed, we need to determine $\bar{\Omega}^0_{\boldsymbol{\mathcal{E}}}$. Differentiating the law of motion of Ω gives

$$\bar{\Omega}^0_{\boldsymbol{\mathcal{E}}}=-\sum_j\int\delta(\bar{z}^j(\boldsymbol{z})-\boldsymbol{y}^j)\prod_i\iota(\bar{z}^i(\boldsymbol{z})\leq\boldsymbol{y}^i)\bar{\boldsymbol{z}}^{j,0}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})\bar{\omega}^0(\boldsymbol{z})d\boldsymbol{z}.$$

The density of $\bar{\Omega}^0_{\boldsymbol{\mathcal{E}}}$ is therefore

$$-\sum_j\frac{\partial}{\partial\boldsymbol{y}^j}\int\prod_i\delta(\bar{z}^i(\boldsymbol{z})-\boldsymbol{y}^i)\bar{z}^{j,0}_{\mathcal{E}}(\boldsymbol{z})\bar{\omega}^0(\boldsymbol{z})d\boldsymbol{z}=-\sum_j\frac{\partial}{\partial\boldsymbol{y}^j}\left(\bar{\boldsymbol{z}}^{j,0}_{\boldsymbol{\mathcal{E}}}\left(\bar{\boldsymbol{z}}^{-1}(\boldsymbol{y})\right)\bar{\omega}^1(\boldsymbol{y})\right)\equiv\sum_j\bar{\omega}^{0,j,1}_{\boldsymbol{\mathcal{E}}},$$

where 1 here represents that the objects are evaluated using the density of the transition path at time 1, $\bar{\omega}^1(\boldsymbol{y})$. If we define $\bar{\Omega}^{0,j,n}_{\boldsymbol{\mathcal{E}}}$ as the measure with density

$$\bar{\omega}^{0,j,n}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y})=-\frac{\partial}{\partial\boldsymbol{y}^j}\left(\bar{\boldsymbol{z}}^{j,0}_{\boldsymbol{\mathcal{E}}}\left(\underbrace{\bar{\boldsymbol{z}}^{-1}\left(\ldots\bar{\boldsymbol{z}}^{-1}(\boldsymbol{y})\right)}_{n\text{ times}}\right)\bar{\omega}^n(\boldsymbol{y})\right),$$

then

$$\partial\bar{\boldsymbol{x}}^1(\boldsymbol{z})\cdot\bar{\Omega}^{0,j,1}_{\boldsymbol{\mathcal{E}}}=\sum_{k=0}^{N-1}\mathsf{C}^{j,1}_k(\boldsymbol{z})\partial\bar{\boldsymbol{X}}^{1+k}\cdot\bar{\Omega}^{0,j,1+k}_{\boldsymbol{\mathcal{E}}}\equiv\sum_{k=0}^{N-1}\mathsf{C}^{j,1}_k\bar{\boldsymbol{X}}'^{j,1+k}_{\boldsymbol{\mathcal{E}}}.$$

Combined with (54) gives a linear system

$$\mathsf{M}^0(\boldsymbol{z})\bar{\boldsymbol{x}}^0_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})=\mathsf{N}^0(\boldsymbol{z})\begin{bmatrix}I & \bar{\boldsymbol{X}}'^{1,1}_{\boldsymbol{\mathcal{E}}} & \bar{\boldsymbol{X}}'^{2,1}_{\boldsymbol{\mathcal{E}}} & \ldots & \bar{\boldsymbol{X}}'^{n_z,N}_{\boldsymbol{\mathcal{E}}}\end{bmatrix}^{\mathsf{T}}$$

and which can be solved for $\bar{\boldsymbol{x}}^0_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})$. To find $\bar{\boldsymbol{X}}'^{j,n}_{\boldsymbol{\mathcal{E}}}$, we note that they satisfy the equation

$$\bar{\boldsymbol{X}}'^{j,n}_{\boldsymbol{\mathcal{E}}} = -\left(\mathsf{D}^{j,n}\right)^{-1}\left(\bar{R}^n_{\boldsymbol{z}^j} + \bar{R}^n_{\boldsymbol{x}}\int\left[\bar{\boldsymbol{x}}^n_{\boldsymbol{z}^j}(\boldsymbol{z})\bar{\boldsymbol{z}}^{j,0}_{\boldsymbol{\mathcal{E}}}\left(\underbrace{\bar{\boldsymbol{z}}^{-1}\left(\ldots\bar{\boldsymbol{z}}^{-1}(\boldsymbol{z})\right)}_{n \text{ times}}\right)\right]d\Omega^n(\boldsymbol{z}) + \sum_{k=1}^{N-n}\mathsf{E}^{j,n}_k\bar{\boldsymbol{X}}'^{j,n+k}_{\boldsymbol{\mathcal{E}}}\right).$$

Combining the previous equation with

$$\bar{\mathsf{R}}^0_{\boldsymbol{x}}\int\bar{\boldsymbol{x}}^0_{\boldsymbol{\mathcal{E}}}(\boldsymbol{z})d\Omega^0(\boldsymbol{z}) + \bar{\mathsf{R}}^0_{\boldsymbol{X}}\bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}} + \bar{\mathsf{R}}^0_{\boldsymbol{X}+}\left(\bar{\boldsymbol{X}}^1_{\boldsymbol{\Theta}} + \bar{\boldsymbol{X}}^1_{\boldsymbol{\mathcal{E}}} + \bar{\boldsymbol{X}}^1_{\Lambda}\mathsf{P}\bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}}\right) + \bar{\mathsf{R}}^0_{\boldsymbol{\mathcal{E}}} = 0$$

yields a linear system

$$\mathsf{O}\cdot\left[\begin{array}{ccccc}\bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}} & \bar{\boldsymbol{X}}'^{1,1}_{\boldsymbol{\mathcal{E}}} & \bar{\boldsymbol{X}}'^{2,1}_{\boldsymbol{\mathcal{E}}} & \ldots & \bar{\boldsymbol{X}}'^{n_z,N}_{\boldsymbol{\mathcal{E}}}\end{array}\right]^{\mathsf{T}} = \mathsf{P}$$

which can be solved for $\bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}}$.

The term $\bar{\boldsymbol{x}}^0_{\boldsymbol{\varepsilon}}(\boldsymbol{z})$ satisfies

$$\bar{\boldsymbol{x}}^0_{\boldsymbol{\varepsilon}}(\boldsymbol{z}) = \left(\bar{F}^0_{\boldsymbol{x}}(\boldsymbol{z}) + \bar{F}^0_{\boldsymbol{x}+}(\boldsymbol{z})\bar{\boldsymbol{x}}^1_{\boldsymbol{z}}(\bar{\boldsymbol{z}}(\boldsymbol{z}))\mathsf{p}\right)^{-1}\bar{F}^0_{\boldsymbol{\varepsilon}}(\boldsymbol{z}).$$

A.3.3 An Alternative Approximation

In this section we present the alternative approach highlighted in section 3.2 where we scale $\{\sigma\boldsymbol{\mathcal{E}}, \sigma\varepsilon, \sigma\boldsymbol{\Theta}, \sigma\theta\}$ and expand with respect to σ instead of just $\{\sigma\boldsymbol{\mathcal{E}}, \sigma\varepsilon\}$.

For this approach, the full policy function for $\boldsymbol{X}$ can be written as $\tilde{\boldsymbol{X}}(\Omega(\sigma), \Lambda, \sigma\boldsymbol{\Theta}, \sigma\boldsymbol{\mathcal{E}}; \sigma)$ where $\Omega(\boldsymbol{y};\sigma)$ incorporates the fact that we are scaling θ with σ and therefore also scaling Ω. Formally, we have (assuming the simplest case where m, μ and θ are the only individual state variables for our problem)

$$\Omega(\mathbf{y};\sigma) = \int\iota(m \leq y_1)\iota(\mu \leq y_2)\iota(\sigma\theta \leq y_3)d\Omega(m,\mu,\theta). \tag{55}$$

The same proof can be used to show that Lemma 2 holds for this approximation as well. There still may be transition dynamics with respect to Λ, at which point it will be necessary to follow sections A.3.1 and A.3.2 to compute the relevant derivatives for the expansion.[38]

$\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{x}}$ can then be approximated using Taylor expansions with respect to σ. For brevity we only report the first order expansion of $\tilde{\boldsymbol{X}}$ which given by

$$\tilde{\boldsymbol{X}}(\Omega(\sigma), \Lambda, \sigma\boldsymbol{\Theta}, \sigma\boldsymbol{\mathcal{E}}; \sigma) = \bar{\boldsymbol{X}}^0 + \sigma(\partial\bar{\boldsymbol{X}}^0\cdot\bar{\Omega}_\sigma + \bar{\boldsymbol{X}}^0_{\boldsymbol{\Theta}}\boldsymbol{\Theta} + \bar{\boldsymbol{X}}^0_{\boldsymbol{\mathcal{E}}}\boldsymbol{\mathcal{E}} + \bar{\boldsymbol{X}}^0_\sigma) + \mathcal{O}(\sigma^2).$$

[38] In the case where there is no endogenous aggregate sate variable only section A.3.1 is required.

To obtain $\partial\bar{\boldsymbol{X}}\cdot\bar{\Omega}_\sigma$, we differentiate (55) with respect to σ to obtain

$$\bar{\Omega}_\sigma(\mathbf{y}) = -\int \iota(m \le y_1)\iota(\mu \le y_2)\delta(0 - y_3)\theta d\Omega(m,\mu,\theta).$$

The density of this object is constructed by applying the derivative $\frac{\partial^3}{\partial y_1 \partial y_2 \partial y_3}$ to get

$$\begin{aligned}
\bar{\omega}_\sigma(\mathbf{y}) &= -\frac{\partial}{\partial y_3}\left(\int \delta(m - y_1)\delta(\mu - y_2)\delta(0 - y_3)\theta d\Omega(m,\mu,\theta)\right)\\
&= -\frac{\partial}{\partial y_3}\left(\delta(0 - y_3)\int \omega(y_1, y_2, \theta)\theta d\theta\right)\\
&= -\frac{\partial}{\partial y_3}\left(E\theta(y_1,y_2)\bar{\omega}(\mathbf{y})\right)
\end{aligned}$$

where in the last equality we defined $E\theta(y_1, y_2) = \frac{\int \omega(y_1,y_2,\theta)\theta d\theta}{\int \omega(y_1,y_2,\theta) d\theta}$ as the cross-sectional mean of θ conditional on $(m,\mu) = (y_1,y_2)$. From this expression, we know that $\partial\bar{\boldsymbol{X}}^0\cdot\bar{\Omega}_\sigma$ can be solved for in the same manner as $\partial\bar{\boldsymbol{X}}^0\cdot\bar{\Omega}^0_{\boldsymbol{\mathcal{E}}}$ using the tools in section (A.3.2).

A.3.4 Simulation and Clustering

To simulate an optimal policy at each date with N agents, we discretize the distribution across agents with K grid points that we find each period using a k-means clustering algorithm. Let $\{\boldsymbol{z}_i\}_{i=1}^N$ represent the current distribution of agents. The k-means algorithm generates K points $\{\bar{\boldsymbol{z}}_k\}_{k=1}^K$ with each agent i assigned to a cluster $k(i)$ to minimize the squared error $\sum_i \|\boldsymbol{z}_i - \bar{\boldsymbol{z}}_{k(i)}\|^2$. We let Ω represent the distribution of N agents and $\bar{\Omega}$ represent our approximating distribution of clusters.[39] At each history, we compute $\bar{\Omega}$ and then apply our algorithm to approximate the optimal policies around $\bar{\Omega}$.[40] When $K = N$ we exactly approximate around Ω, but for $K < N$ we can speed up the computations by a factor of $\frac{N}{K}$.

A.3.5 Solving the $t = 0$ problem

For the Ramsey problem (21), optimality conditions at $t = 0$ are different from $t \ge 1$. The full set of optimality conditions are represented by expanding equations (22)–(24). We describe how to apply our procedure for the section 3.1 simple case. The extension to the general problem in section 2 is straightforward.

We start with some notation. Let Ω^B be a measure over the claims to risk-free debt. Denote the $t = 0$ aggregate policy functions as $\tilde{\boldsymbol{X}}_0\left(\Omega^B, \boldsymbol{\mathcal{E}}_0\right)$ and individual policy functions

[39] Formally $\Omega(\boldsymbol{z})$ has density $\sum_i \frac{1}{N}\delta(\boldsymbol{z} - \boldsymbol{z}_i)$ while $\bar{\Omega}(\boldsymbol{z})$ has density $\sum_i \frac{1}{N}\delta(\boldsymbol{z} - \bar{\boldsymbol{z}}_{k(i)})$.

[40] Similar to section (A.3.3) this is done by constructing a distribution $\Omega(\sigma)$ with density $\sum_i \frac{1}{N}\delta(\boldsymbol{z} - \bar{\boldsymbol{z}}_{k(i)} - \sigma(\boldsymbol{z}_i - \bar{\boldsymbol{z}}_{k(i)}))$ and then computing $\partial\bar{\boldsymbol{X}}^0\cdot\bar{\Omega}_\sigma$in the same manner as section (A.3.3).

as $\tilde{\boldsymbol{x}}_0\left(b, \Omega^B, \varepsilon_0, \boldsymbol{\mathcal{E}}_0\right)$. Augment the system (22)–(24) with mappings F_0 and R_0, capturing the time 0 first order conditions, such that

$$F_0\left(\tilde{\boldsymbol{x}}_0, \mathbb{E}_+\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{X}}_0, \varepsilon_0, \boldsymbol{\mathcal{E}}_0, b_0\right) = \mathbf{0} \tag{56}$$

$$R_0\left(\int \tilde{\boldsymbol{x}}_0 d\Omega^B, \tilde{\boldsymbol{X}}_0, \boldsymbol{\mathcal{E}}_0\right) = \mathbf{0} \tag{57}$$

Policy functions for $t \geq 1$ individual states $\boldsymbol{z}_0 = (m_0, \mu_0)$ are components of $\tilde{\boldsymbol{x}}_{\mathbf{0}}$. Let function $\Omega_0\left(\Omega^B, \boldsymbol{\mathcal{E}}_0\right)$ map the initial condition Ω^B and aggregate shock $\boldsymbol{\mathcal{E}}_0$ to a measure Ω over z using

$$\Omega_0\left(\Omega^B, \boldsymbol{\mathcal{E}}_0\right)(\boldsymbol{z}) = \int \iota\left(\tilde{\boldsymbol{z}}_0\left(\boldsymbol{y}, \Omega^B, \varepsilon_0, \boldsymbol{\mathcal{E}}_0\right) \leq \boldsymbol{z}\right) d\Pr\left(\varepsilon_0\right) d\Omega^B\left(\boldsymbol{y}\right) \quad \forall \boldsymbol{z} \tag{58}$$

Section 3.1 characterizes the small-noise approximations of the $t \geq 1$ policy functions around an arbitrary Ω. We update Ω along the path by iterating between an approximation and a simulation step. At some $t \geq 1$, taking as input Ω_{t-1}, we draw idiosyncratic shocks ε for each agent as well as aggregate shocks $\boldsymbol{\mathcal{E}}$, and use the policy functions approximated around Ω_{t-1} to move to the next period Ω_t. All that remains to be specified is how the $t = 1$ state, Ω_0, is obtained. We do that below by constructing small-noise approximations to $t = 0$ policy functions: $\tilde{\boldsymbol{X}}_0(\Omega^B, \sigma\boldsymbol{\mathcal{E}}_0; \sigma)$ and $\tilde{\boldsymbol{x}}_0(b, \Omega^B, \sigma\varepsilon_0, \sigma\boldsymbol{\mathcal{E}}_0; \sigma)$. We present a first order expansion. Higher order expansions along the lines of A.2 are analogous.

1. Zeroth-order: For some choice of Ω^B, the $\sigma = 0$ allocation consists of $\{\bar{\boldsymbol{x}}_0(b), \bar{\boldsymbol{x}}(b)\}$ for b in support of Ω^B as well as $\left\{\bar{\boldsymbol{X}}_0, \bar{\boldsymbol{X}}\right\}$ such that

$$F_0\left(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}, \bar{\boldsymbol{X}}_0, 0, 0, b_0\right) = \mathbf{0}, \quad R_0\left(\int \bar{\boldsymbol{x}}_0(b) d\Omega^B(b), \bar{\boldsymbol{X}}_0, \boldsymbol{\mathcal{E}}_0\right) = \mathbf{0}$$

$$F\left(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{X}}, 0, 0, \bar{\boldsymbol{z}}\right) = \mathbf{0}, \quad R\left(\int \bar{\boldsymbol{x}}(b) d\Omega^B(b), \bar{\boldsymbol{X}}, 0\right) = \mathbf{0}$$

2. To compute derivatives $\left\{\bar{\boldsymbol{x}}_{0,\varepsilon}(b), \bar{\boldsymbol{x}}_{0,\boldsymbol{\mathcal{E}}}(b), \bar{\boldsymbol{x}}_{0,\sigma}(b), \bar{\boldsymbol{X}}_{0,\boldsymbol{\mathcal{E}}}, \bar{\boldsymbol{X}}_{0,\sigma}\right\}$, we use the formulas from section (A.3.2). The expressions that appear in section (A.3.2) use superscript n to denote the period of transition path for the $\sigma = 0$ allocation. We can obtain $\left\{\bar{\boldsymbol{x}}_{0,\varepsilon}(b), \bar{\boldsymbol{x}}_{0,\boldsymbol{\mathcal{E}}}(\boldsymbol{b}), \bar{\boldsymbol{x}}_{0,\sigma}(b), \bar{\boldsymbol{X}}_{0,\boldsymbol{\mathcal{E}}}, \bar{\boldsymbol{X}}_{0,\sigma}\right\}$ by using those formulas after replacing $F^0_{\bullet}$ with $F_{0,\bullet}$, $F^n_{\bullet}$ with $F_{\bullet}$ for $n \geq 1$ and similarly for $R^0_{\bullet}$ and $R^n_{\bullet}$.

3. Simulation: Draw idiosyncratic shocks ε_0 for each agent as well as aggregate shocks $\boldsymbol{\mathcal{E}}_0$ and use the approximations to policy functions

$$\tilde{\boldsymbol{X}}_0(\Omega^B, \sigma\boldsymbol{\mathcal{E}}_0; \sigma) = \bar{\boldsymbol{X}}_0 + \sigma\left(\bar{\boldsymbol{X}}_{0,\boldsymbol{\mathcal{E}}}\boldsymbol{\mathcal{E}}_0 + \bar{\boldsymbol{X}}_{0,\sigma}\right) + \mathcal{O}(\sigma^2)$$

and

$$\tilde{\boldsymbol{x}}(b, \Omega^B, \sigma\boldsymbol{\varepsilon}_0, \sigma\boldsymbol{\mathcal{E}}_0; \sigma) = \bar{\boldsymbol{x}}_0(b) + \sigma\left(\bar{\boldsymbol{x}}_{0,\boldsymbol{\varepsilon}}(b)\boldsymbol{\varepsilon}_0 + \bar{\boldsymbol{x}}_{0,\boldsymbol{\mathcal{E}}}(b)\boldsymbol{\mathcal{E}}_0 + \bar{\boldsymbol{x}}_{0,\sigma}(b)\right) + \mathcal{O}(\sigma^2)$$

to obtain the $\Omega_0(z_0)$ for $t = 1$.

A.4 Additional details for section 3.3

In this section, we provide more details concerning Acharya and Dogra (2018) "PRANK" economy, which we use as a laboratory to test the accuracy of our algorithm and compare it to alternative methods. We start with equilibrium conditions, next we discuss the calibration, and report the accuracy tests for our method. Finally, we present a simplified version of this economy for which we can solve for all gradients in closed form.

Equilibrium in the PRANK economy To obtain the PRANK setting we impose the following assumptions: (i) labor is supplied inelastically, and period utility function $U(c_t, n_t) = -\exp(-\gamma c_t)$, (ii) the distribution of shares is uniform, (iii) Idiosyncratic productivity shocks are i.i.d, and (iv) All tax rates are constant and the monetary policy follows a Taylor rule given by

$$Q_t^{-1} - 1 = a_0\left(1 + \Pi_t\right)^{a_1} \tag{59}$$

In the PRANK economy, a perfect foresight equilibrium is constructed as follows. For a sequence of innovations to TFP $\{\mathcal{E}_{\Theta,s}\}_{s=0}^{T}$, Agent is consumption $c_{i,t}$ satisfies

$$c_{i,t} = \mathcal{C}_t + \mu_t\left(\frac{b_{i,t-1}}{1+\Pi_t} + y_{i,t}\right), \tag{60}$$

where $y_{i,t} = (1 - \Upsilon_t)W_t\epsilon_{i,t}n_{i,t} + T_t + d_{i,t}$ is the households income at date t. The two parameters $\mathcal{C}_t$ and μ_t that are common to all agents are given by

$$\mu_t = \frac{\mu_{t+1}\left(\frac{Q_t}{1+\Pi_{t+1}}\right)}{1 + \mu_{t+1}\left(\frac{Q_t}{1+\Pi_{t+1}}\right)} \tag{61}$$

$$\mathcal{C}_t\left[1 + \mu_{t+1}\left(\frac{Q_t}{1+\Pi_{t+1}}\right)\right] = -\frac{1}{\gamma}\ln\beta\left(\frac{Q_t}{1+\Pi_{t+1}}\right) + \mathcal{C}_{t+1} + \mu_{t+1}\bar{y}_{t+1} - \frac{\gamma\mu_{t+1}^2\sigma_{y,t+1}^2}{2} \tag{62}$$

where $\bar{y}_{t+1} = \int y_{i,t+1}di$ is the average household income, and $\sigma_{y,t+1}^2$ is the variance in household level income. A perfect foresight equilibrium can by solving equations (59)–(62) along with equations (8), (10),(16), and (17).

We check the accuracy of our approximations using an exact solution to the perfect foresight equilibrium. Section (A.4) discusses how we calibrate the PRANK economy, section

(A.4) and (A.4) compare approximation errors using several diagnostics.

Calibration We study several cases. For the parameters that are common across these cases we use Acharya and Dogra (2018) targets which are are quite standard in the representative agent New Keynesian literature. The discount rate β to 0.96 to get a real rate of 4% per year, the elasticity of substitution parameter, Φ, to 6 to target an average markup of 20%. The share of intermediate inputs α, is set to 0.6 to target a labor income share of 2/3, and we set the adjustment cost parameter ψ to 41.6 to target a slope of the Phillips curve of 0.06. Aggregate productivity follows an AR(1) process with a decay parameter 0.73, and the standard deviation of the innovation is set to 1.23% to be consistent with de trended output per hour and we turn off the markup shocks. For the Taylor rule parameters, we set $a_1 = 1.5$ and choose a_0 to target 0% inflation rate in absence of aggregate risk. We vary the standard deviation of idiosyncratic risk, $\sigma_\epsilon \in \{0.5, 0.75, 1\}$, and the risk aversion parameter, $\gamma \in \{1, 3\}$. Our calibrationscover a range that includes Acharya and Dogra (2018) as well as what we use in our baseline section 4. Since the distribution of assets is non-stationary, we set $\Omega_0(b)$ to be Gaussian and calibrate the parameters to be consistent with the distribution of wealth in the SCF. For simulation, we approximate the distribution with 150 points and the idiosyncratic shocks with 10 point Gaussian quadrature.

Diagnostics In this section, we compare the accuracy of our policy functions in two settings. We start with a stationary environment with no aggregate risk and study the policy function for individual consumption as well as values for the aggregate variables. Then, we study impulse responses of several aggregate variables to a TFP shock. As mentioned before, the advantage of PRANK is that in both cases, the true solution can be solved for exactly.

We report three types of approximation errors for the individual policy functions that are defined in the main text. For all the experiments we use a second-order approximation of our method. As a point of comparison, we report the errors when policy functions are approximated using the Reiter-approach (also used in Acharya and Dogra (2018)) in which the no-aggregate risk economy is solved exactly and then the policy functions are linearized with respect to aggregate shocks. In all our plots, our method will be represented by a bold blue line, the Reiter approximation will be represented by a dashed black line and the exact solution will be a bold black line.

We begin with the approximation errors turning off aggregate risk. By construction, the errors for the Retier-method are zero and so we report the error diagnostics just for our method for several values of $\{\sigma_\epsilon, \gamma\}$ in table 1. The maximum percent errors in the individual policy rules for consumption relative to the exact solution are small starting at 0.0039% for our baseline calibration and raising to only 0.0328% when we double the size of

Maximum Errors (%)	Individual consumption Policy	Euler	Dyn. Euler	Agg. Output	Inflation	Interest Rate
$\gamma = 1, \sigma_\epsilon = 0.5$	0.0039	0.0031	0.0097	5.2e-6	3.1e-5	4.3e-5
$\gamma = 1, \sigma_\epsilon = 0.75$	0.0134	0.0105	0.0207	2.6e-5	1.6e-4	2.2e-4
$\gamma = 1, \sigma_\epsilon = 1.0$	0.0328	0.0249	0.0705	4.9e-4	6.9e-4	8.2e-4
$\gamma = 3, \sigma_\epsilon = 0.5$	0.0453	0.0280	0.1220	4.1e-4	2.4e-3	3.4e-3

TABLE 1: Percentage Errors in policy functions with no aggregate risk. The values reported are the maximum errors across the state space (b, ϵ). The columns "Policy", "Euler", and "Dyn. Euler" refer to the diagnostic measure $E^{pol}_{c,t}(b, \epsilon)$, $E^{EE}_{c,t}(b, \epsilon)$, and $E^{dynEE}_{c,t}(b, \epsilon)$, respectively.

the idiosyncratic risk. The Euler equation errors are comparable. We see a similar pattern in the errors for the aggregate variables, though the errors for those are an order of magnitude smaller.

Next we compare the errors in the policy functions in response to an one time one standard deviation unanticipated shock to aggregate productivity. We report these errors in table 2. For the individual consumption policy functions, the maximum errors (across the state space (b, ϵ) and across time t) for our second-order approach are comparable to the Reiter method. In fact, while the Euler equation errors $E^{EE}_{c,t}(b, \epsilon)$ for the Reiter method are generally smaller than our second-order approximation, the errors relative to the exact solution $E^{pol}_{c,t}(b, \epsilon)$ are an order of magnitude larger (0.0039% vs 0.0404%). The diagnostic errors $E^{pol}_{c,t}(b, \epsilon)$ clearly captures errors coming from aggregate shocks that are not reflected in the Euler equation errors. We also see that the dynamic Euler equation errors remain small and comparable to those of the Reiter approach, which indicates that one should not be too concerned with errors accumulating over time.

Tables 1 and 2 also report errors for the alternative calibrations where we increase risk aversion, γ, to 3. Not surprisingly increasing risk aversion leads to the largest policy errors for both our second-order approximation as well as the Retier methods, but the policy errors remain small and are comparable to those from the Reiter approach. Figure 1 reproduces the impulse responses in figure I of the main text for $\gamma = 3$. For larger values of risk aversion, we see a visible deviation of the Reiter approach from both the exact solution and our second-order approximations. The visible deviation reflects the errors in aggregates of the Reiter approach documented in Table 2.

Long run errors In the PRANK economy, individual assets follow an approximate random walk and, therefore, the distribution of individual savings drifts over time. Since our method approximates with respect to the size of idiosyncratic risk, a diagnostic for whether small errors at a point in time accumulate to large error over time, we check how well the

Maximum Errors (%)	Individual consumption Policy	Euler	Dyn. Euler	Agg. Output	Inflation	Interest Rate
2^{nd} Order						
$\gamma = 1, \sigma_\epsilon = 0.50$	0.0039	0.0031	0.0103	4.2e-6	3.1e-5	4.3e-5
$\gamma = 1, \sigma_\epsilon = 0.75$	0.0134	0.0105	0.0402	2.6e-5	1.5e-4	2.2e-4
$\gamma = 1, \sigma_\epsilon = 1.00$	0.0328	0.0249	0.0853	8.2e-5	4.9e-4	6.9e-4
$\gamma = 3, \sigma_\epsilon = 0.5$	0.0453	0.0280	0.1091	0.0011	0.0024	0.0034
Reiter-based						
$\gamma = 1, \sigma_\epsilon = 0.50$	0.0374	0.0022	0.0153	0.0616	0.0337	0.0505
$\gamma = 1, \sigma_\epsilon = 0.75$	0.0466	0.0022	0.0208	0.0610	0.0335	0.0501
$\gamma = 1, \sigma_\epsilon = 1.00$	0.0492	0.0023	0.0364	0.0602	0.0329	0.0493
$\gamma = 3, \sigma_\epsilon = 0.5$	0.0896	0.0038	0.0462	0.2252	0.1327	0.1991

TABLE 2: Percentage errors in policy functions in response to an one standard deviation unanticipated shock to aggregate TFP. The values reported are the maximum errors across states (b, ϵ) and time t. The columns "Policy", "Euler", and "Dyn. Euler" refer to the diagnostic measure $E^{pol}_{c,t}(b, \epsilon)$, $E^{EE}_{c,t}(b, \epsilon)$, and $E^{dynEE}_{c,t}(b, \epsilon)$, respectively.

approximated distribution of assets tracks the true distribution with our method as well as with the Reiter method.

In figure 2, we plot the distribution of assets obtained at $t = 250$ after a one-standard deviation shock at $t = 0$. We see the second-order approximation lines up very closely with the Reiter method and to the outcomes from the exact solution. This figure also explains the finding in section 3.3, why our method captures the the response of inequality to an unanticipated TFP shock $t = 250$ so well.

We next compare the distribution of assets after a sequence of TFP shocks in a stochastic PRANK economy. The TFP shocks follow an AR(1) and for this exercise we do not have the true solution in an analytic form. However, we can still compare our second-order approximation and the Reiter approach. In addition, we include a "hybrid" method where we take a second-order approximation with respect to idiosyncratic shocks and a first-order approximation with respect to aggregate shocks.

In figure 3, we see that the hybrid and Reiter approaches produce nearly identical distributions after these shocks, but the second-order approach delivers a tighter distribution over time. As the hybrid approach was obtained by dropping the second-order terms with respect to aggregate shocks, we take this as evidence that, in this model, ignoring those second-order terms can lead long run drift away from the true solution.

A Simplified Example To illustrate how our approach from section 3 is applied to the PRANK economy, we present a version of the PRANK economy where we can explicitly

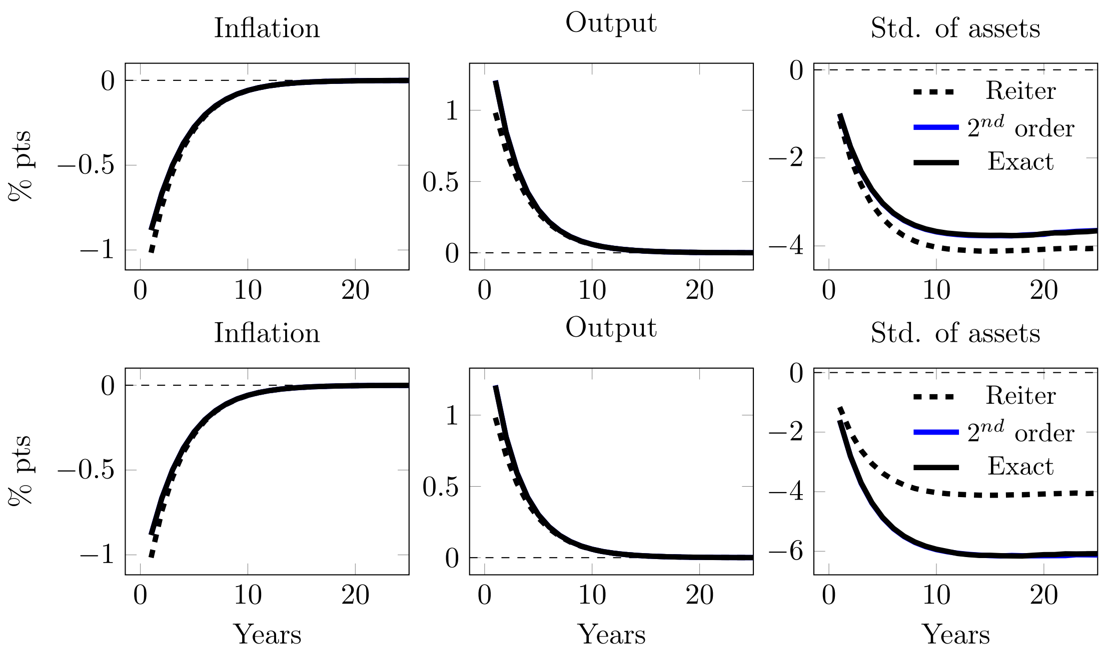

Figure 1: Comparisons for impulse responses to a 1% TFP shock at $t = 1$ in the top panel and $t = 250$ in the bottom panel

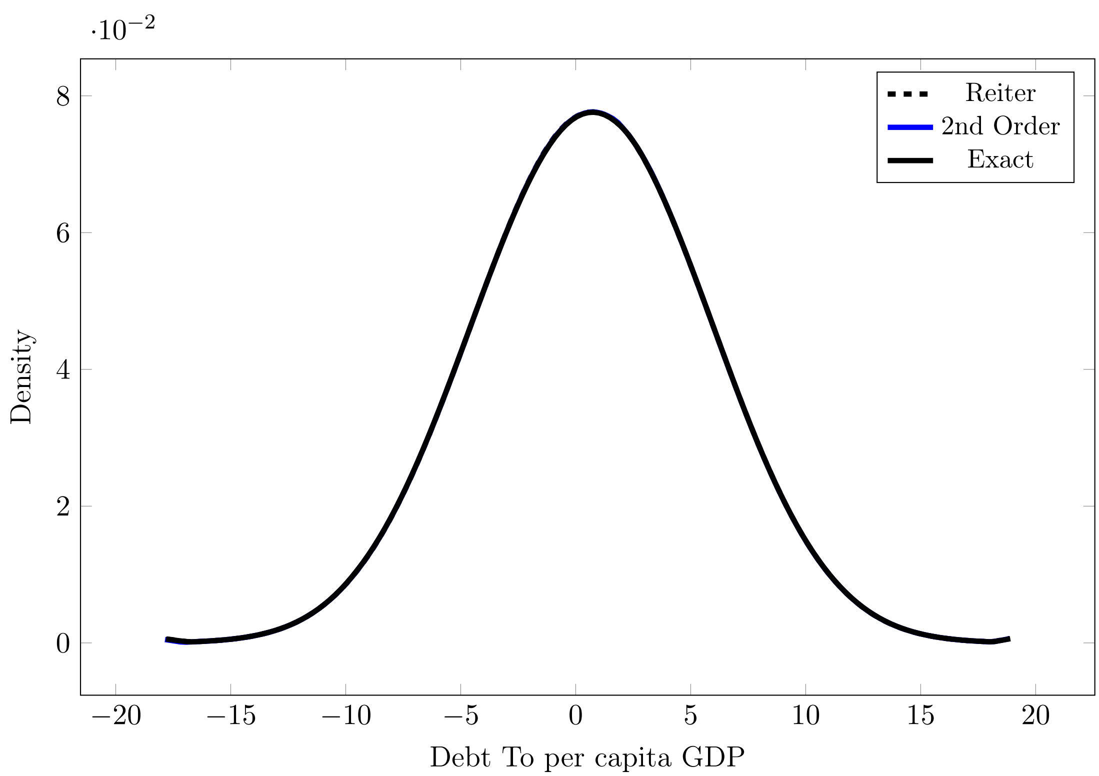

Figure 2: Distribution of assets at $t = 250$ following a one time unanticipated TFP shock at $t = 0$.

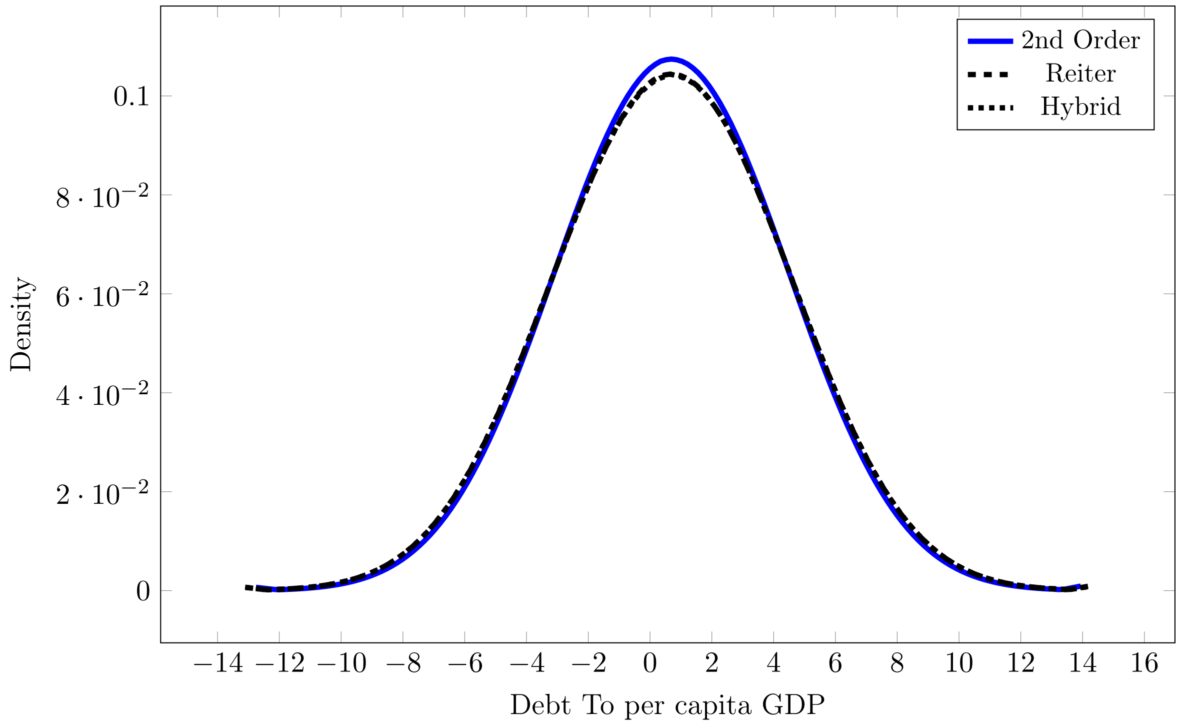

Figure 3: Distribution of assets $t = 250$ in the stochastic PRANK economy after a sequence of TFP shocks.

show how to compute all the gradients that appear that section. We assume aggressive enough monetary policy to ensure $\Pi_t = 0$ for all t; that share of intermediate inputs, $1 - \alpha$, is 0 which ensures that output is linearly related to productivity; and finally that $\Phi \to \infty$ to ensure that there are no markups and dividends. The environment is similar to the well known Huggett (1993) model and can be trivially solved with standard methods; we use it to illustrate transparently how to construct all the objects that appear in Section 3.

The economy is populated with a continuum of infinitely lived consumers who receive endowment shocks. Let $e_{i,t}$ be endowment of consumer i in period t. Endowments are subject to aggregate shock $\mathcal{E}_t$ and idiosyncratic shock $\varepsilon_{i,t}$ and satisfy

$$e_{i,t} = 1 + \varepsilon_{i,t} + \mathcal{E}_t.$$

Shocks $\mathcal{E}_t$ and $\varepsilon_{i,t}$ are mean zero and i.i.d. over time.

Competitive equilibrium in this economy is fully characterized by consumer budget constraint and the Euler equation

$$\begin{aligned} c_{i,t} + Q_t b_{i,t} - 1 - \varepsilon_{i,t} - \mathcal{E}_t - b_{i,t-1} &= 0 \\ Q_t \exp(-\gamma c_{i,t}) - \beta \mathbb{E}_{i,t} \exp(-\gamma c_{i,t}) &= 0 \end{aligned}$$

as well as the feasibility

$$\int c_{i,t} di - 1 - \mathcal{E}_t = 0.$$

We now show how to use our approximation techniques in this simple example to find competitive equilibrium. To make it similar to our notation in Section 3, let $y = \exp(-\gamma c)$ and re-write this problem as

$$\begin{aligned} c_{i,t} + Q_t b_{i,t} - 1 - \varepsilon_{i,t} - \mathcal{E}_t - b_{i,t-1} &= 0 \\ Q_t y_{i,t} - \beta \mathbb{E}_{i,t} y_{i,t+1} &= 0 \\ y_{i,t} - \exp(-\gamma c_{i,t}) &= 0 \end{aligned}$$

The first pair of equations correspond to (22) and define mapping F, the last equation corresponds to (23) and define R. This problem is recursive in the distribution of agents' assets. In our notation of Section 3 we have $\boldsymbol{z} = b$ and Ω is the distribution of b such that $\int b d\Omega = 0$. Vector $\tilde{\boldsymbol{x}}$ of individual policy functions is given by three policy functions $\begin{bmatrix} \tilde{b} & \tilde{c} & \tilde{y} \end{bmatrix}^{\mathrm{T}}$, and $\tilde{Q}$ is the only aggregate policy function in vector $\boldsymbol{X}$. Selection matrix p is simply $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$.

It is immediate to verify that without shocks consumption smoothing implies that $\bar{b}(b) = b$ for all b, so that Lemma 1 holds and equation (27) become

$$\begin{aligned} \bar{c}(b) + \bar{Q}\bar{b}(b) - 1 - b &= 0 \\ \bar{Q}\bar{y}(b) - \beta \bar{y}\left(\bar{b}(b)\right) &= 0 \\ \bar{y}(b) - \exp(-\gamma \bar{c}(b)) &= 0 \end{aligned}$$

and

$$\int \bar{c}(b) d\Omega - 1 = 0,$$

which immediately gives $\bar{Q} = \beta$ and $\bar{c}(b) = 1 + (1-\beta) b$, $\bar{y}(b) = \exp(-\gamma \bar{c}(b))$. From these we construct mappings $\mathsf{R}_{\boldsymbol{x}} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $\mathsf{R}_{\boldsymbol{X}} = 0$, $\mathsf{R}_{\boldsymbol{\mathcal{E}}} = -1$, and $\mathsf{F}_{\boldsymbol{x}-}(b) = \mathbf{0}$,

$$\mathsf{F}_{\boldsymbol{x}}(b) = \begin{bmatrix} \bar{Q} & 1 & 0 \\ 0 & 0 & \bar{Q} \\ 0 & -\gamma \exp(-\gamma \bar{c}(b)) & 1 \end{bmatrix}, \quad \mathsf{F}_{\boldsymbol{x}+}(b) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\beta \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathsf{F}_{\boldsymbol{X}}(b) = \begin{bmatrix} b \\ \bar{y}(b) \\ 0 \end{bmatrix},$$

$$\mathsf{F}_{\boldsymbol{\varepsilon}}(b) = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathsf{F}_{\boldsymbol{\mathcal{E}}}(b) = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathsf{F}_{\boldsymbol{z}}(\boldsymbol{z}) = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}.$$

All elements of these matrices are know from the zeroth order expansion. Using them, we construct first order approximations of policy functions as described in the text.

B Additional details for section 4

In this section we provide details of how we calibrate the initial distribution of nominal and real claims using the Doepke and Schneider (2006) procedure. Then we show the dynamics of the calibrated competitive equilibrium using simulations.

Initial distribution of nominal and real claims We combine the rich house-level data on financial assets from the Survey of Consumer Finances (SCF) and the aggregated Flow of Funds for intermediate investors to obtain nominal and real exposures. We start with the 2007 Wave of the SCF and restrict our sample to married households who work at least 100 hours. We drop observations where equity or bond holdings are more than 100 times the average yearly wage. These turned out to be about 0.5% of the total sample. We extract household-level data on their financial holdings and categorize them into (i) deposits, government bonds, liquid assets (net of unsecured credit), (ii) direct holdings of claims to corporate equities and corporate bonds, and (iii) indirect holdings of (i) and (ii) through mutual funds and retirement accounts.

We then use Flow of Funds data to obtain balance sheet information for private pensions (Table L.118), for state and local pensions (Table L.119) and mutual funds (Table L.122). Since pension funds have a nontrivial exposure to mutual funds and not vice versa, we start with the aggregated mutual fund balance sheet and map it into broad categories that represent deposits, corporate bonds, government bonds, corporate equities. In the year 2007, mutual funds invested 84 percent of their assets in corporate equities and bonds, 16 percent in government bonds, and other liquid claims.

We next turn pension funds and after aggregating private and public pension funds categorize the combined assets into deposits, government-issued debt, corporate debt, corporate equities, and mutual funds. For the year 2007, the pension funds assets were invested 22 percent in mutual funds, 63 percent in corporate equities and bonds, and rest 15 percent in government bonds and other liquid claims.

We define nominal claims as money-like assets plus government issued bonds and claims to real profits as corporate bonds plus corporate equities. Using the information above, we first consolidate the mutual funds into these two categories and then reassign the mutual funds to pension funds, and finally the mutual funds and pensions funds to the individuals in the SCF.

To fit initial states, we sample directly from the SCF log wages, nominal claims, and

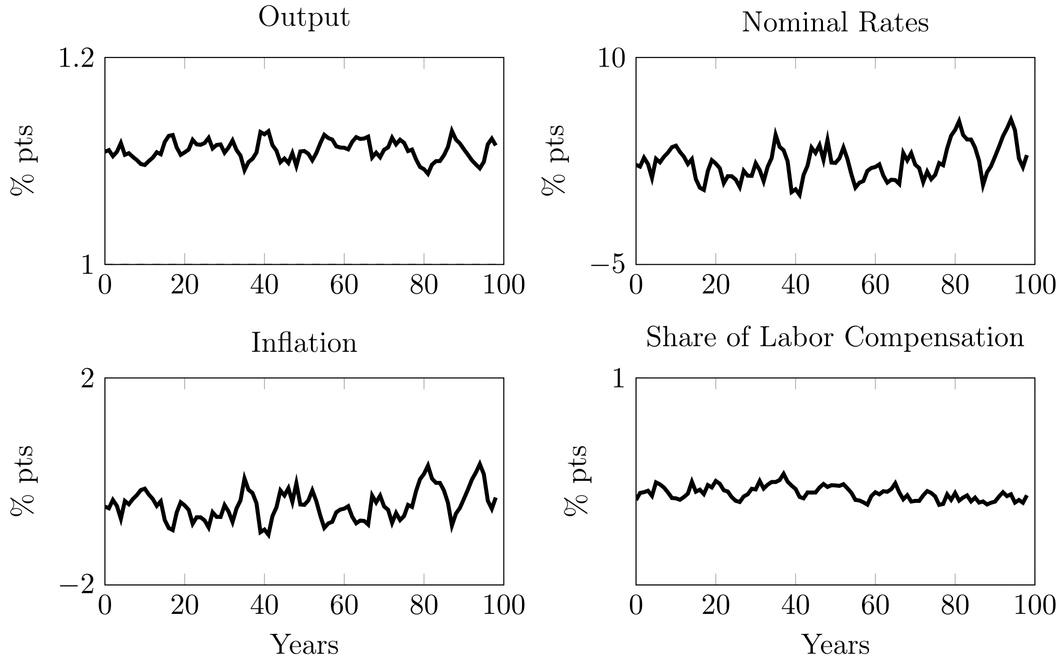

Figure 4: Simulated paths for aggregate variables using the calibrated competitive equilibrium

claims to real profits that we constructed. The SCF provides population weights for each observation. Given these weights, we set the initial condition by drawing with replacement a random sample of 100000 agents from a discrete distribution.

Properties of the competitive equilibrium In this section, we report several moments from our calibrated competitive equilibrium along a transition path. We draw a sequence of markup and TFP shocks of length 100 and simulate the competitive equilibrium policies using 100000 agents. When we simulate the competitive equilibria, we keep the tax rates $\Upsilon_t = \bar{\Upsilon}$ and use a Taylor rule with $Q_t^{-1} = \frac{1}{\beta}\Pi_t^{2.5}$.

In figure 4, we plot the time series for aggregate output and labor share. We see that the aggregates are quite stationary and exhibit small fluctuations due to productivity and markup shocks. In the table 3 we report cross-sectional moments at dates $t \in \{10, 25, 50, 75\}$. Here we notice a small drift in the distribution of the risk-free assets. The more significant drifts are in the correlations of of log wages and dividend shares as well as risk-free assets and dividend shares which steadily declines over time and the correlation between bonds and log wages that increase over time. These patterns are the outcomes of the features in the baseline that claims to equity are not traded and households are subject to natural debt limits.

TABLE 3: DISTRIBUTIONAL MOMENTS ALONG THE PATH

Moments	DATA	MODEL			
		$t = 10$	$t = 25$	$t = 50$	$t = 75$
Std. share of equities	2.63	2.62	2.62	2.62	2.62
Std. bond	6.03	6.18	6.46	7.06	7.31
Std. ln wages	0.80	0.81	0.81	0.80	0.80
Std. ln hours	0.42	0.42	0.45	0.49	0.51
Corr(share of equities, ln wages)	0.40	0.37	0.33	0.27	0.22
Corr(share of equities, bond holdings)	0.62	0.59	0.50	0.33	0.22
Corr(bond, ln wages)	0.33	0.40	0.44	0.48	0.50

Notes: The data moments correspond to SCF 2007 wave with sample restrictions explained in the text and after scaling wages, equity holdings, and debt holdings by the average yearly wage in our sample. The share of equities refers to the ratio of individual equity holdings to the total in our sample such that the weighted sum of shares equals one. The model columns correspond to simulated sample of 100000 agents using the baseline calibration from section 4.

C Additional details for section 5

C.1 Cyclical properties of optimal policies

In this section, we present the counterpart of III in the main text for two intermediate economies between our baseline HANK and RANK: (i) first, we turn off the idiosyncratic shocks, and (ii) then, we additionally allow agents trade a full set of Arrow securities. In the top panel of table 4, we see that the moments are very similar between the baseline HANK and the HANK with no idiosyncratic risk. In the bottom panel of 4, we see that HANK with complete markets is quite similar to RANK, and very different from either the baseline HANK or the HANK with no idiosyncratic risk. From this, we can deduce that optimal policies are driven mainly by how much of the aggregate shocks agents can hedge using private markets.

C.2 Example with perfectly aligned distribution of equity shares

As noted in section 6.1, the quantitative driver of the need for insurance concerns against the markup shocks is the misalignment of dividend income from labor income. To illustrate this point, we construct a calibration with non trivial amount of inequality but in which these shares are perfectly aligned. To achieve this, we take the distribution of labor productivities from the benchmark calibration; assume Pareto weights such that optimal tax rates $\bar{\Upsilon}$ equal zero; and then assign dividend shares such that individuals initial share of labor income $\frac{\epsilon_{i,o}n_{i,0}}{\int_i \epsilon_{i,0}n_{i,0}}$ equals their share of dividend income s_i. Figure 5 plots the optimal response and its when the shares are aligned. We see that the alignment of shares nearly removes all need

TABLE 4: MOMENTS

	HANK					HANK NO IDIOSYNCRATIC				
	Std.	Correlations				Std.	Correlations			
	Dev(%)	i_t	Π_t	W_t	$\ln Y_t$	Dev(%)	i_t	Π_t	W_t	$\ln Y_t$
Nominal Rate i_t	1.82	1				1.50	1			
Inflation Π_t	0.46	-0.94	1			0.41	-0.93	1		
Labor Share W_t	2.13	-0.78	0.78	1		1.83	-0.69	0.73	1	
Log Output $\ln Y_t$	0.88	-0.31	0.10	0.12	1	0.82	-0.31	0.02	0.04	1

	HANK COMPLETE MARKETS					RANK				
	Std.	Correlations				Std.	Correlations			
	Dev(%)	i_t	Π_t	W_t	$\ln Y_t$	Dev(%)	i_t	Π_t	W_t	$\ln Y_t$
Nominal Rate i_t	0.87	1				0.87	1			
Inflation Π_t	0.03	0.01	1			0.03	-0.01	1		
Labor Share W_t	1.20	-0.14	-0.25	1		1.18	-0.09	-0.32	1	
Log Output $\ln Y_t$	0.92	-0.98	-0.1	0.30	1	0.92	-0.98	-0.09	0.24	1

Notes: Moments are computed using allocations under HANK (top left); HANK without idiosyncratic shocks (top right); HANK with complete markets (bottom left); and RANK (bottom right) optimal monetary policies.

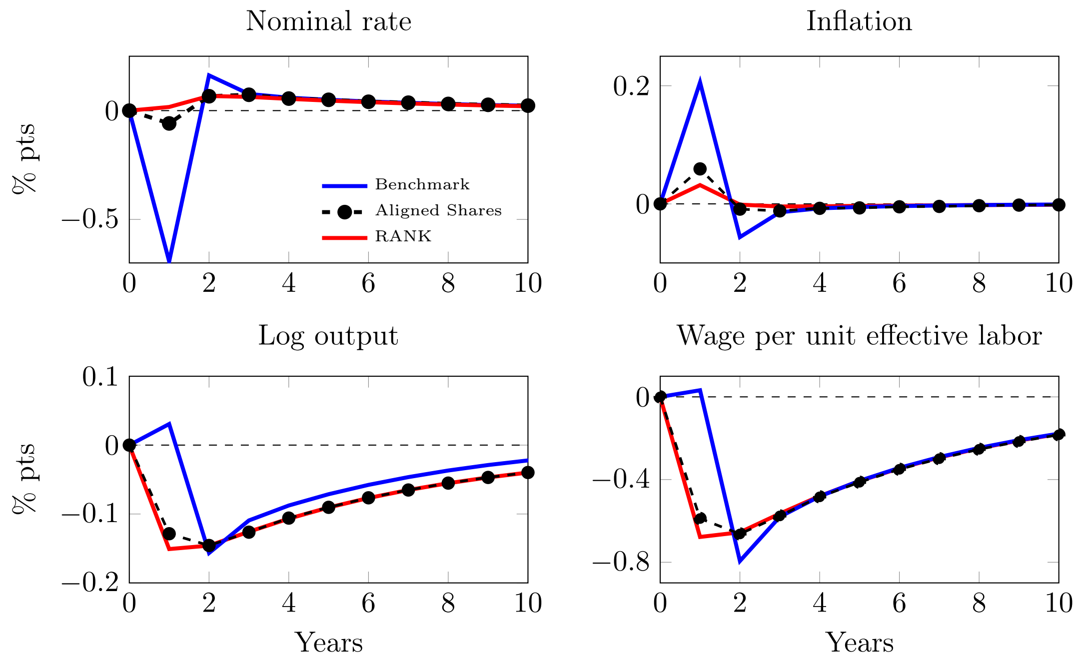

Figure 5: Optimal monetary response to a markup shock. The bold blue and red lines are the calibrated HANK and RANK responses respectively. The dashed black lines with circles are responses under HANK when the shares of labor and dividend income are aligned.

for insurance bringing the policy responses in line with those of the representative agent.[41]

D Additional details for section 6

In the main body, we focused on the results under the baseline calibration and briefly discuss sensitivity checks and special cases. In this section, we provide all the omitted details.

D.1 Sensitivity with respect to price adjustment costs

In this section, we present the impulse responses under alternative choices for the price adjustment cost parameter ψ. As mentioned in section 6.1, we vary ψ from twice the baseline calibration to one quarter of the baseline calibration, and also when ψ is approximately zero. Figure 6 plots the responses to a markup shock while figure 7 plots responses to a productivity shock.

As is readily apparent in both figures the effect on inflation is roughly linear for a large range of ψ. Doubling ψ leads to a halving of inflation while halving ψ leads to a doubling of inflation. The effect on the nominal rate is quite small. In the limit as ψ approaches zero, the planner can no longer effect real variables through monetary policy and instead

[41]There is some difference in insurance needs arising from differential labor responses to the markup shock.

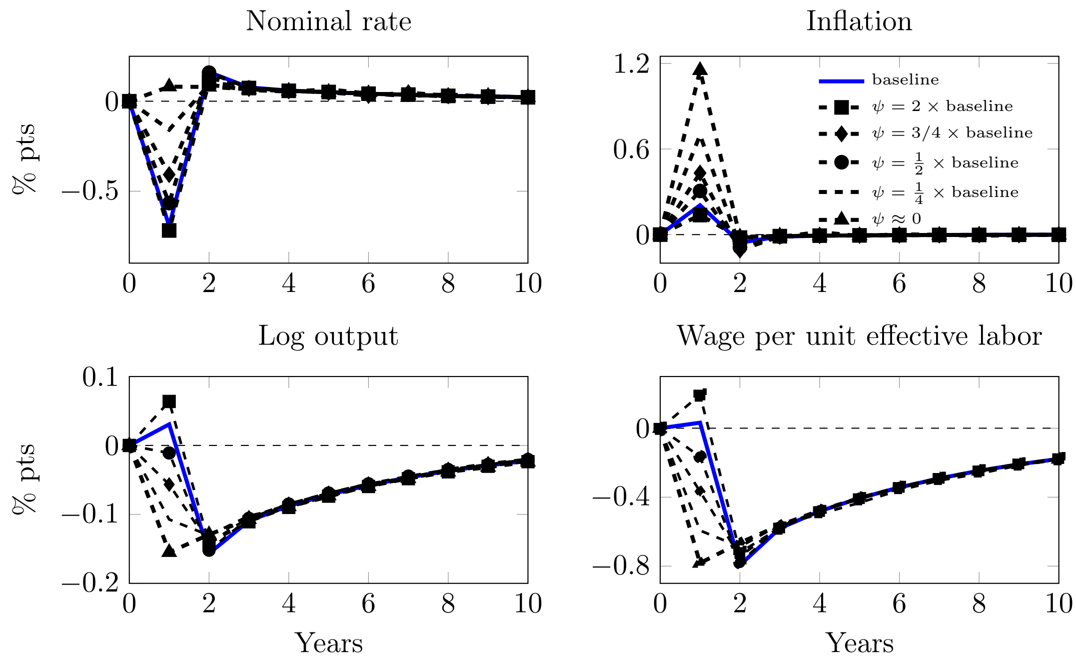

Figure 6: Optimal monetary response to a markup shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares, circles and triangles are responses under a calibration in which we double the price adjustment costs parameter, half the price adjustment cost parameter, and finally set it near zero, respectively.

relies more on unexpected inflation to provide insurance through the ex-post real return as instead of distorting the allocation by varying the ex-ante real rate.

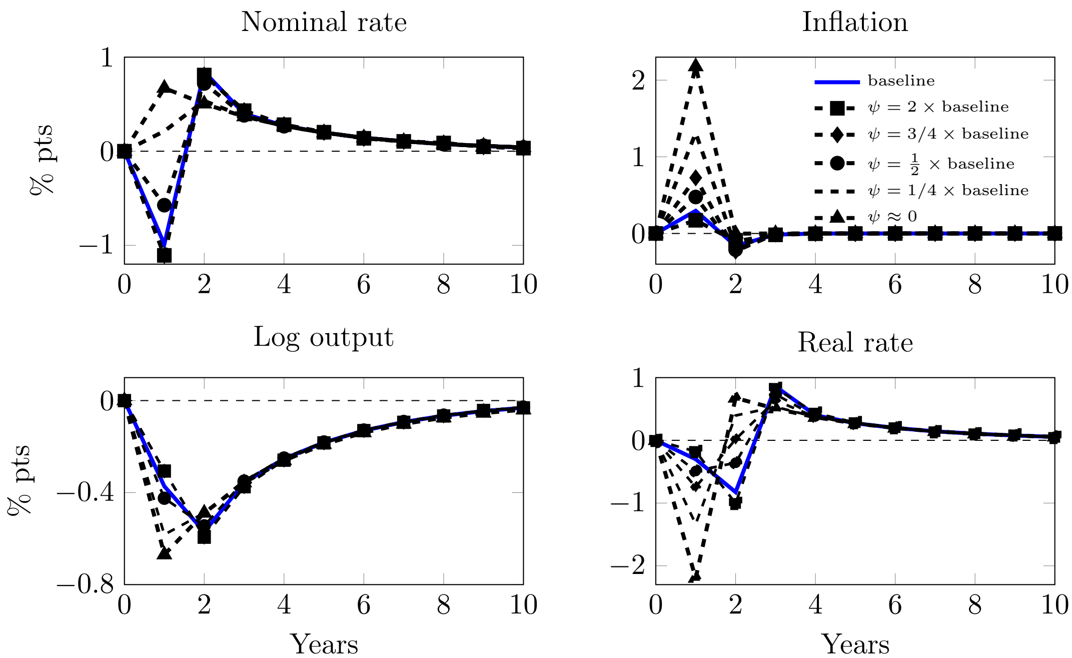

Figure 7: Optimal monetary response to a TFP shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares, circles and triangles are responses under a calibration in which we double the price adjustment costs parameter, half the price adjustment cost parameter, and finally set it near zero, respectively.

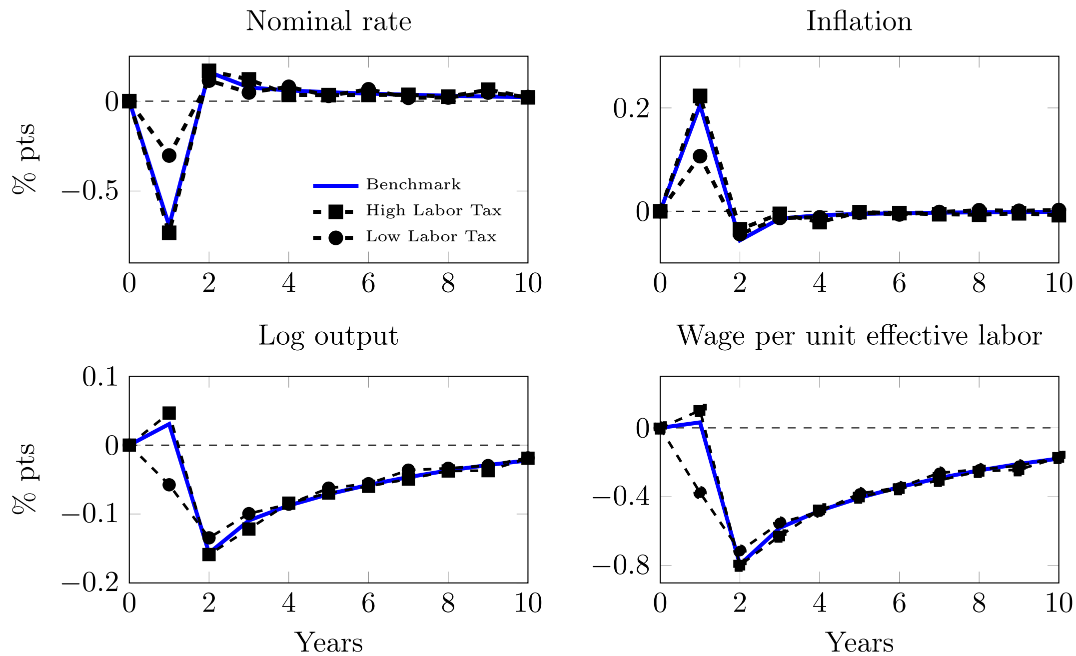

Figure 8: Optimal monetary response to a markup shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares and circles are responses under a calibration with higher and lower labor taxes respectively.

D.2 Sensitivity with respect to choice of Pareto weights

Here we present sensitivity to the choice of Pareto weights. As mention in the main text, we set Pareto weights using a thee parameter exponential specification, which loads on the three dimensions of initial heterogeneity and maps to optimal levels of tax rates $\bar{\Upsilon}$, on labor income, dividend income, and bond income. For the purpose of sensitivity, we vary these implied tax rates in a large range: from 0% to 50%. In addition, we also study a Utilitarian planner that weights all agents equally.

We start with the experiments that vary the labor income tax rate and the responses are depicted in figures 8 and 9 to markup and productivity shocks, respectively. We see that the responses to both the shocks are larger when labor tax rates are higher and lower when labor taxes are lower. Raising the labor tax compresses labor shares pushing the economy further away from full insurance, while decreasing the labor tax pushes the economy closer to full insurance. In line with this, we see that the increasing the labor tax leads to a stronger policy response while decreasing the labor tax diminishes the response.

Next we vary the tax on dividend income and report the results in figures 10 and 11. Our baseline calibration exhibits are far more unequal distribution of dividend share than labor shares. Increasing the dividend tax brings the economy closer to full insurance while decreasing the dividend tax pushes the economy away from full insurance. As such, we

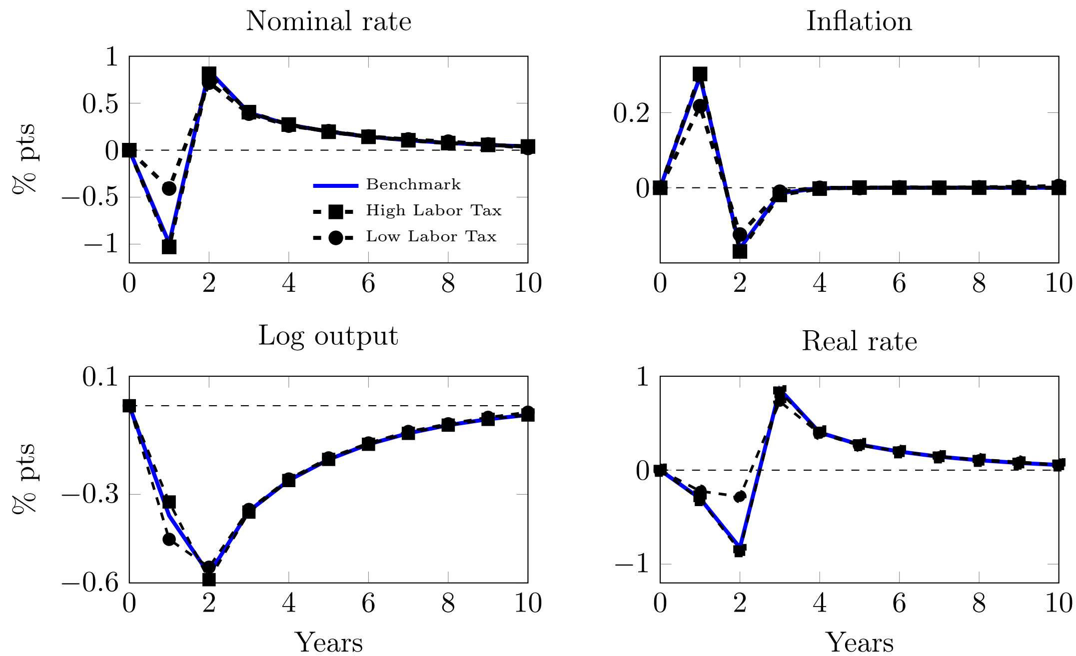

Figure 9: Optimal monetary response to a TFP shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares and circles are responses under a calibration with higher and lower labor taxes respectively.

see that the response of inflation and other variables is stronger when the dividend tax is low and weaker when it is high when we look at markup shocks. Since productivity shocks affect wages and dividends symmetrically, we should expect that the responses are not very different across cases that vary in the level of tax on dividend income. This prior is confirmed in figure 11.

On the contrary, a bond tax directly controls the dispersion in after-tax bond income which is key statistic for insurance against a productivity shock. In figures 12 and 13, we see that a higher bond tax lowers the response to the productivity shock and leaves the response to markup shock barely unchanged. To make our plots comparable with the baseline case we report impulse responses to the after-tax nominal and real interest rates.

Finally, we study the utilitarian Planner who sets Pareto weights equal. In our setup a utilitarian planner would set labor tax rate of 68%, a dividend tax rate of 116% and a bond tax rate of 118%. These effects go in offsetting directions but overall we find little deviation from the baseline responses. The results are summarized in figures 14 and 15.

D.3 Sensitivity with respect to the period of the shock

In this section, we compare optimal responses to a shock that occurs at $t = 25$ as well as $t = 50$ with our baseline in which the shocks occur at $t = 1$. For brevity we only report the

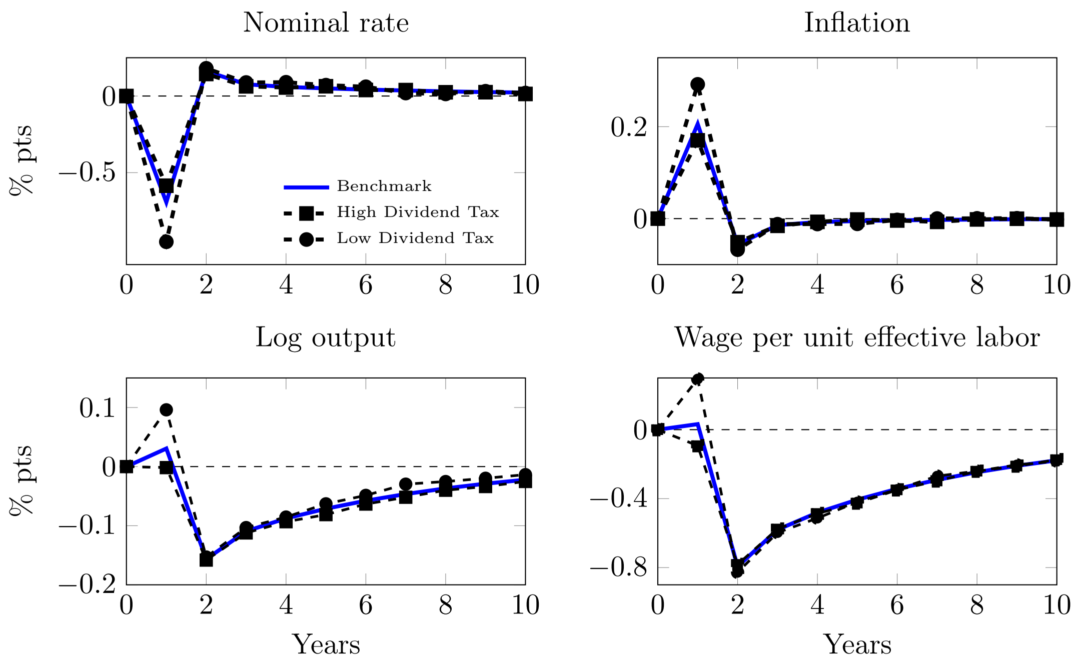

Figure 10: Optimal monetary response to a markup shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares and circles are responses under a calibration with higher and lower dividend taxes respectively.

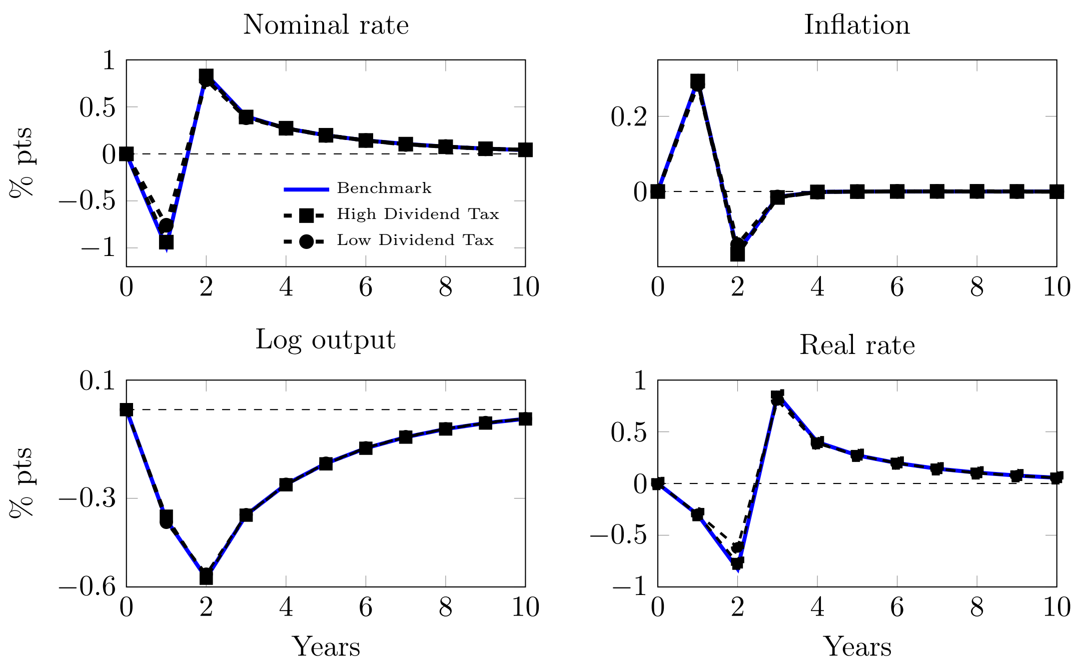

Figure 11: Optimal monetary response to a TFP shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares and circles are responses under a calibration with higher and lower dividend taxes respectively.

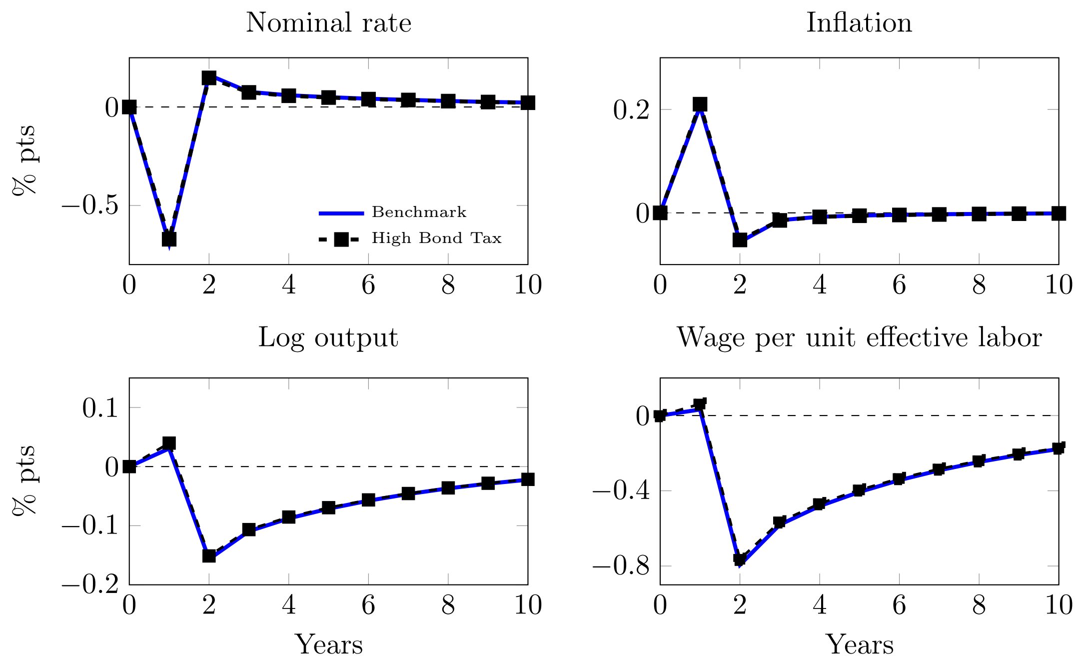

Figure 12: Optimal monetary response to a markup shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares are responses under a calibration with higher bond taxes.

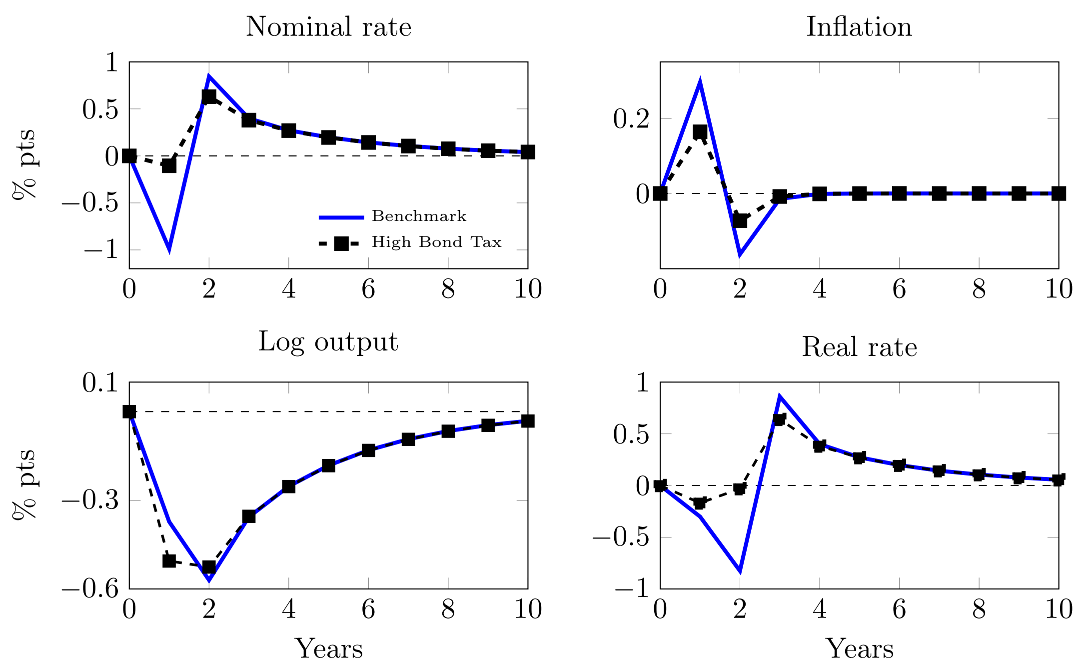

Figure 13: Optimal monetary response to a TFP shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares are responses under a calibration with higher bond taxes.

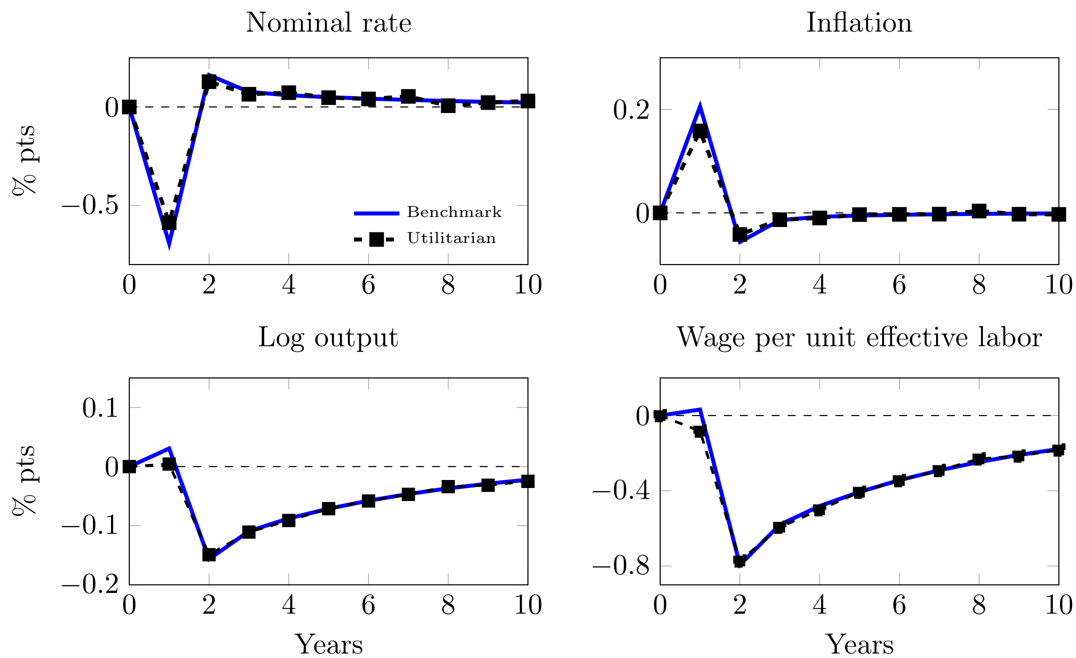

Figure 14: Optimal monetary response to a markup shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares are responses under utilitarian Pareto weights.

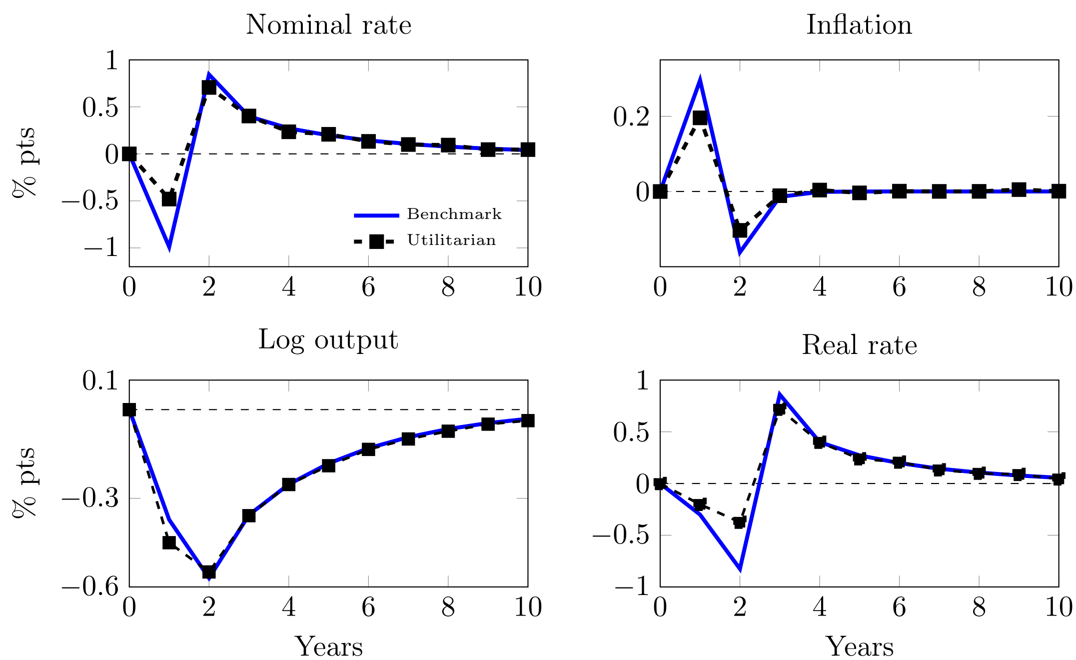

Figure 15: Optimal monetary response to a TFP shock. The bold blue lines are the responses for the baseline calibration. The dashed black lines with squares are responses under utilitarian Pareto weights.

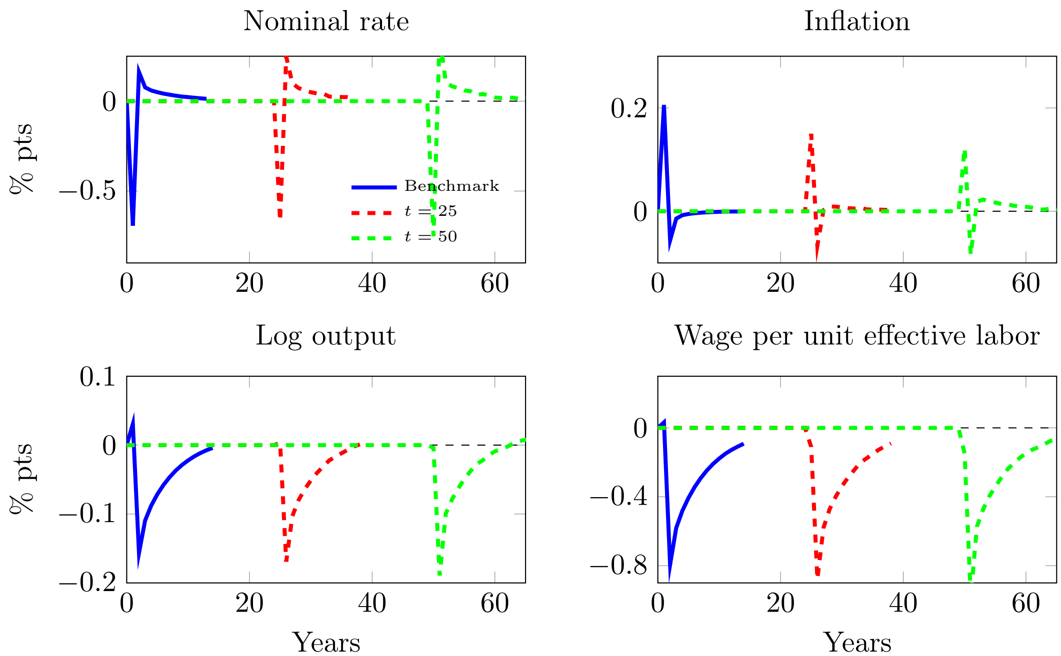

Figure 16: Optimal monetary response to a markup shock. The bold blue lines are the responses for the baseline calibration. The dashed lines are from the policies after initializing the the Ramsey allocation with $t = 1, 25, 50$ years of the competitive equilibrium.

optimal monetary response. Figure 16 plots the response to a markup shock, and figure 17 plots a response to a TFP shock. We find the responses to be very similar. The response to the TFP shock are slightly larger with time because the distribution of risk-free debt spreads out with idiosyncratic shocks and therefore there is a larger role for providing insurance.

D.4 Sensitivity with respect to choice of initial conditions

In the main text, we set the initial distribution of productivities, risk-free nominal bonds claims, and equity claims using the observed SCF distribution. Here we redo the optimal policy starting at a joint distribution of wealth and productivities that arises after simulating 100 years in the calibrated competitive equilibrium with fixed policies. The results are summarized in figures 18 and 19. The response to a markup shock is a balance of two forces. On the one hand the passage of time diminishes the correlation between stock holdings and labor earnings which renders inequality more misaligned according to our distance measure. That increases the planner's gains from providing insurance. On the other hand the correlation of shares of equities and bond holdings diminishes which diminishes the insurance gains from the unanticipated inflation. The increase in the spread of nominal debt leads the planner to be more responsive to a TFP shock.

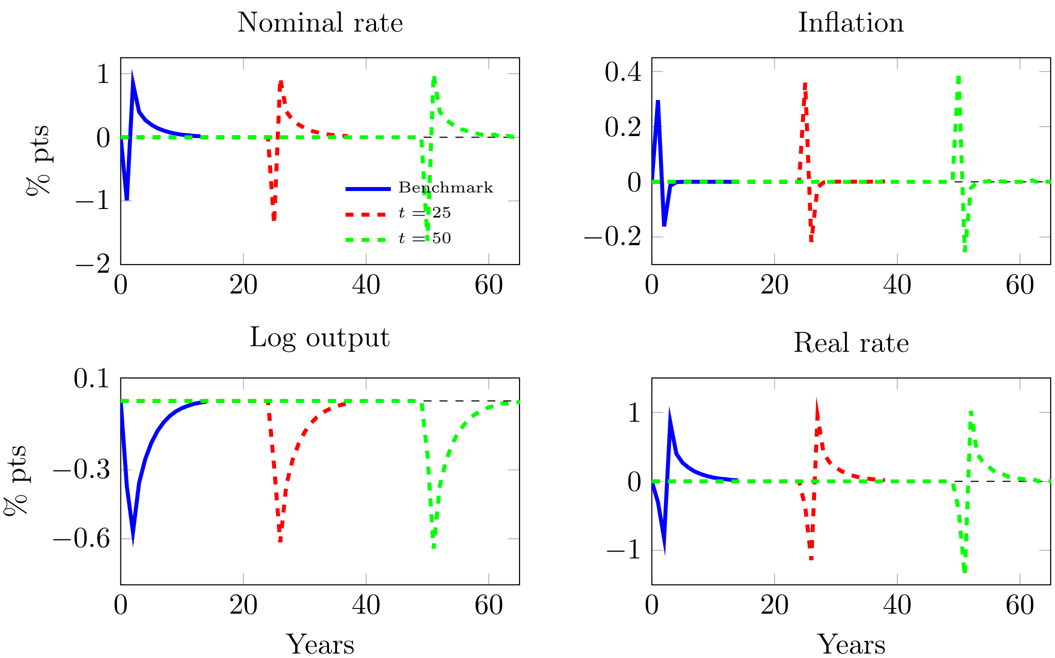

Figure 17: Optimal monetary response to a TFP shock. The bold blue lines are the responses for the baseline calibration. The dashed lines are from the policies after initializing the the Ramsey allocation with $t = 1, 25, 50$ years of the competitive equilibrium.

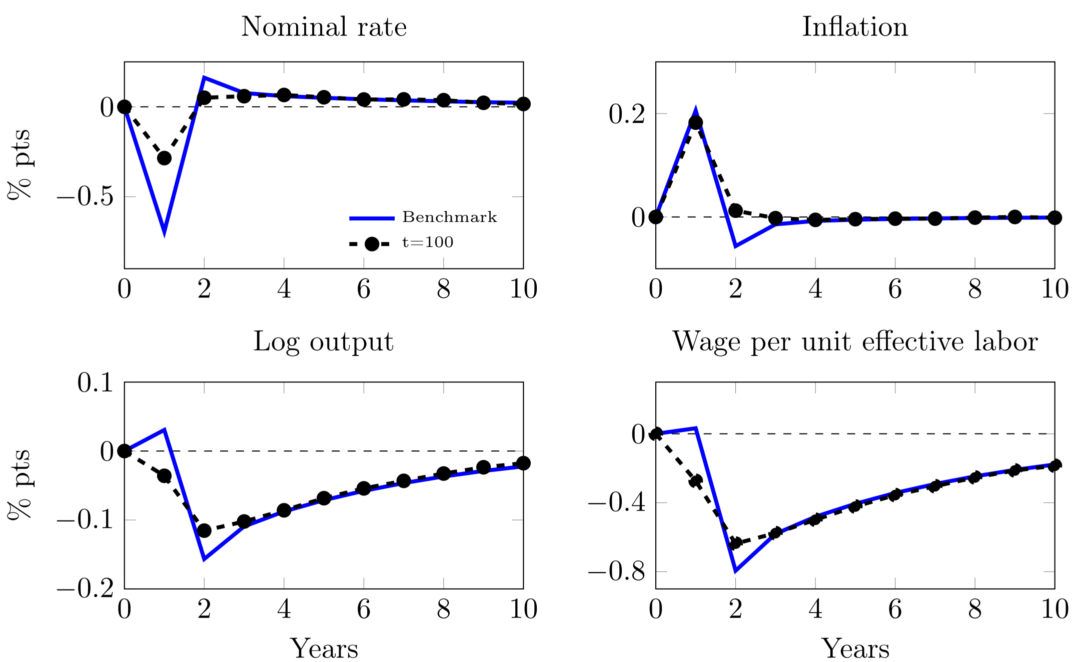

Figure 18: Optimal monetary response to a markup shock. The bold blue lines are the responses for the baseline calibration. The dashed lines are from the policies after initializing the the Ramsey allocation with $t = 100$ years of the competitive equilibrium.

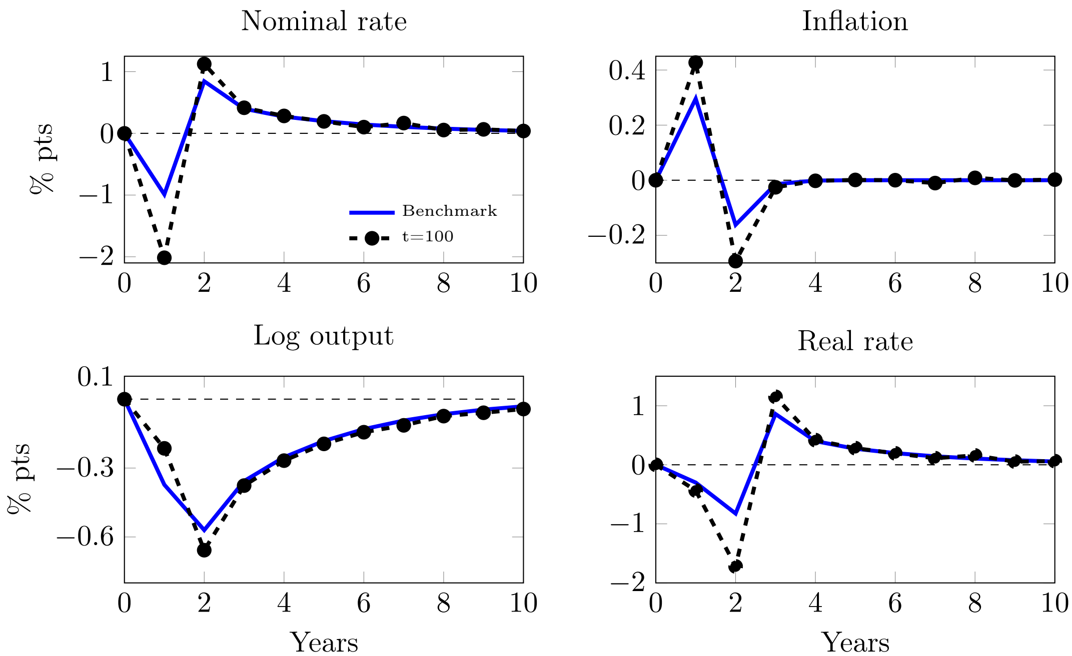

Figure 19: Optimal monetary response to a TFP shock. The bold blue lines are the responses for the baseline calibration. The dashed lines are from the policies after initializing the the Ramsey allocation with $t = 100$ years of the competitive equilibrium.

D.5 Example with poor hand-to-mouth agents

We can also consider an alternative calibration of the hand-to-mouth agents were we restrict bottom 15% of the cash-in-hand distribution to be hand-to-mouth. This environment is similar in spirit to what would arise in an standard Aiyagari model as the new hand to mouth agents more homogeneous and are almost entirely reliant on labor income. We plot the optimal policy response with only poor hand-to-mouth agents using the dashed red line in figure 20. As opposed to the hand-to-mouth setting in the main text that is calibrated to the evidence in Jappelli and Pistaferri (2014), the optimal policy with only poor hand to mouth agents is almost identical to that of the baseline economy as the government can construct a transfer scheme to smooth the consumption of the hand to mouth agents by mirroring the path of wages.

D.6 Heterogeneous marginal propensity to consume from dividend income and wage income

In this section, we study optimal monetary responses in a variant in which liquidity constrained agents can smooth dividend income. This results in a lower marginal propensity to consume out of income from capital income relative to income from labor. To model this,

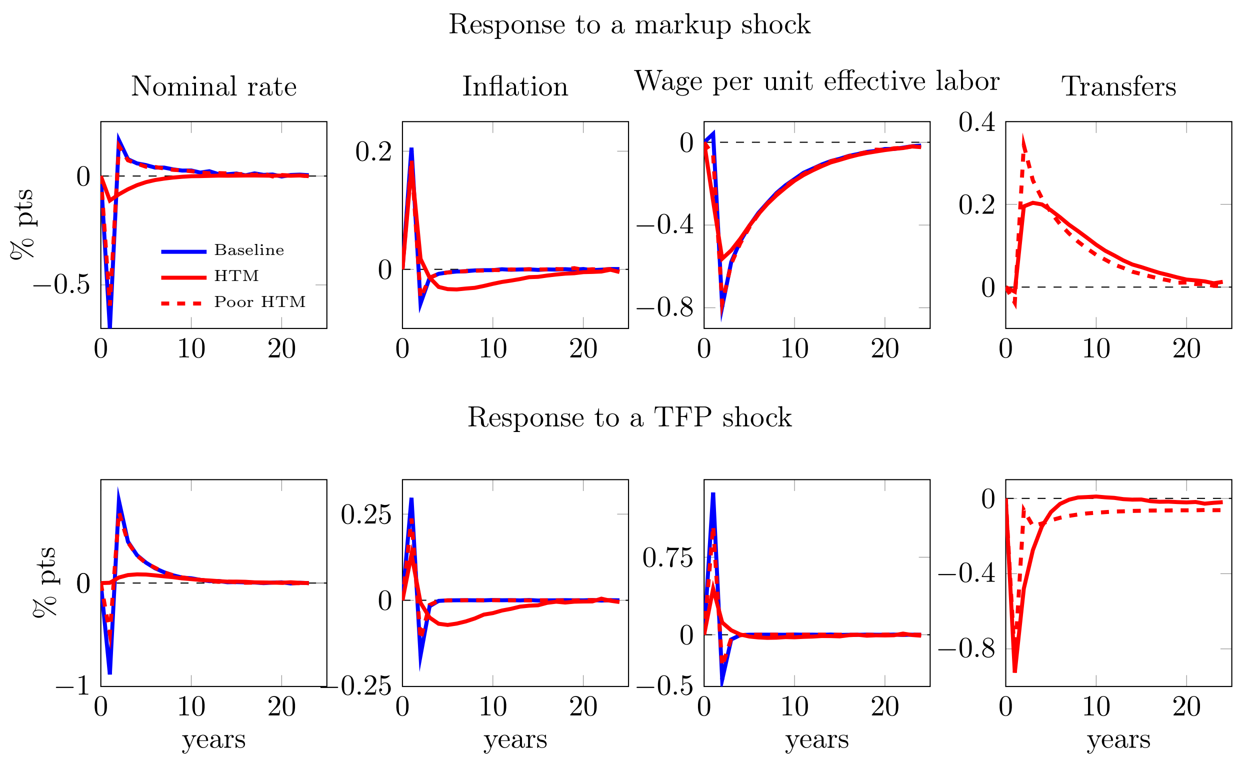

Figure 20: Optimal monetary responses with hand-to-mouth agents. The top panel plots responses to a markup shock and the bottom panel plots responses to a productivity shocks.

we change the savings rule for agents with the $h_i = 1$ from $P_t Q_t b_{i,t} = P_0 Q_0 b_{i,0}$ to

$$P_t Q_t b_{i,t} = P_0 Q_0 b_{i,0} + s_i P_t \tilde{D}_t$$

$$\tilde{D}_t = \tilde{D}_{t-1} + (1 - \text{divMPC}) \times \left(D_t - \overline{D}_t\right)$$

where $\bar{D}_t$ is the long run dividend level. The state variable $\tilde{D}_t$ is similar to holdings of mutual fund in which households save the fluctuations in their dividend income and are paid at a risk-free on the balance in return. For the rest of the section, we set divMPC=0. In figure 21 and 22, we plot optimal monetary responses to the markup and the productivity shock, respectively.

As we describe in the main text in section 6.2, heterogeneity in marginal propensity to consume across households makes the path of the optimal interest rate smoother. Manipulating the timing of lump sum transfers is not sufficient to insure the consumption path of the constrained agents who differ in their holdings on stocks and bonds. The planner uses monetary policy to directly smooth real returns and wages. While implementing such smoothing, there is a tension: smoothing the wage and dividend share that helps liquidity constrained stockholders requires movements in natural rates that hurt liquidity constrained bond holders. Allowing for the ability to additionally smooth dividends relaxes this tension and the results in paths of nominal rates that are even more smooth. Quantitatively, this effect is larger for markup shocks than for productivity shocks.

D.7 Optimal monetary-fiscal response with mutual fund

In this section, we present optimal monetary response to productivity shock, as well as the optimal monetary-fiscal response to both under the mutual fund calibration. The optimal monetary response to the productivity shock is in figure 23.

One aspect of the mutual fund calibration is that it enforces a perfect correlation between bond and dividend wealth following any history of shocks. As a result, the optimal policy the bond and dividend tax rates are indeterminate as the planner can achieve the same effective returns with either instrument. To make the results comparable with our benchmark calibration, we assume that the planner adjusts the dividend tax in response to a markup shock and the bond rate in response to a markup shock. The results are plotted in figures 24 and 25. In both cases the optimal policy under the mutual fund is almost identical to the benchmark calibration.

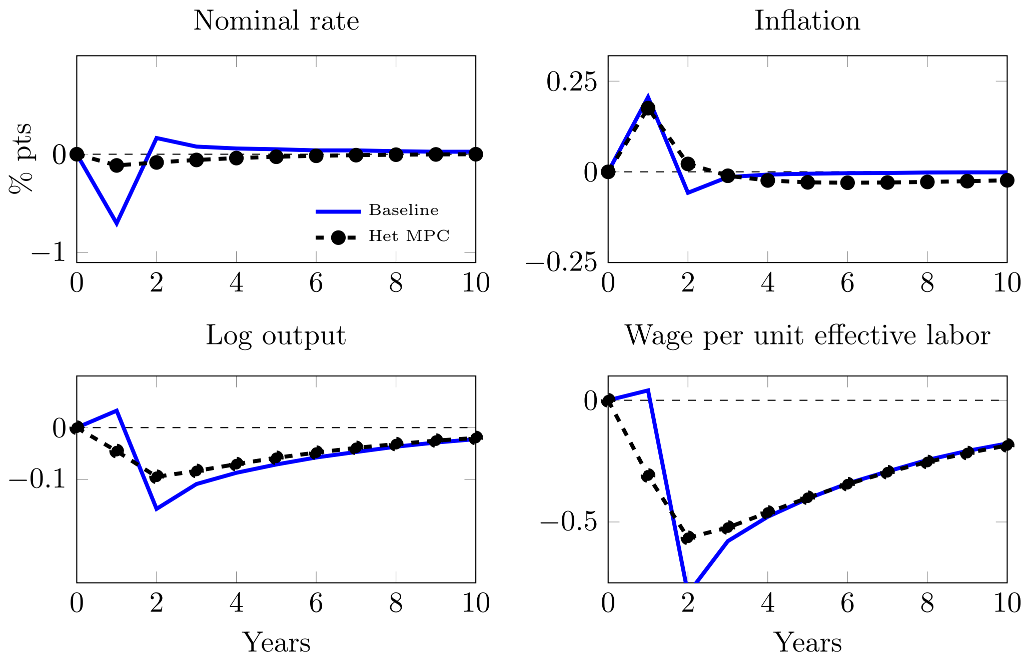

Figure 21: Optimal monetary responses to the markup shock. The bold blue lines are responses under the baseline and the dashed black lines with circles are responses with heterogeneous marginal propensities to consume out of dividend and labor incomes.

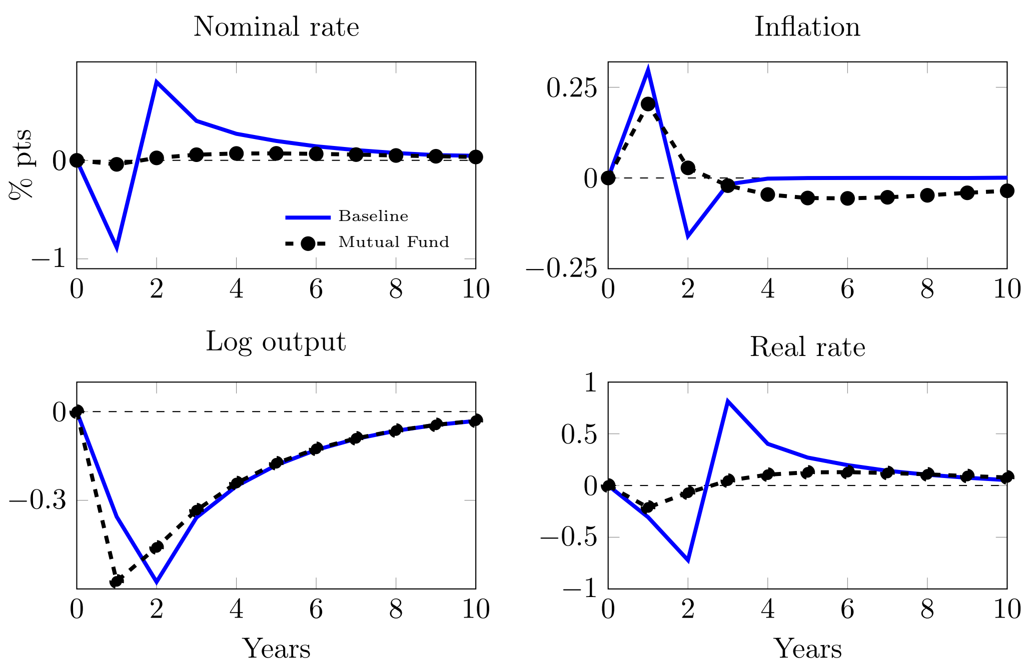

Figure 22: Optimal monetary responses to productivity shock. The bold blue lines are responses under the baseline and the dashed black lines with circles are responses with heterogeneous marginal propensities to consume out of dividend and labor incomes.

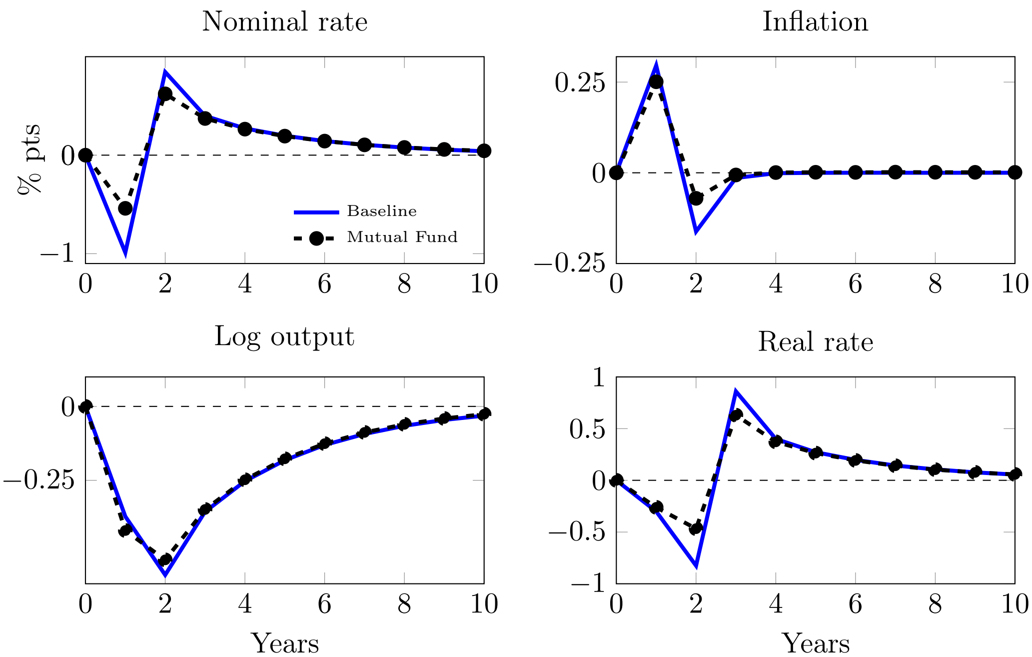

Figure 23: Optimal monetary responses to productivity shock. The bold blue lines are responses under the baseline and the dashed black lines with circles are responses under the mutual fund setting.

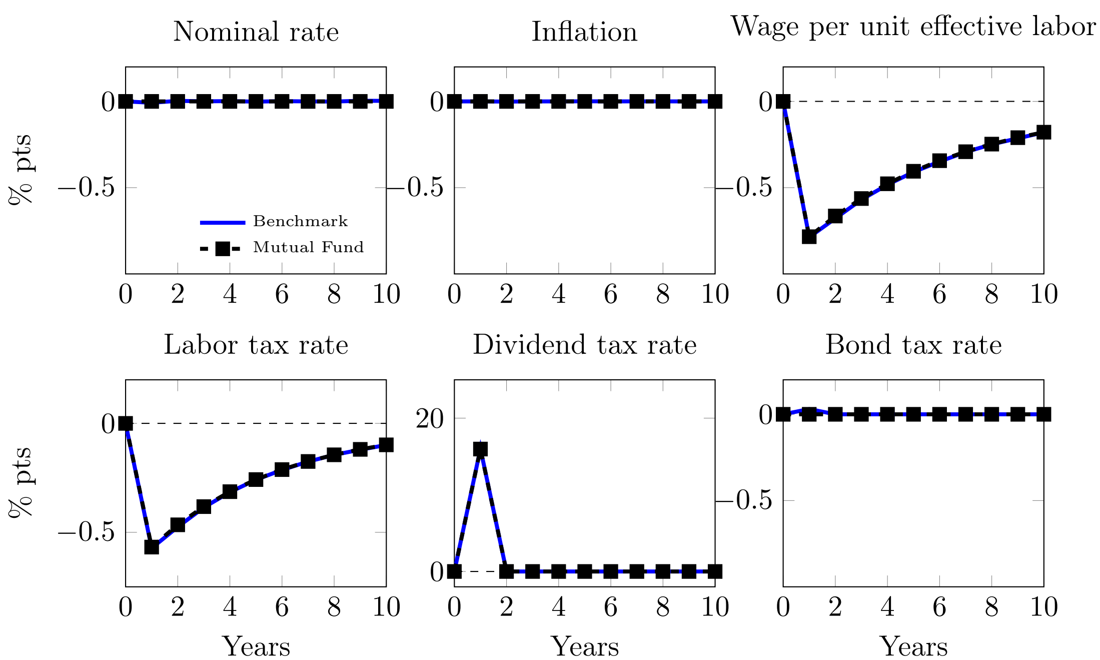

Figure 24: Optimal monetary-fiscal response to a markup shock. The bold red are the benchmark response while the bold blue lines are the responses for the mutual fund calibration.

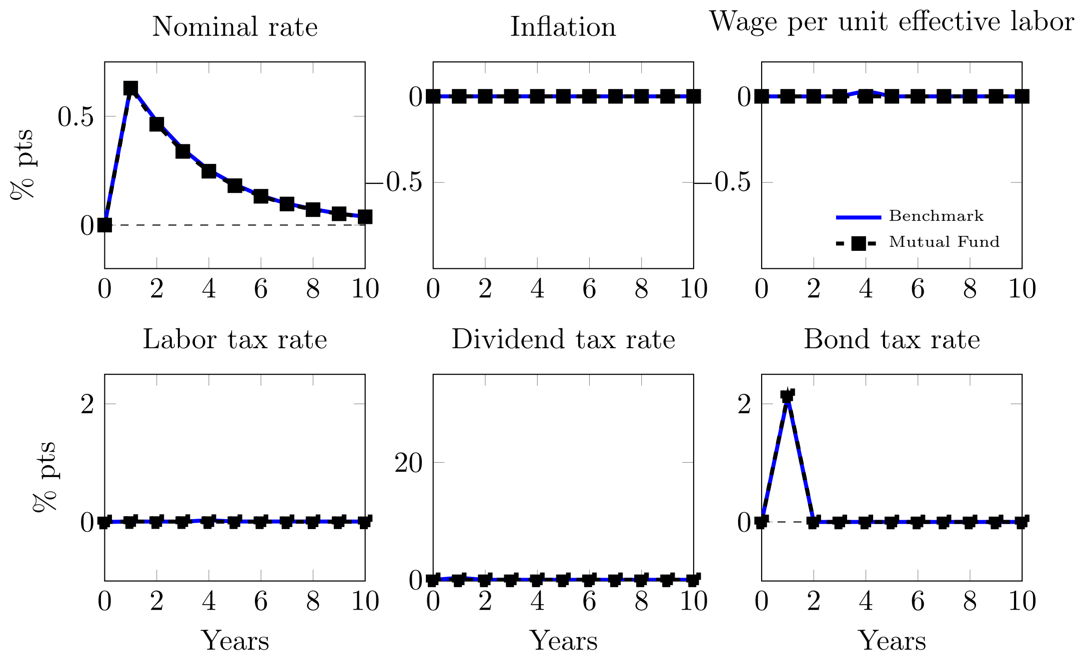

Figure 25: Optimal monetary-fiscal response to a TFP shock. The bold blue lines are the response under the baseline while the dashed black lines are the responses for the mutual fund calibration.

D.8 Optimal monetary and monetary-fiscal response to a TFP shock with heterogeneous labor income exposures

As noted in section 6.4 we calibrate the coefficients of $f(\theta) = f_0 + f_1\theta + f_2\theta^2$ by simulating the competitive equilibrium for 30 periods and extracting "recessions" as consecutive periods where the growth rate of output one standard deviation below zero. Following the empirical procedure in Guvenen et al. (2014), we rank workers by percentiles of their average log labor earnings 5 years prior to the shock and compute the percent earnings loss for each percentile relative to the median. The parameters f_1, f_2 are set to match earnings losses of the 5th, 95th percentiles. The parameter f_0 is set so that agent with the median productivity faces a drop similar to the aggregate TFP. Figure 26 plots the earnings losses by percentile of the income distribution relative to those found by Guvenen et al. (2014).

Figure 27 plots the responses of the monetary-fiscal policy. When the government has access to fiscal policy, it no longer needs to rely solely on monetary policy. In figure 27 we see that in response to an inequality shock the planner raises the labor tax rate by nearly 1% and then allows it to mean revert back as the TFP shock dissipates. This mean reversion arises because the level of inequality loads on TFP and partly captures the forces laid out in Werning (2007) where the planner responds to changes in relative labor productivity though changes in the labor tax rate. Unlike in baseline, case when nearly all insurance can be

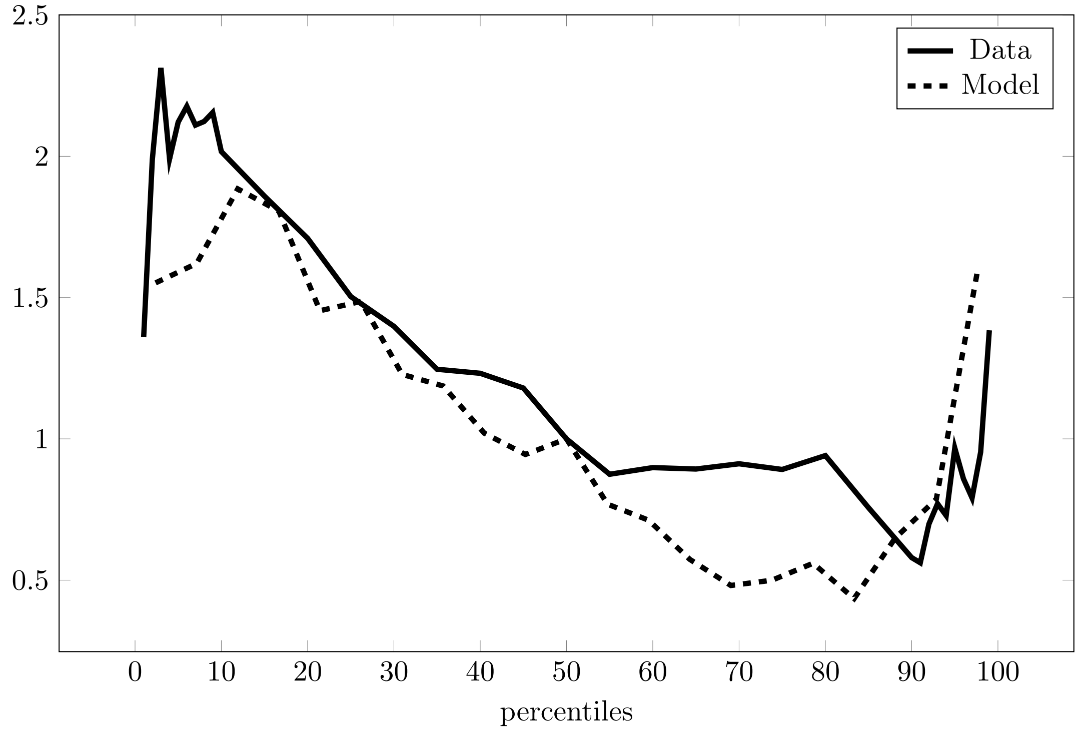

Figure 26: Relative income losses after recessions in data (solid line) and model (dashed line).

provided through a surprise tax on bond income the planner must also rely on a surprise increase in the dividend tax rate to partially provide insurance. This highlights a feature of heterogeneous agent models. Unlike representative agent models where a single tax on returns can complete markets, with heterogeneous agents one tax may not provide insurance for all agents and the planner may exploit multiple different asset taxes.

In figure 28, we apply our decomposition to the monetary response with setting with heterogeneous exposures. The small difference in the rate of inflation in the HANK complete market relative to RANK case captures the redistribution. In figure 27, we saw that labor income taxes are used to respond to inequality even with complete markets. When the planner cannot adjust labor income tax, wages and thereby inflation is used to attain similar objectives.

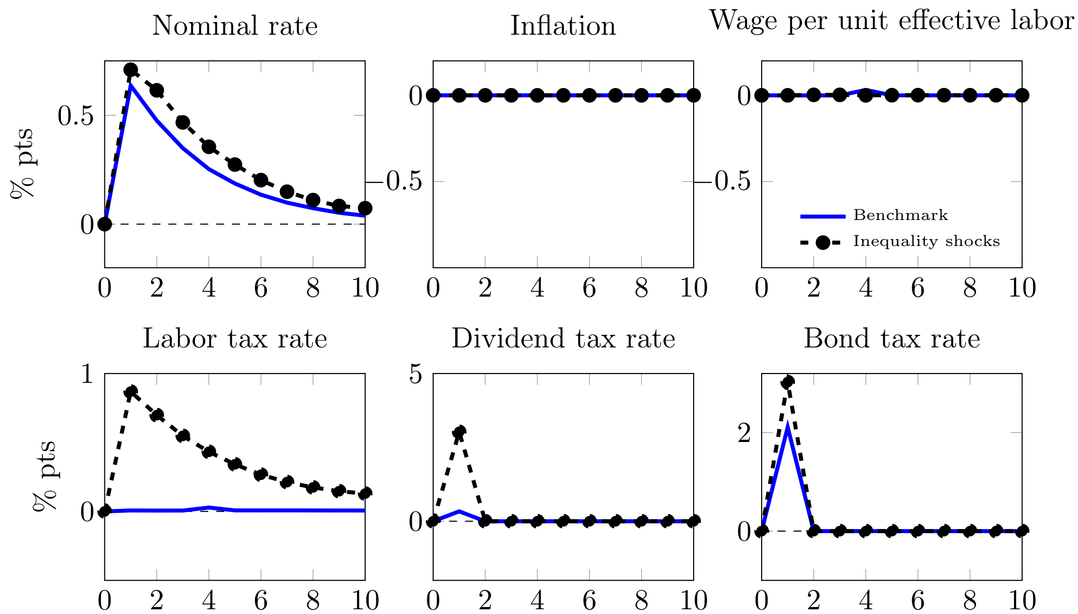

Figure 27: Optimal monetary-fiscal response to a TFP shock. The bold blue lines are the response under the baseline calibration while the dashed black lines are the responses calibration with heterogeneous exposures to TFP shocks

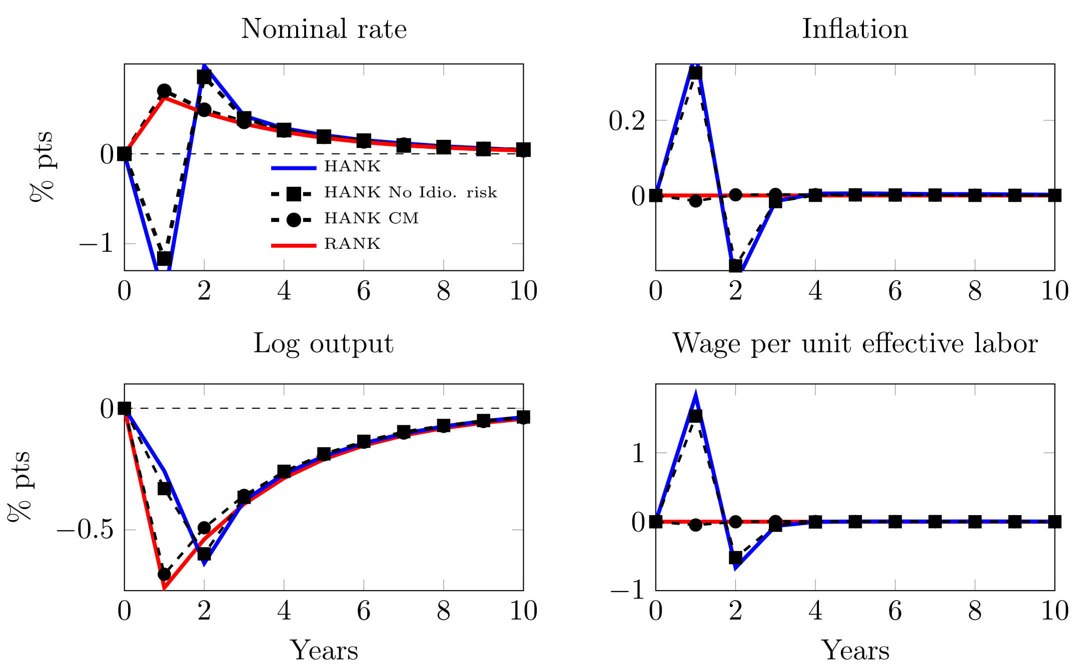

Figure 28: Decomposition of the optimal monetary response to a TFP shock with heterogeneous exposures.The bold blue and red lines are the calibrated HANK and RANK responses respectively. The dashed black lines with squares and circles are responses under HANK with idiosyncratic shocks shutdown and with complete markets, respectively.