How Small is Small? Non-linearities in Heterogeneous Agent Models *

Javier Bianchi[†] Greg Kaplan[‡]

July 7, 2026

Abstract

In plausibly calibrated heterogeneous-agent models, marginal propensities to consume (MPCs) are highly non-linear in wealth, falling sharply away from borrowing constraints. As a result, the aggregate consumption response to a fiscal transfer is strongly concave in its size: larger transfers shift households out of high-MPC regions, dampening the consumption response. Across partial- and general-equilibrium settings, linear methods substantially overstate the effects of fiscal stimulus at empirically relevant sizes. Local methods of any order are unlikely to be reliable in settings where a failure of Ricardian equivalence from high MPCs is important.

Keywords: Heterogeneous agents, non-linearities, HANK

JEL classification: D15, D31, D52, E21, E62, G51, C63

*We thank Manuel Amador, Alisdair McKay and Ben Moll for helpful conversations and Xing Xu for excellent research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

[†]Federal Reserve Bank of Minneapolis, javier.i.bianchi@gmail.com.

[‡]University of Chicago, e61 Institute and NBER, gkaplan@uchicago.edu.

1 Introduction

There has been an explosion of work studying the effects of fiscal and monetary interventions in settings with heterogeneous agents (HA) and incomplete markets. These frameworks are useful because they do a good job of capturing relevant features of household-level spending and savings behavior. In particular, suitably calibrated versions can match both the level and distribution of marginal propensities to consume (MPC).

These aspects of consumption behavior are especially relevant for studying fiscal interventions, because a failure of Ricardian equivalence lies at the heart of the aggregate and distributional effects of such policies. The size and pattern of MPCs underscore these departures from Ricardian equivalence and drive much of the difference between heterogeneous-agent and representative-agent models. Indeed, the effects of fiscal stimulus programs funded by future taxes are driven entirely by such departures.

A large part of the recent literature has analyzed these settings using various forms of first-order methods.[1] The most common approach, labeled "sequence-space Jacobian" (SSJ) by Auclert, Bardóczy, Rognlie and Straub (2021), constructs an aggregate consumption function by aggregating household consumption policies, and computes equilibrium transition paths using a linear approximation of this mapping in a suitably chosen vector of aggregate prices and policy variables, evaluated around the steady-state.[2] It is well understood that such methods approximate non-linear equilibria well when the policy innovations and exogenous disturbances being studied are small. But how small is small?

We argue that for commonly studied fiscal interventions, small is indeed very small—much smaller than many of the aggregate shocks and interventions that macroeconomists are motivated to explore based on real-world experience.

Our starting point is the observation that for a funded fiscal stimulus program, the bulk of the aggregate effects are driven by the strength of departures from Ricardian equivalence, which in turn are driven by the extent to which households' MPCs exceed those implied by the permanent income hypothesis (PIH).

[1] Examples include Aggarwal, Auclert, Rognlie and Straub (2023), Bardoczy, Sim and Tischbirek (2024), Rachel and Summers (2019), Hänsel (2024), Campos, Fernández-Villaverde, Nuño and Paz (2024), Eichenbaum, Guerreiro and Obradovic (2025); Angeletos, Lian and Wolf (2024a,b). Studies that allow for non-linear effects of fiscal policy in incomplete markets are Hagedorn, Manovskii and Mitman (2019); Brinca, Faria-e Castro, Ferreira, Holter and Nóbrega (2025); Mongey and Waugh (2025). Relative to this work we systematically evaluate the degree of non-linearity in a canonical HANK model and assess the accuracy of linearization and higher-order local methods.

[2] See also Boppart, Krusell and Mitman (2018) for a related approach in the sequence space using finite-differences, and Reiter (2009) for a projection-based method.

In plausibly calibrated heterogeneous-agent models, MPCs decline sharply with wealth and may fall discontinuously when a borrowing constraint ceases to bind. Larger transfers therefore move households out of high-MPC regions, making the aggregate consumption response highly concave in the size of the transfer. As a result, linear methods can substantially overstate the effects of fiscal stimulus at empirically relevant transfer sizes.

Higher-order local approximations do not remedy the inaccuracy of the first-order approximation. The main source of non-linearity is not the smooth curvature in the consumption function, but rather the movement of households across the kink induced by the borrowing constraint. A second-order approximation accounts for this movement and improves the quality of the approximation for very small transfers. However, for larger empirically relevant transfer amounts, we find that second- and third-order approximations can perform worse than the linear approximation because they assume that the share of constrained households continues to decline with transfer size, eventually becoming negative, and thus substantially understate the response relative to the non-linear solution.

These features lead us to conjecture that inferring the aggregate effects of a fiscal intervention on the basis of the aggregate consumption response to very small disturbances may be misleading. In the remainder of the paper, we investigate when this is and is not the case. Overall, our findings do not bode well for using first-order methods, such as SSJ, to study the effects of fiscal interventions.

We start in Section 2 with a partial equilibrium exercise that highlights the key mechanism behind the non-linearity and illustrates its quantitative significance. In Section 3, we then analyze various general equilibrium models with different assumptions about production, nominal rigidities, fiscal financing and monetary policy. In Section 4 we provide examples of shocks and other policy interventions where linear solutions approximate non-linear solutions much better. Our simulations suggest that the non-linearities we emphasize are significant when a failure of Ricardian equivalence is an important component of the transmission mechanism. Consequently, those shocks and policies for which local methods are accurate in HA models tend to be those where aggregate responses are similar in RA and HA economies.

2 Partial Equilibrium

2.1 Model

Demographics Time is discrete and infinite $t = 0, 1, \ldots$. The economy is populated by a measure-one continuum of households that face uninsurable idiosyncratic shocks. We assume the path for aggregate disturbances is known at $t = 0$, so there is no aggregate uncertainty.

Preferences Households have utility over consumption, where $u(c_{it})$ is strictly increasing and concave, and discount the future with factor β; discount-factor heterogeneity is introduced in the calibration.

Income and assets Each period households receive earnings y_{it} whose realizations are independent across households. We describe the stochastic process for earnings below. Households pay a proportional tax at rate τ on their earnings and receive a lump-sum transfer T_t from the government. They can save and borrow in a one-period risk-free real asset which pays interest rate r.

Household problem Households enter period t with assets $a_{i,t-1}$ and choose next-period assets $a_{i,t}$. They face the borrowing constraint $a_{i,t} \geq 0$. The household problem is

$$\max_{\{c_{i,t},a_{i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{i,t})$$

subject to

$$c_{i,t} + a_{i,t} \leq (1+r)a_{i,t-1} + (1-\tau)y_{i,t} + T_t \quad \text{and} \quad a_{i,t} \geq 0.$$

The first-order condition of this problem yields the intertemporal Euler equation:

$$u'(c_{i,t}) \geq \beta(1+r)\mathbb{E}_t u'(c_{i,t+1})$$

with equality if $a_{i,t} > 0$.

Aggregation The solution to the household problem yields a consumption policy that we express as a function of cash on hand, $x \equiv (1+r)a_{i,t-1} + (1-\tau)y_{i,t} + T_t$, and the income state y, namely $c_t(x, y)$. Given an initial distribution $\Gamma_0(x, y)$ and a transfer path $\{T_s\}_{s\geq 0}$, the policy induces a sequence of distributions $\{\Gamma_t(x, y)\}_{t\geq 0}$, and total consumption in period t is

$$\mathcal{C}_t(\{T_s\}) = \int c_t(x, y)\, \mathrm{d}\Gamma_t(x, y).$$

2.2 Steady-state calibration

Demographics A period represents a quarter. The numeraire in the model is quarterly GDP per household which was around \$166,000/4 = \$41,500 in 2019.

Income and assets We set the interest rate to 2% p.a. and the borrowing limit to $\underline{a} = 0$. We model the process for log earnings $\log y_{it}$ as the sum of two orthogonal components, an AR(1) component and an IID component. Following Kaplan and Violante (2022), we assume that shocks to both components arrive stochastically with a Poisson arrival rate of 1/4, so that shocks are received on average once a year. We set the parameters of the earnings process to match moments of the household labor income distribution from the Panel Study of Income Dynamics. See Appendix A for details.

Preferences We assume $u(c) = \log(c)$. We assume the discount factor β_i is heterogeneous across households and can take five equally spaced values, each with equal probability. We calibrate the mean discount factor so that the ratio of mean wealth to annual GDP is 100% (\$166,000). We calibrate the dispersion in the discount factors so that the average quarterly MPC out of a one-time \$1,000 windfall is 25%. The implied range of discount factors is between 0.947 and 0.993.

Tax and transfers The tax system consists of a proportional tax rate and lump-sum transfer, whose values in steady state we denote by τ and T, respectively. Net earnings are given by $y_{it} - \mathcal{T}(y_{it}) = (1-\tau)y_{it} + T$. We set $\tau = 0.274$ to match total government revenue of 27.4% of GDP in 2019. We set the steady-state lump-sum transfer $T = 15\%$ of GDP, which is around \$25,000 per household per year.

Wealth and MPC distributions Figure 1a shows the stationary wealth distribution overlaid with the MPC as a function of assets, for households with different income levels.[3] Figure 1b shows a histogram of these MPCs. Two features of MPCs in this class of models that drive our aggregate findings are apparent. First, the average MPC is declining in wealth and approaches the PIH MPC as assets rise. This decline reflects two forces: (i) at each income level the MPC is weakly declining in wealth; and (ii) very low-wealth households are disproportionately those with low income, who are borrowing constrained at low asset

[3] We plot the MPC out of an infinitesimal windfall because this is the MPC that is relevant for the linear vs non-linear comparisons that are the focus of this paper. We calibrate to the MPC out of a \$1,000 windfall because that is more consistent with the empirical evidence we match to.

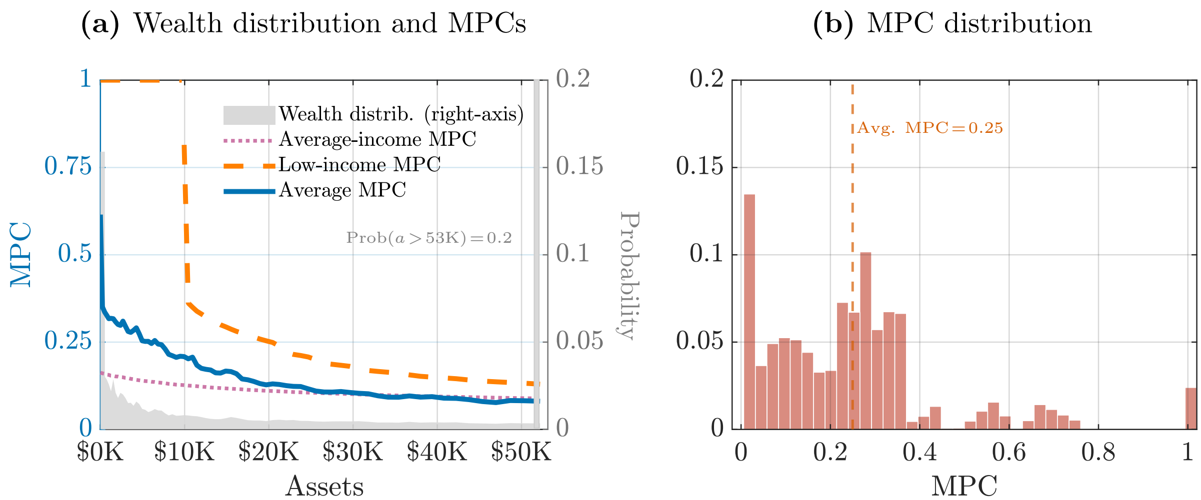


Figure 1: Steady-state wealth distribution and MPCs

Note: The left axis of panel (a) presents the MPC for average-income and low-income households both for the lowest β_i and the average MPC across households, each as a function of assets. The right axis of panel (a) presents the wealth distribution up to \$50K. Panel (b) presents the distribution of MPCs.

levels and have an MPC of 100%. As wealth increases, the mix shifts towards high-income households who have much lower MPCs.

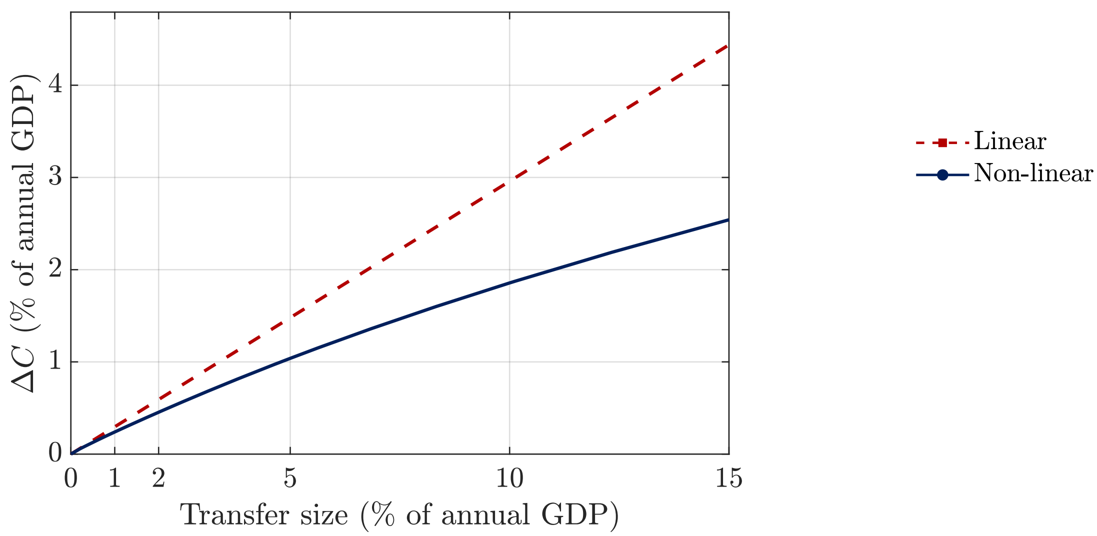
Second, the distribution of MPCs is highly dispersed and approximately bi-modal. A subset of households with both low assets and low income have very high MPCs, close to 100%. For these households, the MPC function is flat at 100% in a small region above the borrowing constraint, after which it drops discontinuously. The next section explores the implications of these patterns for the shape of the aggregate consumption function.

2.3 Aggregate Transfers in Partial Equilibrium

Experiment Starting from the steady-state distribution, we study an unexpected one-time increase in the lump-sum transfer to all households at $t = 0$. We implement this by setting the transfer to $T_0 = T + \Delta$ and keeping $T_t = T$ for $t > 0$. Since the only change in the tax-and-transfer schedule is at $t = 0$, we denote aggregate consumption at time t by $\mathcal{C}_t(\Delta)$. Our interest is the aggregate consumption response, $\mathcal{C}_t(\Delta) - \mathcal{C}_t(0)$.

Computation We compute steady-state decision rules using the method of endogenous grid points. Since we are studying a one-time transfer at $t = 0$ in partial equilibrium, we can compute the non-linear response to the transfer shock using the steady-state decision rules

(a) Aggregate effects of a one-time transfer in partial equilibrium



(b) Linear to non-linear aggregate consumption response

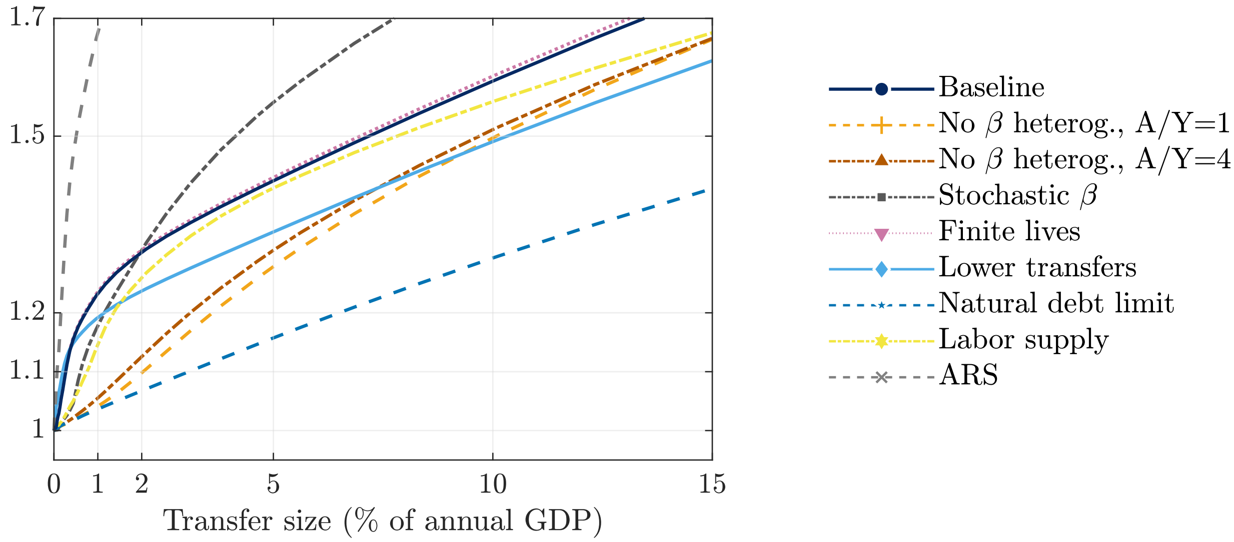


Figure 2: Partial equilibrium effects of a one-time transfer

Notes: Panel (a): ΔC is expressed as a percentage of annual GDP. Panel (b) presents the ratio of the linear to the non-linear consumption response for different model calibrations. See Table A.3 in Appendix A.2 for more details.

and propagating forward the household wealth distribution.

We compute the linear response using a finite difference in Δ to approximate the derivative of the aggregate consumption function $\frac{\partial \mathcal{C}_t(\Delta)}{\partial \Delta}$. In partial equilibrium this is equivalent to how the transfer is treated in the SSJ method that we use for general equilibrium experiments in Section 3. We have also verified that computing $\frac{\partial \mathcal{C}_t(\Delta)}{\partial \Delta}$ using automatic differentiation yields the same result as our central finite-difference approximation, as well as the analytical individual MPC in (B.6), aggregated as in the expression for $\mathcal{C}'(0)$ in (2), with differences of order 10^{-7}. See Appendix D for details of computation.

Linear vs. non-linear responses Figure 2 contains the main results of the partial equilibrium exercise. Figure 2a shows the first-quarter increase in aggregate consumption $\mathcal{C}_0(\Delta) - \mathcal{C}_0(0)$, expressed as a percentage of annual GDP per household, as a function of the size of the transfer Δ. The blue solid line labeled "Non-Linear" shows the results computed using the non-linear solution and the red dashed line labeled "Linear" shows the results from the linear solution. The aggregate consumption response is concave in the size of the transfer. For very small transfers, the linear and non-linear solutions produce similar aggregate consumption responses. But for larger transfers, the linear solution substantially overstates the consumption response. For example, for a transfer that is 5% of annual household GDP (around \$8,300), the linear solution overstates the non-linear solution by around 42%. Even for a transfer of 0.5% of annual GDP (around \$830), the linear solution overstates the non-linear solution by more than 17%.

Alternative calibrations Figure 2b shows that the concavity of $\mathcal{C}_0(\Delta)$ is a robust feature of this class of models and is not specific to our particular calibration and choice of model features. The figure shows the ratio of the linear consumption response to the non-linear response for transfers of different sizes, for a range of alternative calibrations and versions of the model. We report the results in terms of the ratio of the linear to non-linear solutions because the size of the response differs across models.

The versions of the model included in the figure include those without discount factor heterogeneity, with stochastic discount factors, with different levels of aggregate wealth, a life-cycle version with finite lives, with lower lump-sum transfers, with a natural borrowing limit, with endogenous labor supply, and with the idiosyncratic earnings process from Auclert et al. (2024). In all these versions, and many others we have experimented with, the linear solution overstates the aggregate consumption response by an economically meaningful magnitude for transfer sizes larger than 2% of GDP. In many cases, including our baseline, the errors appear at much smaller transfer sizes.

2.4 Inspecting concavity

The aggregate consumption response is concave for two reasons. The first is familiar: unconstrained household consumption policies are concave in cash on hand (e.g., (Carroll and Kimball, 1996)). The second source, central to our analysis, is that a transfer moves households off the borrowing constraint, replacing an MPC of one with a lower unconstrained MPC. To make this precise, we decompose the aggregate response.

Let $\tilde{x}$ denote baseline cash on hand before the one-time transfer, and write total current cash on hand under the transfer as

$$x = \tilde{x} + \Delta.$$

Let $G(\tilde{x} \mid y)$ denote the baseline distribution of cash on hand conditional on income y and let $g(\tilde{x} \mid y)$ its density. In addition, let $F(y)$ denote the income distribution, and let $\bar{x}(y)$ denote the cash-on-hand kink at which the borrowing constraint stops binding. For simplicity, we assume a common β across households.

Aggregate consumption under a one-time transfer Δ at $t = 0$ is

$$\mathcal{C}(\Delta) = \int_Y \left[\underbrace{\int_{-\infty}^{\bar{x}(y)-\Delta} c(\tilde{x}+\Delta, y)\, \mathrm{d}G(\tilde{x} \mid y)}_{\text{constrained after transfer: } \tilde{x}+\Delta \leq \bar{x}(y),\ \text{MPC}=1} + \underbrace{\int_{\bar{x}(y)-\Delta}^{\infty} c(\tilde{x}+\Delta, y)\, \mathrm{d}G(\tilde{x} \mid y)}_{\text{unconstrained after transfer: } \tilde{x}+\Delta > \bar{x}(y)} \right] \mathrm{d}F(y). \tag{1}$$

where we use that a household remains constrained after the transfer if and only if

$$\tilde{x} + \Delta \leq \bar{x}(y),$$

. Thus, as Δ rises, the mass of households below the post-transfer constraint cutoff falls. This is the movement-off-the-constraint channel.

Individual MPCs are given by

$$c_x(x, y) = \begin{cases} 1, & x < \bar{x}(y), \\ c_x^+(x, y), & x > \bar{x}(y), \end{cases}$$

where $c_x^+(x, y) < 1$ is the unconstrained MPC, derived in Appendix B in equation (B.6). The MPC therefore drops discontinuously at the kink: below $\bar{x}(y)$, constrained households consume one-for-one out of additional cash on hand; above $\bar{x}(y)$, their MPC falls to $c_x^+ < 1$.

This discontinuity is visible in the orange dashed low-income MPC line in Figure 1.[4]

Differentiating equation (1) at $\Delta = 0$ gives

$$\mathcal{C}'(0) = \underbrace{\int_Y G(\bar{x}(y) \mid y)\,\mathrm{d}F(y)}_{\text{constrained (MPC = 1)}} + \underbrace{\int_Y \int_{\bar{x}(y)}^{\infty} c_x^+(\tilde{x}, y)\,\mathrm{d}G(\tilde{x} \mid y)\,\mathrm{d}F(y)}_{\text{unconstrained}}\,. \tag{2}$$

The first term is the mass of households initially constrained. The second term is the weighted average MPC for unconstrained households. Note that because consumption is continuous at the kink, the effect of Δ in the boundary terms cancel.[5] Hence, movement across the kink does not affect the first-order aggregate MPC.

Consider now the second-order effects from transfers. We have

$$\mathcal{C}''(0) = \underbrace{-\int_Y \overbrace{\left(1 - c_x^+(\bar{x}(y), y)\right)}^{\text{MPC drop at the kink}} g(\bar{x}(y) \mid y)\,\mathrm{d}F(y)}_{\text{movement off constraint}} + \underbrace{\int_Y \int_{\bar{x}(y)}^{\infty} c_{xx}^+(\tilde{x}, y)\,\mathrm{d}G(\tilde{x} \mid y)\,\mathrm{d}F(y)}_{\text{individual-policy concavity}}\,. \tag{3}$$

The first term is the movement-off-constraint channel. A marginally larger transfer moves households located just below the kink $\bar{x}(y)$ across it, from the constrained region, where their MPC is one, into the unconstrained region, where their MPC is $c_x^+(\bar{x}(y), y) < 1$. Because the density is continuous, the mass exactly at the kink is zero; the term $g(\bar{x}(y) \mid y)$ measures the marginal mass of households near the kink that cross into the unconstrained region.

The second term is the familiar individual-policy concavity in incomplete market models. A higher transfer moves an unconstrained individual household toward a lower MPC. Quantitatively, the movement-off-constraint channel dominates: it accounts for about 93% of the gap between the linear and non-linear impact responses at a transfer of 1% of annual GDP, and about 82% at a 5% transfer (Figure F.2). [6]

[4] See Figure F.4 in Appendix F for the corresponding consumption function and Appendix B for a proof that c_x is discontinuous at the boundary of the linear segment.

[5] Note also that because the density is continuous, the mass at the kink is zero.

[6] Figure F.2 quantifies these two forces. The exercise applies the exact non-linear consumption policy for unconstrained households, while holding fixed the share of constrained households at its steady-state level and assigning constrained households an MPC of one. Thus, the main source of aggregate concavity is the movement of households off the constraint rather than the smooth concavity of unconstrained consumption policies.

2.5 Higher-order approximation

It is natural to conjecture that a more accurate solution could be obtained by taking a higher-order expansion around $\Delta = 0$. Figure 3 shows, however, that second- and third-order approximations to $\mathcal{C}_0(\Delta)$ deliver little improvement over the linear approximation over the empirically relevant range of shocks.[7] Panel (a) shows that, for transfers below 1% of annual GDP, the second- and third-order approximations improve on the linear solution, consistent with the concavity of the aggregate consumption function discussed above. Panel (b), however, shows that this improvement quickly disappears: for moderately larger shocks, the higher-order approximations become less accurate than the linear approximation. Thus, the results suggest that the relevant radius of convergence is small, making local approximations quantitatively unreliable for empirically relevant shocks.

The second-order expansion effectively represents the constrained share as declining at a constant rate. In the exact solution, however, the constrained share is highly nonlinear: it falls as transfers relax borrowing constraints, but eventually flattens at zero once all agents become unconstrained. As a result, the local approximation extrapolates poorly and can worsen relative to the linear solution even for relatively small shocks. The third-order term brings in the slope of the density at the kink, $g'(\bar{x}(y) \mid y)$, together with the curvature of the consumption policy at the kink and an interior third-derivative term; Appendix C derives this decomposition explicitly. This improves accuracy only in a narrow neighborhood of $\Delta = 0$. Beyond that neighborhood, the implied approximation to the constrained share deteriorates quickly, and the third-order approximation becomes less accurate than the second-order approximation once transfers exceed about 1.5% of annual GDP.

In other contexts, the analytical characterization of the second-order approximation may not be feasible. An alternative approach is to use finite differences to evaluate (1) to a second order. In Figure F.1 we report a wide range of examples that use Richardson extrapolation to combine approximations at different step sizes to cancel leading error terms. The figures show that finite difference approximations are extremely sensitive to the choice of step size and are typically very poorly behaved even for moderate shocks. Without knowledge of the non-linear solution it is unclear how one could use higher-order finite difference approximations around $\Delta = 0$ to obtain reliable solutions.

[7] We construct the second-order Taylor expansion of $\mathcal{C}_0(\Delta)$ using (3). Appendix C derives the third derivative of $\mathcal{C}_0(\Delta)$.

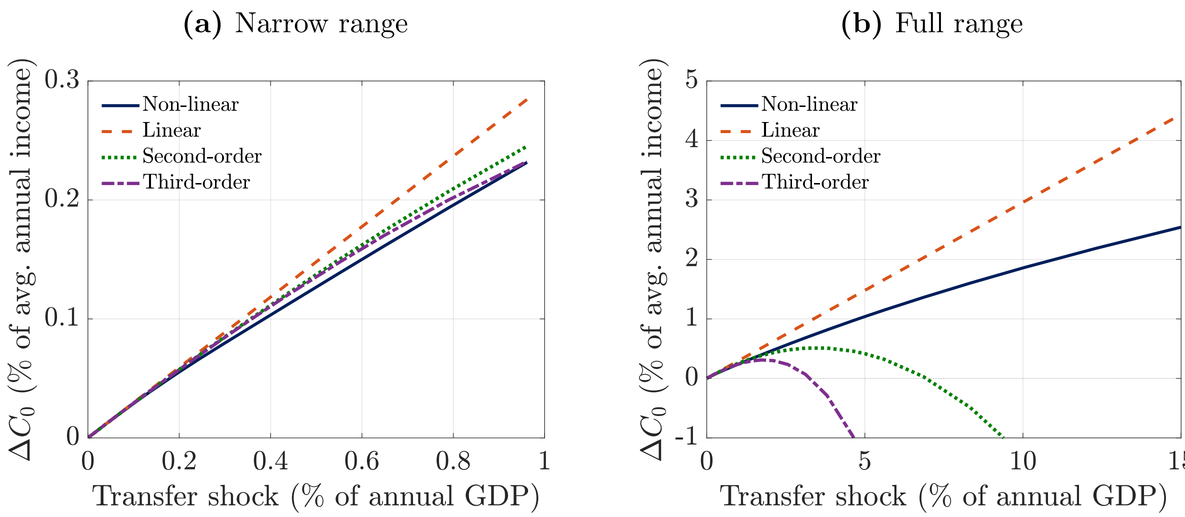


Figure 3: Local higher-order approximations of the impact consumption response.

3 General Equilibrium Model

In this section, we study the extent to which the non-linearity of the aggregate consumption function impacts the general equilibrium effects of one-time fiscal stimulus programs. We extend the model from the previous section to include production, labor supply, asset market clearing and nominal rigidities, as in canonical heterogeneous agent New Keynesian models (e.g., Kaplan and Violante, 2014; Auclert et al., 2023).

3.1 Model

Preferences Households have utility over real consumption c_{it} and hours worked n_{it}

$$\mathbb{E}_0 \sum_{t \geq 0} \beta_i^t \left[u(c_{it}) - v(n_{it}) \right]$$

where $v(\cdot)$ is a strictly increasing convex function in hours worked.

Income and assets Real household income consists of earnings $e_{it} \frac{W_t}{P_t} n_{it}$ and dividend income d_{it}, where e_{it} is an idiosyncratic productivity shock that follows a process analogous to the one for y_{it} in Section 2. In this section we denote total non-asset income as $y_{it} = e_{it} \frac{W_t}{P_t} n_{it} + d_{it}$. With this notation, the borrowing constraint and budget constraint are the same as in Section 2, with the return on inherited assets given by r_{t-1}.

Labor supply In our baseline model we assume flexible prices and sticky wages. The aggregate effective labor input $N_t = \int_i n_{it} e_{it} di$ is chosen by a labor union and is allocated across households so that each household works the same number of hours, implying that $n_{it} = N_t$, where we normalize the cross-sectional mean of efficiency units to one, $\int_i e_{it}\, di = 1$. In Section 3.4 we also report results from a version of the model with sticky prices, flexible wages, and a labor supply choice at the household level.

Production A representative firm produces the final good Y_t with a linear production function that uses effective labor N_t as its only factor of production. Equilibrium therefore implies a real wage of $\frac{W_t}{P_t} = 1$ per effective unit of labor and zero profits so that $d_{it} = 0$.

Wage Phillips curve We adopt the wage setting model in Auclert et al. (2024), in which monopolistically competitive labor unions set nominal wages subject to quadratic adjustment costs to maximize the welfare of a household with aggregate consumption $C_t = \int_i c_{it} di$ and hours N_t. This leads to the wage Phillips curve

$$\pi_t = \kappa \left(v'(N_t) - \frac{1-\tau}{C_t} \right) + \bar{\beta}\, \pi_{t+1}, \tag{4}$$

where π_t denotes wage (and price) inflation, κ depends on the wage adjustment cost and $\bar{\beta}$ is the average discount factor. This is the same linearized wage Phillips curve that arises in an analogous representative agent economy (e.g., Erceg et al., 2000). We choose to work directly with the linearized Phillips curve so as to focus entirely on the non-linearity of the consumption function stemming from the borrowing constraint.[8]

Fiscal policy The government issues real bonds B_t subject to the budget constraint[9]

$$B_t - B_{t-1} = G + T_t - \tau Y_t + r_{t-1} B_{t-1}, \tag{5}$$

where G is a fixed level of government consumption. Outside of steady state, the government adjusts the lump-sum transfer component of the tax-and-transfer schedule according to the rule

$$T_t = T - \phi_B \bar{r}(B_{t-1} - \bar{B}) \tag{6}$$

[8]Eggertsson and Singh (2019) examine this type of non-linearity in a representative agent New Keynesian model at the zero lower bound and find that these non-linearities play a very modest role.

Other examples of studies on the importance of non-linearities in different contexts with incomplete markets include Cao and Nie (2017) and de Groot, Durdu and Mendoza (2025).

[9]We present results for an economy in which the government issues nominal bonds in Section 3.4.

where $\phi_B > 1$ and $\bar{r}$ and $\bar{B}$ are the steady-state levels for the interest rate and government debt. This restriction implies that transfers are set so as to ensure that debt always returns to its steady-state level. Hence, fiscal policy is passive in the language of Leeper (1991).

Monetary policy The central bank sets the nominal interest rate according to the Taylor rule

$$i_t = \bar{i} + \phi_\pi(\pi_t - \bar{\pi}), \tag{7}$$

where $\bar{i}$ is the steady-state nominal rate, $\bar{\pi}$ is steady-state inflation and $\phi_\pi > 1$. Hence, monetary policy is active in the language of Leeper (1991). A standard Fisher equation connects real and nominal rates, $1 + i_t = (1 + r_t)(1 + \pi_{t+1})$ for $t \geq 0$.

Equilibrium Given fiscal and monetary policy rules (6) and (7), and an initial distribution of households over assets, productivity and discount factors, Γ_0, an equilibrium is a sequence of household policy functions $\{c_t(a, e, \beta), a_t'(a, e, \beta)\}_{t=0}^{\infty}$ and associated distributions $\{\Gamma_t(a, e, \beta)\}_{t=1}^{\infty}$, aggregate quantities $\{Y_t, C_t, B_t, N_t, d_t\}_{t=0}^{\infty}$ and prices $\{\pi_t, w_t, i_t, r_t\}_{t=0}^{\infty}$ such that (i) households and firms optimize; (ii) the wage Phillips curve (4) is satisfied; (iii) the government budget constraint (5) holds; (iv) the asset market clears

$$B_t = \int a_t'(a, e, \beta) d\Gamma_t(a, e, \beta) \quad \text{for } t \geq 0;$$

and (v) aggregate consumption is given by

$$C_t = \int c_t(a, e, \beta) d\Gamma_t(a, e, \beta) \quad \text{for } t \geq 0.$$

Walras' law implies the goods market clearing condition $C_t + G = Y_t$ holds.

3.2 Calibration

The calibration follows closely the parameterization of the partial equilibrium model in Section 2. Below we describe the features that are specific to the general equilibrium version. See Table A.4 in Appendix A.2 for further details.

Preferences We assume the following functional form for the disutility of labor

$$v(n) = \chi \frac{n^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}},$$

where ν is the Frisch elasticity of labor supply, which we set equal to 1. We normalize χ so that quarterly output equals one in steady state. As in Section 2, we choose the set of discount factors to match an average wealth to annual GDP ratio of 100% and an average quarterly MPC out of $1,000 of 25%.

Idiosyncratic income process The stochastic process for individual efficiency units of labor e_{it} is the same as the process for household labor income in Section 2.

Wage setting We set the slope of the wage Phillips curve $\kappa = 0.01$, as in Auclert et al. (2024).

Fiscal policy We set government consumption G to 10.4% of annual GDP. The steady-state tax rate and lump-sum transfer are the same as in Section 2. Hence the primary surplus as a percentage of annual GDP is $\tau - T - G = (0.274 - 0.15) - 0.104 = 2\%$. With a debt-to-GDP ratio of 100% this implies a steady-state interest rate of 2% p.a. We set the parameter governing the speed of repayment outside of steady-state to $\phi_B = 8$. This implies that the half-life of an increase in government debt is roughly 5 years.

Monetary policy We set steady-state inflation, $\bar{\pi} = 0$ so that the steady-state nominal rate is $\bar{i} = 2\%$. In our baseline experiments we set the Taylor rule coefficient $\phi_\pi = 1.5$.

3.3 Aggregate Transfers in General Equilibrium

Experiment As in Section 2, we study an unexpected one-time increase in the lump-sum transfer to all households at $t = 0$, which we implement by setting $T_0 = T + \Delta$. In our baseline experiment we assume that the stimulus is financed by lowering future lump-sum transfers for $t > 0$ according to the fiscal rule (6).

Computation We obtain the linear solutions using the SSJ, following Auclert et al. (2021). We compute non-linear solutions by finding the path of aggregate variables that satisfy the market clearing conditions for each size stimulus Δ.

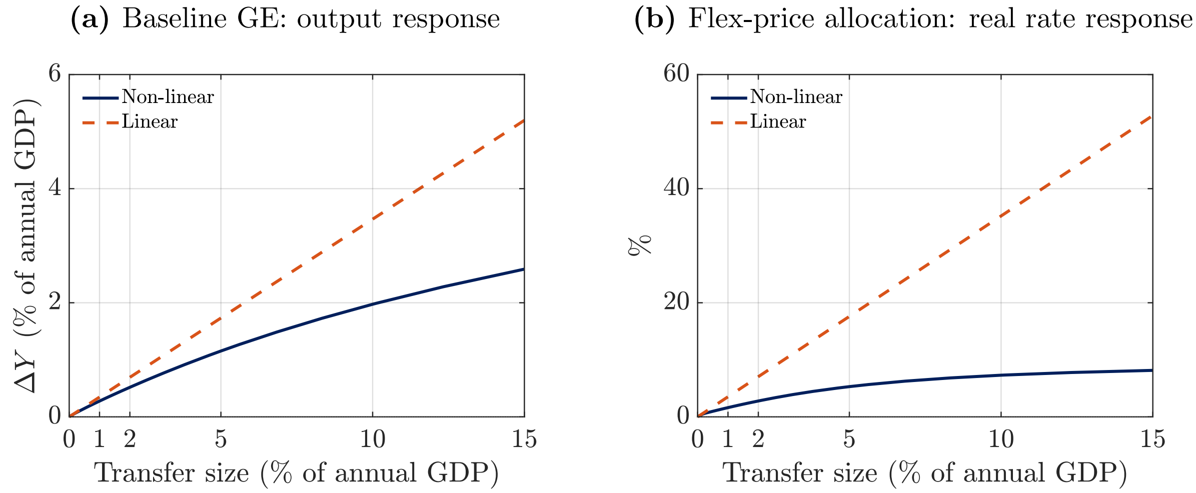


Figure 4: Aggregate effects of a one-time transfer in general equilibrium

Notes: The left panel presents first-quarter output response with sticky wages. The right panel presents first-quarter real interest rate response with flexible wages. The difference in rates is annualized and expressed in percent.

Baseline HANK results Figure 4a shows the aggregate output response for different values of Δ for the linear and non-linear solution methods. The non-linear solution is concave in the size of the shock. For a stimulus that is 2% of steady-state GDP (around \$3,320 per household), the first-quarter output response is around 0.53% of GDP according to the non-linear solution and around 0.71% according to the linear solution. The error in the linear solution is thus around 34%. For a stimulus of size 5% of steady-state GDP (around \$8,300), the first-quarter error in the linear solution is around 50%.

For reference, the fiscal transfers deployed during the Covid-19 pandemic were far larger—on the order of 12.5% of annual GDP, measured by transfers to households and firms together with the increase in government debt.[10] In practice, however, transfers are directed towards lower income households. In Figure G.1 in Appendix G, we show that the errors from linearization become much larger when conducting targeted transfers to the lower tail of the distribution instead of rebating lump-sum to all households.

Figure G.5 in Appendix G presents the impulse responses of output and other macroeconomic variables. As the figure shows, the linear solution implies that output reverts faster to steady state. Because the non-linear solution generates a more persistent output response,

[10] In the United States the main programs were the CARES Act (March 2020, $\approx$ \$2.2 trillion: direct payments to households, expanded unemployment insurance, and forgivable loans to firms under the Paycheck Protection Program), the December 2020 relief package ($\approx$ \$0.9 trillion), and the American Rescue Plan (March 2021, $\approx$ \$1.9 trillion)

this implies that over longer horizons the average error from the linear solution becomes smaller, but it remains economically significant. Figure G.4 reports the analogous first-year responses.

To understand where the errors in the linear solution stem from, it is useful to recap the mechanism for why a one-time transfer raises output in this model. Because of a failure of Ricardian equivalence, the combination of a higher transfer at $t = 0$ and lower future transfers leads households to want to increase their consumption, which leads to a rise in the natural interest rate. In the presence of nominal rigidities this leads to an inflationary output boom as long as the central bank does not fully accommodate the rise in the natural rate by raising the nominal rate.[11]

The strength of this mechanism depends on the initial upward pressure on household consumption, which is determined entirely by the departure from Ricardian equivalence, as reflected in the subset of households with MPCs above the PIH. Recall that in a representative agent version of this experiment there would be no effect on output for any size of stimulus. But as we saw in Figure 1, the high-MPC households are those who are on or close to the borrowing constraint and so have a consumption response that is highly non-linear in the size of the transfer. Thus the same forces that drive the concave response of consumption in the partial equilibrium experiments in Section 2 also lead to concavity in the output response in general equilibrium.

Interest rate response with flexible wages Before investigating the robustness of these findings to alternative model specifications, Figure 4b shows the interest rate response when there are no wage adjustment costs. The red dashed and blue solid lines show the linear and non-linear first-quarter response when unions continue to choose the hours on behalf of households. In this economy, the fiscal stimulus policy has no effects on employment, aggregate output or consumption. However, it does lead to movements in the real interest rate. The funded stimulus program is a mechanism to relieve the effects of borrowing constraints. Constrained households with high MPCs increase their consumption in response to the shock. Since output is unaffected, the interest rate must rise to induce higher wealth households to cut their consumption. Once again, because the strength of these effects is tied to the shape of the consumption function for constrained households, the effect is weaker for larger values of Δ.

[11] The precise mechanism by which output rises depends on the particular model of nominal rigidities. In this example with a monopolistic labor union, nominal wage rigidities and a fixed real wage, the union trades off the cost of wage adjustments with the utility cost of supplying labor in excess of the utility-maximizing level.

Figure G.3 in Appendix G shows the interest rate response in an analogous version of the model without a labor union, in which individual households can choose labor supply freely. In this model there are some small effects on output because of different effects on labor supply incentives for households with different productivity and asset levels. But since the extent of non-linearity is ultimately tied to the failure of Ricardian equivalence, the aggregate effect looks similar to the version with a union.

3.4 Other Versions of the HANK Model

The baseline general equilibrium model in the previous section features sticky wages, flexible prices, real government debt and an active monetary, passive fiscal policy configuration. In this section we investigate whether the extent of non-linearity in the effects of fiscal stimulus is affected by changes in these model features. The findings are summarized in Figure 5.

Faster fiscal adjustment. The green dash-dot line labeled "Faster fiscal adjustment" in Figure 5 shows that the non-linearity is more pronounced when the fiscal authority stabilizes debt more rapidly with a fiscal rule that sets $\phi_B = 16$ compared with $\phi_B = 8$ in the baseline. Since future transfers are lowered sooner it is only the most constrained households, those whose departure from Ricardian equivalence is most severe, whose consumption is affected. And as we have seen, these households have the most non-linear consumption functions.

Fixed real rate The orange dashed line labeled "Constant real rate rule" in Figure 5 shows results from an economy where the central bank keeps the real rate constant. This economy has slightly less non-linearity than the baseline economy.

Sticky prices and flexible wages An alternative model of nominal rigidities is one with sticky prices and flexible wages. We model sticky prices by introducing an intermediate-goods sector with a CES aggregator, monopolistic competition and quadratic price adjustment costs as in Rotemberg (1982). This version of the model features a standard New Keynesian Phillips curve of the form

$$\pi_t = \kappa \left(\frac{W_t}{P_t} - 1 \right) + \bar{\beta} \pi_{t+1}$$

Figure 5 displays results for two versions, one in which the union chooses a common number of hours for all households (light-blue solid line), and one in which households freely choose

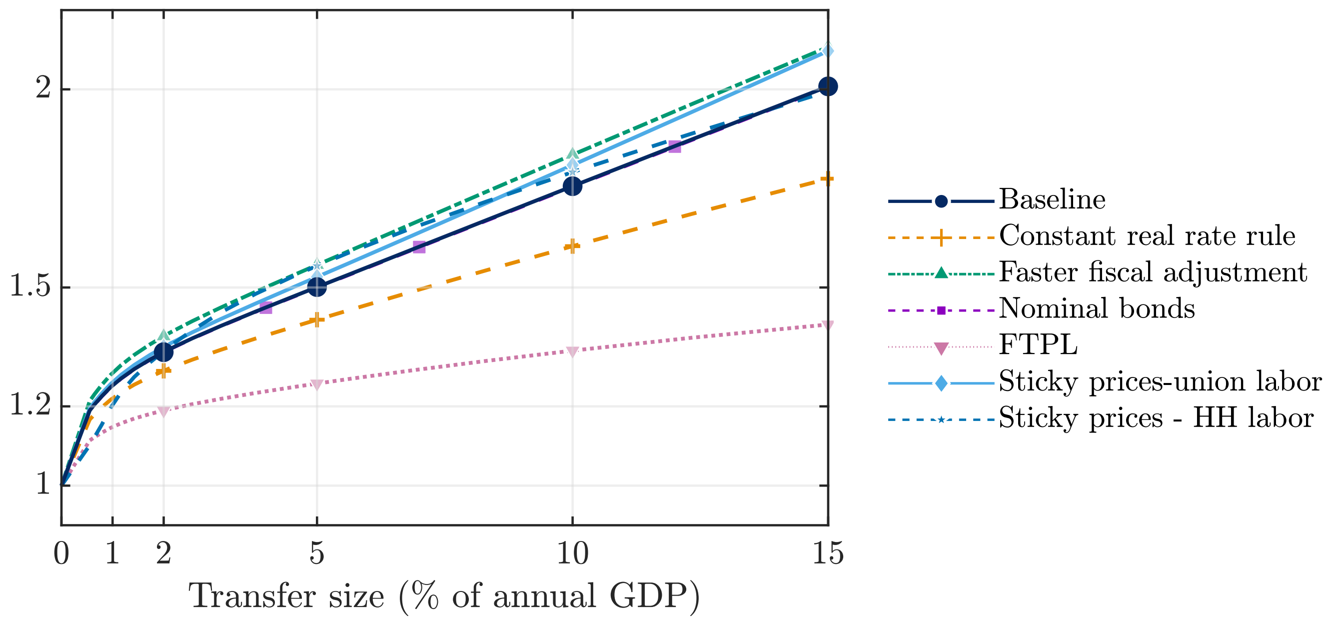


Figure 5: Sensitivity in HANK

Notes: The figure presents the ratio of the linear to the non-linear first-quarter output response for different model calibrations.

how much labor to supply (blue dashed line).[12] Both versions exhibit a degree of non-linearity similar to that in the baseline model.

Unfunded fiscal stimulus We also consider a version of the model in which future inflation, rather than future primary surpluses, pays for the stimulus, as in the Fiscal Theory of the Price Level. To implement this version we make two changes to the baseline model.

First, we assume that the government issues nominal, rather than real debt. This on its own has a negligible effect on the extent of non-linearity, as illustrated by the purple dashed line labeled "Nominal bonds" in Figure 5. The only effect of this change is that the unexpected inflation in the first period following the stimulus lowers the real value of debt that must be repaid with lower future surpluses.

Second, we alter the fiscal and monetary rules so that $\phi_B = \phi_\pi = 0$. This has the effect of pegging the nominal rate at $i_t = \bar{\imath}$ and holding real transfers fixed at their steady-state level $T_t = T$ for $t > 0$. Hence, in the language of Leeper (1991), fiscal policy is active and monetary policy is passive.

The ratio of the linear to non-linear first-quarter output response is shown with the pink

[12] When households choose labor supply, they set n_{it} so that $\frac{1}{c_{it}} e_{it}(1-\tau)\frac{W_t}{P_t} = v'(n_{it})$. When the labor union chooses labor supply they set N_t so that $\frac{1}{C_t}\frac{W_t}{P_t}(1-\tau) = v'(N_t)$.

dotted line labeled "FTPL" in Figure 5. The figure shows that for an unfunded stimulus the extent of non-linearity is much less severe than for the other versions of the model.[13] The reason is that a large part of the output effect of unfunded stimulus does not rely on a failure of Ricardian equivalence or households having MPCs above the PIH. Even in a representative agent model, an unfunded stimulus of this type has an expansionary effect on output. The upward pressure on consumption arises because households seek to exchange nominal government debt for real consumption, as explained in Cochrane (2023). Redistribution towards high-MPC households plays a role as well, as explained by Kaplan, Nikolakoudis and Violante (2024), but it is a small part of the overall effect on output.

4 Other Shocks

We have focused on a particular policy experiment—a one-time lump-sum fiscal transfer—in which a failure of Ricardian equivalence is crucial for the aggregate effects. In this section we discuss other types of shocks in which Ricardian equivalence plays less of a role, and show that for these shocks, the non-linearities are less severe.

Monetary shocks. We compare the first-quarter output response to a one-quarter cut in the nominal rate of different sizes, using linear and non-linear solution methods and find that the two solution methods yield almost identical responses, consistent with simulations in Auclert et al. (2021) (see Figure H.1a). This should not be surprising, since we know from Werning (2015) and Kaplan et al. (2018) that heterogeneity has only a small impact on the aggregate effects of monetary policy shocks. Rather, the nature of MPCs affects the transmission mechanism and decomposition between direct and indirect effects.

Productivity shocks. We compare the first-quarter response to an aggregate productivity shock using linear and non-linear solution methods. We model the shock by modifying the production function to be $Y_t = e^{Z_t} N_t$ and consider a one-time shock to Z_0. The two solution methods yield almost identical responses, which are very close to those in an analogous representative agent economy (see Figure H.1b). The reason is that failures of Ricardian equivalence, and hence the size and nature of MPCs, play a minimal role in the transmission of productivity shocks.

[13] Figure G.6 in Appendix G reports the aggregate output effect of the unfunded stimulus, which is larger than for the funded stimulus.

Additional shocks. For similar reasons, shocks to the household discount factor and to the labor tax rates yield far less non-linearity than the fiscal stimulus shock (see panels [c] and [d] of Figure H.1). [14]

5 Conclusions

Heterogeneous agent models are useful for studying the effects of fiscal stimulus because the interaction between borrowing constraints and incomplete markets for idiosyncratic risk leads to a failure of Ricardian equivalence. However, the resulting MPCs are highly non-linear in household wealth. As a result, the aggregate effects of fiscal transfers to households with high MPCs decline rapidly with their size. Using first-order methods in this class of models effectively treats constrained households as if they had an MPC of 1 regardless of the size of the transfer and overstates their true response. For some shocks and policy experiments, this approximation is accurate. But if local linear methods such as SSJ are to yield reliable quantitative results when studying fiscal stimulus—particularly redistributive stimulus targeted to high-MPC households—then "small" must be very small indeed.

[14] Auclert et al. (2021) also find small non-linearities with respect to labor tax cuts. Through the lens of our analysis, the reason for the lower concavity is that a cut in the labor tax shifts resources from high MPC low-income households to low MPC high-income households.

References

Aggarwal, Rishabh, Adrien Auclert, Matthew Rognlie, and Ludwig Straub, "Excess savings and twin deficits: The transmission of fiscal stimulus in open economies," *NBER Macroeconomics Annual*, 2023, *37* (1), 325–412.

Angeletos, George-Marios, Chen Lian, and Christian K Wolf, "Can deficits finance themselves?," *Econometrica*, 2024, *92* (5), 1351–1390.

_ , _ , and _ , "Deficits and Inflation: HANK meets FTPL," 2024.

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 2021, *89* (5), 2375–2408.

_ , M. Rognlie, and L. Straub, "The trickling up of excess savings," *AEA Papers and Proceedings*, 2023, *113*, 70–75.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub, "The intertemporal keynesian cross," *Journal of Political Economy*, 2024, *132* (12), 4068–4121.

Bardoczy, Bence, Jae Sim, and Andreas Tischbirek, "The Macroeconomic Effects of Excess Savings," 2024. Board of Governors of the Federal Reserve System.

Boppart, Timo, Per Krusell, and Kurt Mitman, "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative," *Journal of Economic Dynamics and Control*, 2018, *89*, 68–92.

Brinca, Pedro, Miguel Faria e Castro, Miguel H Ferreira, Hans A Holter, and Valter Nóbrega, "The nonlinear effects of fiscal policy," *Journal of Public Economics*, 2025, *252*, 105517.

Campos, Rodolfo G, Jesús Fernández-Villaverde, Galo Nuño, and Peter Paz, "Navigating by falling stars: monetary policy with fiscally driven natural rates," Technical Report, National Bureau of Economic Research 2024.

Cao, Dan and Guangyu Nie, "Amplification and asymmetric effects without collateral constraints," *American Economic Journal: Macroeconomics*, 2017, *9* (3), 222–266.

Carroll, Christopher D, "The method of endogenous gridpoints for solving dynamic stochastic optimization problems," *Economics Letters*, 2006, *91* (3), 312–320.

— **and Miles S Kimball**, "On the concavity of the consumption function," *Econometrica*, 1996, *64* (4), 981–992.

Cochrane, John H., *The Fiscal Theory of the Price Level*, Princeton University Press, 2023.

de Groot, Oliver, C. Bora Durdu, and Enrique G. Mendoza, "Why Global and Local Solutions of Open-Economy Models with Incomplete Markets Differ and Why It Matters," *Journal of International Economics*, 2025, *158*, 104142.

Eggertsson, Gauti B. and Sanjay R. Singh, "Log-Linear Approximation versus an Exact Solution at the ZLB in the New Keynesian Model," *Journal of Economic Dynamics and Control*, 2019, *105*, 21–43.

Eichenbaum, Martin, Joao Guerreiro, and Jana Obradovic, "Ricardian non-equivalence," Technical Report, Working Paper 2025.

Erceg, Christopher J, Dale W Henderson, and Andrew T Levin, "Optimal monetary policy with staggered wage and price contracts," *Journal of Monetary Economics*, 2000, *46* (2), 281–313.

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman, "The fiscal multiplier," Technical Report, National Bureau of Economic Research 2019.

Hänsel, Matthias, "Idiosyncratic Risk, Government Debt and Inflation," *arXiv preprint arXiv:2403.00471*, 2024.

Kaplan, Greg and Giovanni L Violante, "A model of the consumption response to fiscal stimulus payments," *Econometrica*, 2014, *82* (4), 1199–1239.

— **and** — , "The marginal propensity to consume in heterogeneous agent models," *Annual Review of Economics*, 2022, *14* (1), 747–775.

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

— **, Evangelos Nikolakoudis, and Giovanni L. Violante**, "Non-linearities and the Redistributive Effects of Unfunded Fiscal Stimulus," Technical Report, Working Paper 2024.

Leeper, Eric M, "Equilibria under active and passive monetary and fiscal policies," *Journal of Monetary Economics*, 1991, *27* (1), 129–147.

Mongey, Simon and Michael E Waugh, "Pricing inequality," Technical Report, National Bureau of Economic Research 2025.

Rachel, Lukasz and Lawrence H Summers, "On falling neutral real rates, fiscal policy and the risk of secular stagnation," in "Brookings Papers on Economic Activity BPEA Conference Drafts, March," Vol. 7 2019, p. 66.

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

Rotemberg, Julio J., "Sticky prices in the United States," *Journal of Political Economy*, 1982, *90* (6), 1187–1211.

Werning, Ivan, "Incomplete markets and aggregate demand," Technical Report, National Bureau of Economic Research 2015.

Online Appendix to "How Small is Small? Non-linearities in Heterogeneous Agent Models"

Javier Bianchi and Greg Kaplan

A Calibration

A.1 Income Process

This appendix describes the estimation of the income process used in our quantitative analysis. We follow closely the approach of Kaplan and Violante (2022) (hereafter KV).

Panel Study of Income Dynamics

We use data from the Panel Study of Income Dynamics (PSID) on total annual household labor income for households with heads aged 25 to 65 from 1968 to 2008, following KV. We drop households with annual labor income less than $7,250 in 2016 dollars, which corresponds to 1,000 hours per year at $7.25 per hour (part-time employment at the minimum wage). We remove age and year effects in a first stage by regressing household labor income on a full set of year and age dummies and we construct the empirical counterparts to $m_{2,d}$ using the residuals from this regression.

Let $\log y_t^{ann}$ be annual labor income in year t, and let annual income growth at lag d be

$$\Delta_d \log y_t^{ann} = \begin{cases} \log y_t^{ann} & \text{if } d = 0, \\ \log y_{t+d}^{ann} - \log y_t^{ann} & \text{if } d > 0. \end{cases}$$

Define cross-sectional moments of annual income growth of order j at lag d as

$$m_{j,d} = \mathbb{E}\Big[\big(\Delta_d \log y_t^{ann}\big)^j\Big]$$

and the kurtosis of income growth at lag d as

$$\kappa_d = \frac{m_{4,d}}{(m_{2,d})^2}.$$

Table A.1 reports the empirical estimates for the cross-sectional moments that we use in estimation, which coincide with those in KV.

Table A.1: Empirical moments of annual income growth at different lags.
Source: PSID 1968–2008.

Lag (d)	$m_{2,d}$	$m_{4,d}$	κ_d
0	0.504	0.930	3.65
1	0.142	0.220	10.90
2	0.207	0.369	8.57
3	0.235	0.410	7.42
4	0.280	0.544	6.96
5	0.295	0.557	6.39

Estimated Income Processes

We model the discrete-time quarterly income process, as in KV:

$$\log y_t = \begin{cases} z_t + \varepsilon_t & \text{with probability } \lambda_\varepsilon, \qquad \varepsilon_t \sim \mathcal{N}\left(-\frac{\sigma_\varepsilon^2}{2},\, \sigma_\varepsilon^2\right), \\ z_t & \text{with probability } 1-\lambda_\varepsilon, \end{cases}$$

$$z_t = \begin{cases} \phi_z\, z_{t-1} + \eta_t & \text{with probability } \lambda_\eta, \qquad \eta_t \sim \mathcal{N}\left(-\frac{\sigma_\eta^2}{2},\, \sigma_\eta^2\right), \\ \phi_z\, z_{t-1} & \text{with probability } 1-\lambda_\eta. \end{cases}$$

We define annual income y^{ann} as the sum of the four quarterly incomes within the year. Based on this definition, we construct the model counterparts of all the empirical moments in Table A.1.

With $\lambda_\varepsilon = \lambda_\eta = 1/4$ fixed—so that each type of shock arrives on average once per year—we have three free parameters $(\phi_z,\, \sigma_\eta^2,\, \sigma_\varepsilon^2)$. We estimate them by minimizing the sum of squared deviations between model and data second moments $(m_{2,0},\, m_{2,1},\, m_{2,5})$.

We discretize the shock process using $n_z = 11$ Rouwenhorst states for the persistent component and $n_\varepsilon = 5$ points for the transitory component, for a total of 55 income states.

Table A.2 reports the estimated parameters. The persistent component has quarterly autocorrelation $\phi_z = 0.9878$ and innovation variance $\sigma_\eta^2 = 0.0439$; the transitory shock has variance $\sigma_\varepsilon^2 = 0.6376$, implying large but infrequent income fluctuations.

Specifically, the transitory mixture distribution $(1-\lambda_\varepsilon)\,\delta_0 + \lambda_\varepsilon\,\mathcal{N}(-\sigma_\varepsilon^2/2,\, \sigma_\varepsilon^2)$ is approximated with $n_\varepsilon = 5$ points by solving for the grid and probability weights that match its second and fourth central moments. The resulting grid spans $\log y^\varepsilon \in \{-2.85,\ -1.42,\ 0,\ 1.42,\ 2.85\}$ with probability weights $\{\approx 0,\ 0.042,\ 0.917,\ 0.042,\ \approx 0\}$, placing 91.7% of the mass at the no-shock point, consistent with the 75% no-shock probability from $\lambda_\varepsilon = 0.25$ plus the mass contributed by the discretized normal at zero.

Table A.2: Parameter estimates for the quarterly income process.

Parameter	Description	Symbol	Value
A. Structural parameters			
AR(1) persistence	Persistent component	ϕ_z	0.9878
Persistent shock variance	Innovation variance	σ^2_η	0.0439
Transitory shock variance	Transitory variance	σ^2_ε	0.6376
Persistent arrival rate	Quarterly probability	λ_η	0.250
Transitory arrival rate	Quarterly probability	λ_ε	0.250
B. Discretization			
Persistent states	Rouwenhorst grid	n_z	11
Transitory states	Mixture approximation	n_ε	5
Total income states		$n_z \times n_\varepsilon$	55
C. Targeted moments			
Variance, lag 0		$m_{2,0}$	0.504
Variance, lag 1		$m_{2,1}$	0.142
Variance, lag 5		$m_{2,5}$	0.295

A.2 Further Details on Calibration

Table A.3: Calibration: parameter values and steady-state moments across PE scenarios

	Parameters		Moments	
Scenario	$\bar{\beta}$	σ_β	Avg. MPC	A/Y
Baseline	0.970	0.011	0.25	1.00
No β heterog., $A/Y = 1$	0.991	0	0.09	1.00
No β heterog., $A/Y = 4$	0.993	0	0.05	4.00
Stochastic β	0.993	0.066	0.25	1.00
Finite lives	0.978	0.011	0.25	1.00
Lower transfers	0.966	0.013	0.25	1.00
Natural debt limit	0.961	0.016	0.25	1.00
Endogenous labor supply	0.956	0.018	0.32	1.00
ARS income process	0.976	0.006	0.25	1.00

Note: Avg. MPC is the steady-state average marginal propensity to consume. The versions of the model are as follows: *No β heterog. ($A/Y = 1$)* and *No β heterog. ($A/Y = 4$)* are versions without discount-factor heterogeneity, calibrated to match aggregate wealth-to-(annualized) income ratios of 1 and 4, respectively. *Stochastic β*: discount factors follow an idiosyncratic stochastic process AR(1) with autocorrelation $\rho_\beta = 0.990$. *Finite lives*: a life-cycle version with overlapping generations and an annual mortality probability of 2%. *Lower transfers*: steady-state lump-sum transfers are set to half of the baseline value. *Natural debt limit*: households face a 0.1% probability of transitioning to a zero-income state, and their lump-sum transfers are set to zero. *Endogenous labor supply*: households choose hours worked endogenously with isoelastic disutility from labor and unitary Frisch elasticity. *ARS*: replaces the baseline idiosyncratic earnings process with the discretized process from Auclert et al. (2024).

Table A.4: Summary of parameters for general equilibrium model

Parameter	Value	Description
$\bar{B}$	4	Steady-state government debt
$\bar{r}$	0.02	Steady-state annual real interest rate (model equations use the quarterly rate $\bar{r}/4 =$
$\bar{\pi}$	0	Steady-state inflation
κ	0.01	Phillips curve slope
ν	1	Frisch elasticity of labor supply
ϕ_π	1.5	Taylor coefficient
ϕ_B	8	Fiscal rule
G	0.104	Government spending
T	0.15	Government transfers
χ	1	Labor disutility

B Analytical Characterization of MPCs

This appendix derives analytical first-order and higher-order derivatives of the individual consumption function.

Recall from Section 2 that x denotes cash on hand and $c(x, y)$ the consumption policy, with y the income state; we write the MPC as $c_x(x, y)$.[15] Next-period cash on hand is

$$x' = (1 + r)[x - c(x, y)] + y',$$

where y' is next-period non-asset income.

Let $\bar{x}(y)$ denote a borrowing-constraint boundary point: locally, the constraint binds for $x < \bar{x}(y)$ and is slack for $x > \bar{x}(y)$. Define the one-sided MPCs at the boundary by

$$c_x^-(\bar{x}(y), y) \equiv \lim_{\varepsilon \downarrow 0} c_x(\bar{x}(y) - \varepsilon, y), \qquad c_x^+(\bar{x}(y), y) \equiv \lim_{\varepsilon \downarrow 0} c_x(\bar{x}(y) + \varepsilon, y).$$

Let $V(a, y)$ denote the beginning-of-period value function, with asset state a. Given cash-on-hand x, the household solves

$$\max_{a' \geq \underline{a}} \left\{ u(x - a') + \beta \mathbb{E}[V(a', y') \mid y] \right\}.$$

The consumption policy is

$$c(x, y) = x - a'(x, y).$$

For the slack branch, use the notation

$$c \equiv c(x, y), \qquad c' \equiv c(x', y'), \qquad c_x \equiv c_x(x, y), \qquad (c_x)' \equiv c_x(x', y'),$$

and

$$(c_{xx})' \equiv c_{xx}(x', y'), \qquad (c_{xxx})' \equiv c_{xxx}(x', y').$$

Define

$$\Lambda \equiv \beta(1 + r)^2 \mathbb{E}\left[u''(c')(c_x)' \mid y\right],$$

$$\Psi \equiv \beta(1 + r) \mathbb{E}\left[u'''(c')((c_x)')^2 + u''(c')(c_{xx})' \mid y\right],$$

$$\Theta \equiv \beta(1 + r) \mathbb{E}\left[u''''(c')((c_x)')^3 + 3u'''(c')(c_x)'(c_{xx})' + u''(c')(c_{xxx})' \mid y\right].$$

[15] The consumption policy varies across households with different β; we do not index it to simplify notation.

Constrained branch. First consider the constrained branch. If the borrowing constraint binds, then

$$a'(x,y) = \underline{a}.$$

Therefore

$$c(x,y) = x - \underline{a}.$$

Differentiating with respect to cash-on-hand gives

$$c_x(x,y) = 1, \qquad c_{xx}(x,y) = 0, \qquad c_{xxx}(x,y) = 0. \tag{B.1}$$

Thus

$$c_x^-(\bar{x}(y), y) = 1.$$

Slack branch: first derivative. Now consider the slack branch, where $a'(x,y) > \underline{a}$. The first-order condition is

$$u'(c(x,y)) = \beta \mathbb{E}\left[V_a(a'(x,y), y') \mid y\right]. \tag{B.2}$$

At smooth continuation points, differentiating (B.2) with respect to x gives

$$u''(c)c_x = \beta \mathbb{E}\left[V_{aa}(a'(x,y), y') \mid y\right] a'_x.$$

Since

$$a'_x = 1 - c_x(x,y),$$

we obtain

$$u''(c)c_x = \widetilde{\Lambda}(1 - c_x), \tag{B.3}$$

where

$$\widetilde{\Lambda} \equiv \beta \mathbb{E}\left[V_{aa}(a'(x,y), y') \mid y\right].$$

If the continuation value is strictly concave at the smooth continuation points reached from the slack branch, then

$$\widetilde{\Lambda} < 0.$$

Since $u''(c) < 0$, equation (B.3) implies

$$c_x(x,y) = \frac{\widetilde{\Lambda}}{u''(c) + \widetilde{\Lambda}}, \qquad 0 < c_x(x,y) < 1. \tag{B.4}$$

Thus the slack-side MPC is strictly below the constrained-side MPC.

The expression above can be written in terms of the Euler equation. At a smooth continuation point, the envelope condition gives

$$V_a(a', y') = (1 + r)u'(c(x', y')), \qquad x' = (1 + r)a' + y'. \tag{B.5}$$

Differentiating (B.5) with respect to a' gives

$$V_{aa}(a', y') = (1 + r)^2 u''(c')(c_x)'.$$

Therefore

$$\widetilde{\Lambda} = \beta(1 + r)^2 \mathbb{E}\left[u''(c')(c_x)' \mid y\right] \equiv \Lambda.$$

Substituting this into (B.4) gives

$$c_x(x, y) = \frac{\Lambda}{u''(c) + \Lambda}. \tag{B.6}$$

Slack branch: second derivative. The Euler equation on the slack branch is

$$u'(c) = \beta(1 + r)\mathbb{E}\left[u'(c') \mid y\right]. \tag{B.7}$$

Define

$$F(z, y') \equiv u'(c(z, y')).$$

Then, evaluated at $z = x'$,

$$F_z = u''(c')(c_x)', \tag{B.8}$$
$$F_{zz} = u'''(c')((c_x)')^2 + u''(c')(c_{xx})', \tag{B.9}$$
$$F_{zzz} = u''''(c')((c_x)')^3 + 3u'''(c')(c_x)'(c_{xx})' + u''(c')(c_{xxx})'. \tag{B.10}$$

Also,

$$x'_x = (1 + r)(1 - c_x), \qquad x'_{xx} = -(1 + r)c_{xx}, \qquad x'_{xxx} = -(1 + r)c_{xxx}. \tag{B.11}$$

Differentiating (B.7) twice with respect to x gives

$$u'''(c)c_x^2 + u''(c)c_{xx} = \beta(1 + r)\mathbb{E}\left[F_{zz}(x'_x)^2 + F_z x'_{xx} \mid y\right].$$

Using (B.8), (B.9), and (B.11), this becomes

$$u'''(c)c_x^2 + u''(c)c_{xx} = [(1+r)(1-c_x)]^2\Psi - \Lambda c_{xx}.$$

Rearranging yields

$$c_{xx}(x,y) = \frac{[(1+r)(1-c_x)]^2\Psi - u'''(c)c_x^2}{u''(c)+\Lambda}. \tag{B.12}$$

Slack branch: third derivative. Differentiating (B.7) three times with respect to x gives

$$u''''(c)c_x^3 + 3u'''(c)c_x c_{xx} + u''(c)c_{xxx} = \beta(1+r)\mathbb{E}\left[F_{zzz}(x'_x)^3 + 3F_{zz}x'_x x'_{xx} + F_z x'_{xxx} \mid y\right].$$

Using (B.8), (B.9), (B.10), and (B.11), the right-hand side is

$$[(1+r)(1-c_x)]^3\Theta - 3(1+r)^2(1-c_x)c_{xx}\Psi - \Lambda c_{xxx}.$$

Therefore

$$u''''(c)c_x^3 + 3u'''(c)c_x c_{xx} + u''(c)c_{xxx} = [(1+r)(1-c_x)]^3\Theta - 3(1+r)^2(1-c_x)c_{xx}\Psi - \Lambda c_{xxx}.$$

Rearranging gives

$$c_{xxx}(x,y) = \frac{[(1+r)(1-c_x)]^3\Theta - 3(1+r)^2(1-c_x)c_{xx}\Psi - u''''(c)c_x^3 - 3u'''(c)c_x c_{xx}}{u''(c)+\Lambda}. \tag{B.13}$$

Summary. Combining the constrained-branch derivatives in (B.1) with the slack-branch formulas in (B.6), (B.12), and (B.13), we obtain

$$c_x(x,y) = \begin{cases} 1, & x < \bar{x}(y), \\ \dfrac{\Lambda}{u''(c)+\Lambda}, & x > \bar{x}(y), \end{cases}$$

$$c_{xx}(x,y) = \begin{cases} 0, & x < \bar{x}(y), \\ \dfrac{[(1+r)(1-c_x)]^2\Psi - u'''(c)c_x^2}{u''(c)+\Lambda}, & x > \bar{x}(y), \end{cases}$$

$$c_{xxx}(x,y) = \begin{cases} 0, & x < \bar{x}(y), \\ \dfrac{[(1+r)(1-c_x)]^3\Theta - 3(1+r)^2(1-c_x)c_{xx}\Psi - u''''(c)c_x^3 - 3u'''(c)c_x c_{xx}}{u''(c)+\Lambda}, & x > \bar{x}(y). \end{cases}$$

In particular,

$$c_x^-(\bar{x}(y),y) = 1, \qquad 0 < c_x^+(\bar{x}(y),y) < 1,$$

so the MPC is discontinuous at the point where the borrowing constraint becomes slack.

C Derivatives of $\mathcal{C}(\Delta)$ up to third order

We impose the following regularity conditions, under which the differentiation below is valid. For each income y, the conditional distribution of cash on hand $G(\cdot \mid y)$ is absolutely continuous in a neighborhood of the kink $\bar{x}(y)$, with a density $g(\cdot \mid y)$ that is continuous there (and continuously differentiable for the third-order term); in particular, no positive mass of households sits exactly at the kink. The unconstrained branch $c^{\text{unc}}(\cdot, y)$ is three-times continuously differentiable above the kink. These conditions justify differentiating under the integral and letting the boundary terms enter as densities rather than discrete jumps. In the numerical model, where income takes discrete states and asset policies are interpolated, they are to be understood as holding for the smoothed cash-on-hand distribution used in the derivative computations.

Fix an income y and write $g(x) = g(x \mid y)$, G for its c.d.f., and $b(\Delta) = \bar{x}(y) - \Delta$ (so $b' = -1$). The inner bracket in (1) is

$$I(\Delta) = \int_{-\infty}^{b} c^{\text{con}}(x{+}\Delta)\, g(x)\, \mathrm{d}x + \int_{b}^{\infty} c^{\text{unc}}(x{+}\Delta)\, g(x)\, \mathrm{d}x, \qquad \mathcal{C}(\Delta) = \int_{Y} I(\Delta)\, \mathrm{d}F(y).$$

Each derivative combines an *interior* term (differentiating the integrand) with a *boundary* term (differentiating the moving limit b). Because consumption is continuous at the kink the boundary terms in I' cancel; using the constrained slope $c_x^{\text{con}} = 1$,

$$I'(\Delta) = G(b) + \int_{b}^{\infty} c_x^{\text{unc}}(x{+}\Delta)\, g\, \mathrm{d}x.$$

Differentiating again, the slope is discontinuous at the kink so the boundary term now survives:

$$I''(\Delta) = -\big[1 - c_x^{\text{unc}}(\bar{x})\big]\, g(b) + \int_{b}^{\infty} c_{xx}^{\text{unc}}(x{+}\Delta)\, g\, \mathrm{d}x.$$

A third differentiation acts on the density $g(b)$ (producing its slope g'), on the surviving curvature boundary term, and on the interior:

$$I'''(\Delta) = \big[1 - c_x^{\text{unc}}(\bar{x})\big]\, g'(b) + c_{xx}^{\text{unc}}(\bar{x})\, g(b) + \int_{b}^{\infty} c_{xxx}^{\text{unc}}(x{+}\Delta)\, g\, \mathrm{d}x.$$

Evaluating at $\Delta = 0$ (so $b = \bar{x}(y)$) and integrating over y yields

$$\mathcal{C}'(0) = \int_Y G(\bar{x} \mid y)\,\mathrm{d}F + \int_{x>\bar{x}} c_x\,\mathrm{d}\Gamma,$$

$$\mathcal{C}''(0) = -\int_Y \big(1 - c_x^+(\bar{x})\big)\, g(\bar{x} \mid y)\,\mathrm{d}F + \int_{x>\bar{x}} c_{xx}\,\mathrm{d}\Gamma,$$

$$\mathcal{C}'''(0) = \underbrace{\int_Y \big(1 - c_x^+(\bar{x})\big)\, g'(\bar{x} \mid y)\,\mathrm{d}F}_{\text{density slope}} + \underbrace{\int_Y c_{xx}(\bar{x})\, g(\bar{x} \mid y)\,\mathrm{d}F}_{\text{curvature at the kink}} + \underbrace{\int_{x>\bar{x}} c_{xxx}\,\mathrm{d}\Gamma}_{\text{interior}}.$$

D Numerical Algorithm

D.1 Partial Equilibrium Algorithm

In partial equilibrium, the path for prices is exogenous, so the problem reduces to solving the household block and aggregating over the induced distribution. The algorithm is as follows.

1. Discretize the idiosyncratic income on Markov chains (see Appendix A), choose an asset grid $\mathcal{A}$ and set a horizon $\mathbb{T}$.

2. For a given transfer path, solve the household problem using the endogenous-grid method (Carroll, 2006), interpolating linearly for assets not in the grid.

3. Given the policy rules, propagate the cross-sectional distribution forward until it converges to the ergodic distribution $\Gamma(a, y, \beta)$.

4. Compute the linear response by finite differences. We solve the full household transition under a small transfer ε at $t = 0$. At $t = 0$,

$$s_0 \;=\; \frac{C_0^{nl}(+\varepsilon) - C_0^{nl}(-\varepsilon)}{2\varepsilon},$$

where

$$C_0^{nl}(\varepsilon) \equiv \int c_i((1+r)a + y_i + \varepsilon,\, y_i;\, \beta_i)\,\mathrm{d}\Gamma_0(a, i),$$

where the first argument of the policy is period-0 cash on hand: the transfer ε augments cash on hand $(1 + r)a + y_i$, not beginning-of-period assets. Here β_i is the household's discount factor and Γ_0 is the (given) steady-state distribution over assets a and types i. We use $\varepsilon = 1.0 \times 10^{-5}$. Automatic differentiation and central difference yield differences in the order of 10^{-7}. For ε that is not in the grid for $c(\cdot)$, we interpolate linearly.

Then, using the first-order derivative s_0 compute the linear impulse response to a transfer of size Δ as

$$\Delta C_0^{lin} \;=\; \Delta \cdot s_0.$$

For $t \geq 1$ the transfer has already been paid at $t = 0$, so it propagates only through the distribution, which is advanced from the period-0 perturbed distribution under the (time-invariant) steady-state policies:

$$C_t^{nl}(\varepsilon) \equiv \int c_i((1+r)a + y_i,\, y_i;\, \beta_i)\, \mathrm{d}\Gamma_t(a, i;\, \varepsilon),$$

where $\Gamma_t(\cdot\,;\varepsilon)$ is the date-t distribution along the transition induced by the $t = 0$ transfer ε. The horizon-t linear response is then

$$s_t \;=\; \frac{C_t^{nl}(+\varepsilon) - C_t^{nl}(-\varepsilon)}{2\varepsilon}, \qquad \Delta C_t^{lin} \;=\; \Delta \cdot s_t.$$

5. Compute the non-linear response by evaluating the individual consumption policies at the transfer and aggregating using the distribution, which is propagated forward. That is,

$$C_t^{nl}(\Delta) \equiv \int c_i\big((1+r)a + y_i + \Delta\, \mathbf{1}\{t = 0\},\, y_i;\, \beta_i\big)\, \mathrm{d}\Gamma_t(a, i;\, \Delta),$$

where in partial equilibrium the policy functions are time-invariant (prices are fixed): the transfer enters period-0 cash on hand directly and, for $t \geq 1$, only through the propagated distribution $\Gamma_t(\cdot\,;\Delta)$.

For the asset grid we use an exponential grid with $n_a = 200$ points on $[0, \bar{a}]$, where $\bar{a} = 4{,}000$. We use $\mathbb{T} = 32$.

E General Equilibrium

Let the unknown transition path be

$$X \equiv \{C_t, N_t, Y_t, \pi_t, i_t, r_t, B_t\}_{t=0}^{\mathbb{T}-1}.$$

We compute the first-order transition using the sequence-space Jacobian method of Auclert et al. (2021). We use their Python toolkit to compute it with $\mathbb{T} = 150$.

To solve the household problem non-linearly, we first conjecture a vector of sequences X, and then solve backward for policy functions using EGM and then propagate the distribution forward. This yields household asset demand $A_t(X; s)$ at each date. We then evaluate the residuals from the equilibrium conditions and iterate with a modified Newton method.

Specifically, we conjecture $\{Y_t, \pi_t\}_{t=0}^{\mathbb{T}-1}$ and then obtain $\{C_t, N_t, i_t, B_t, r_t\}_{t=0}^{\mathbb{T}-1}$ using $Y_t = N_t = C_t + G$ and

$$
\begin{aligned}
i_t &= \bar{i} + \phi_\pi(\pi_t - \bar{\pi}), \\
1 + r_t &= \frac{1 + i_t}{1 + \pi_{t+1}}. \\
B_t &= (1 + r_{t-1})B_{t-1} + G + T_t - \tau Y_t.
\end{aligned}
$$

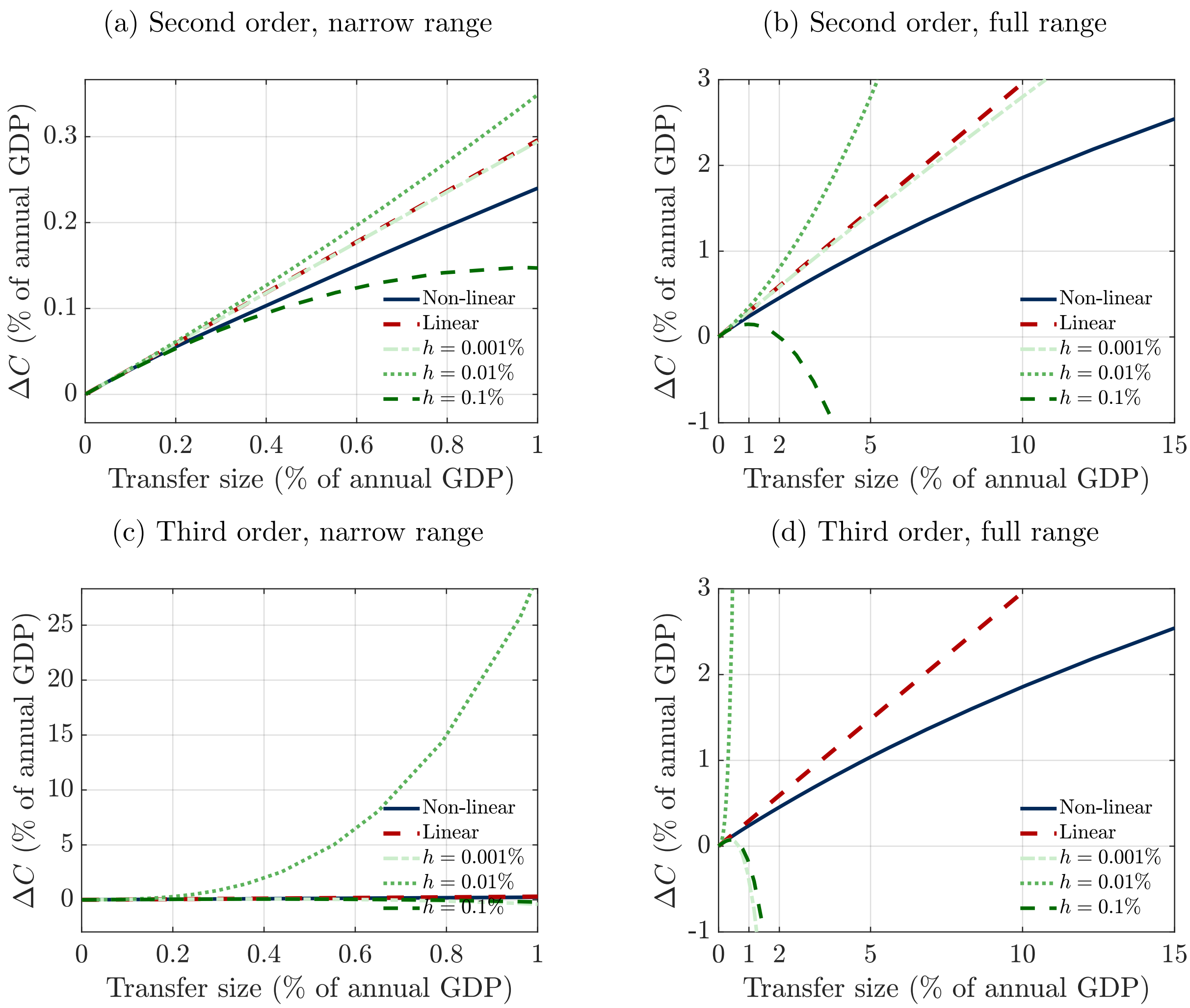
The non-linear equilibrium residuals for all $t \geq 0$ are

$$
\begin{aligned}
F_t^A(X; s) &= A_t(X; s) - B_t(X; s), \\
F_t^\pi(X; s) &= \kappa \left(v'(N_t) - \frac{1 - \tau}{C_t} \right) + \bar{\beta}\, \pi_{t+1} - \pi_t,
\end{aligned}
$$

Given the conjectured X, we solve backward for policy functions using EGM, propagate the distribution forward, compute asset demands and evaluate residuals. We update the conjectured sequences by a modified Newton method.

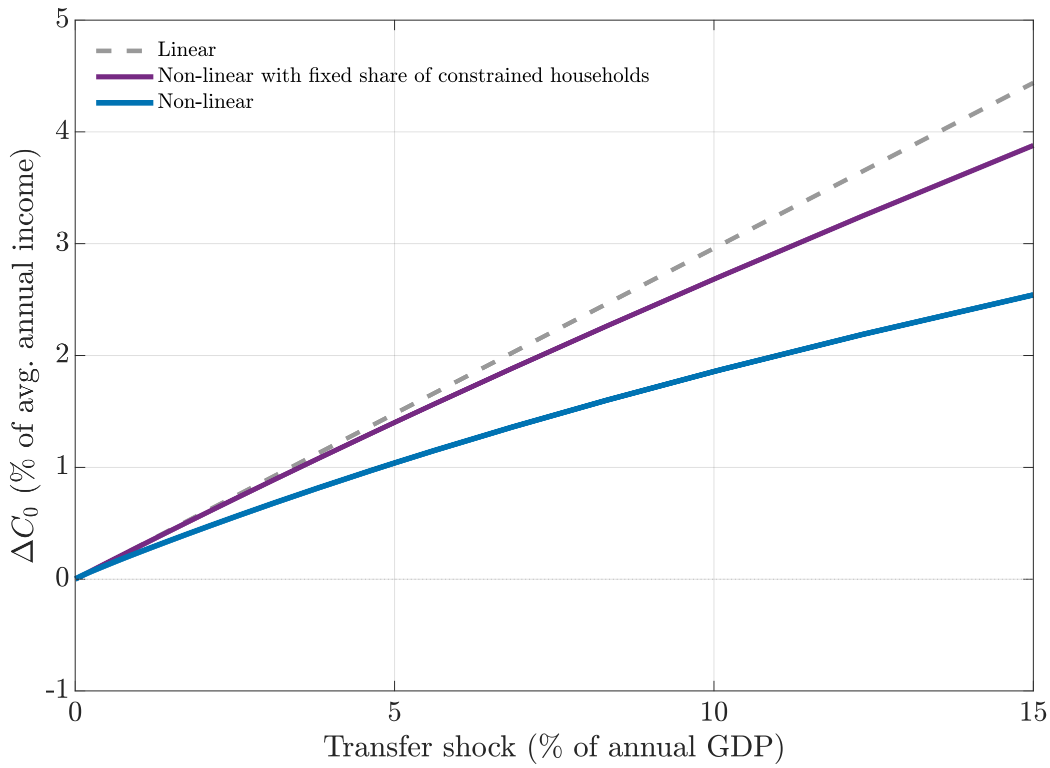
F Additional Figures: Partial Equilibrium

Figure F.1: Higher-order finite-difference approximations



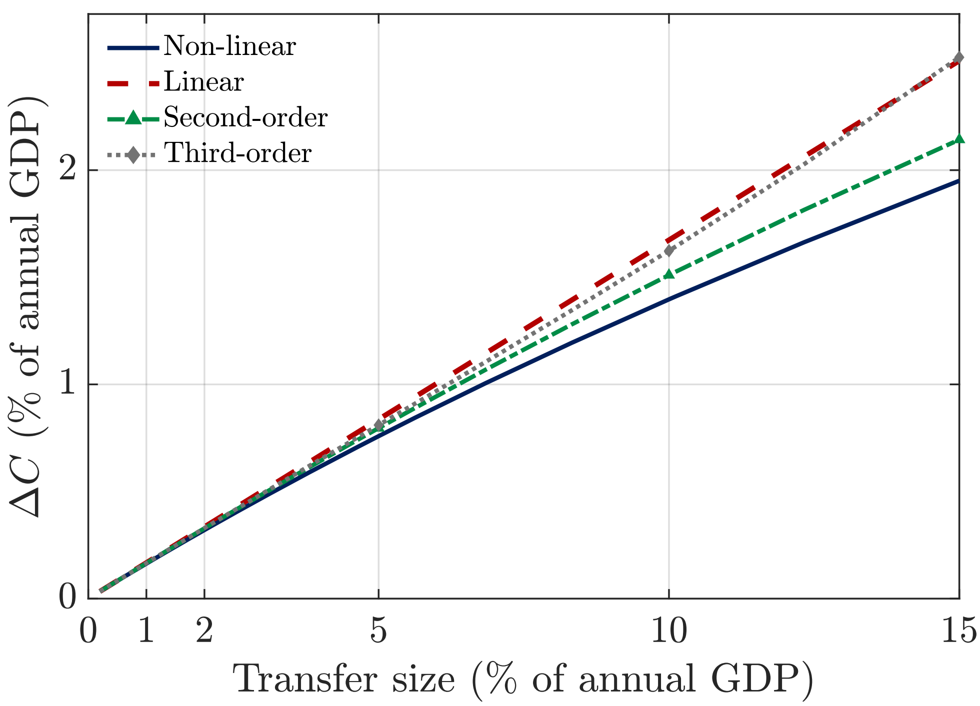
Note: Finite differences under different Richardson windows: the narrow range shown side by side with the full range (up to 15%, right column), for the second-order (top) and third-order (bottom) approximations. Each panel compares the non-linear PE response, the linear response, and local higher-order approximations using three different windows: $h = 0.001\%$, 0.01%, and 0.1% of annual GDP

Figure F.2: Extensive vs. Intensive Margin



Note: Non-linear with share-constrained fixed corresponds to the sum of the fully non-linear solution for unconstrained households and constrained households consumption, holding the shares fixed.

Figure F.3: Consumption for Unconstrained households



Note: Consumption response for initially unconstrained households, as a percentage of annual GDP per capita.

Figure F.4: Consumption function for low income

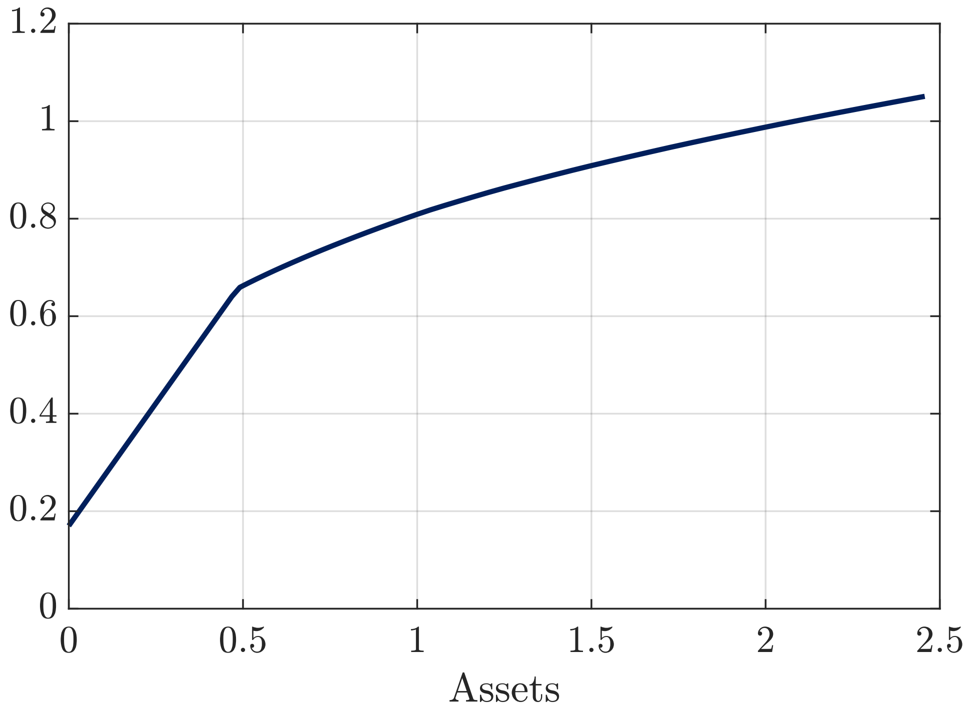


Figure F.5: Change in the Share of Constrained Households at $t = 0$

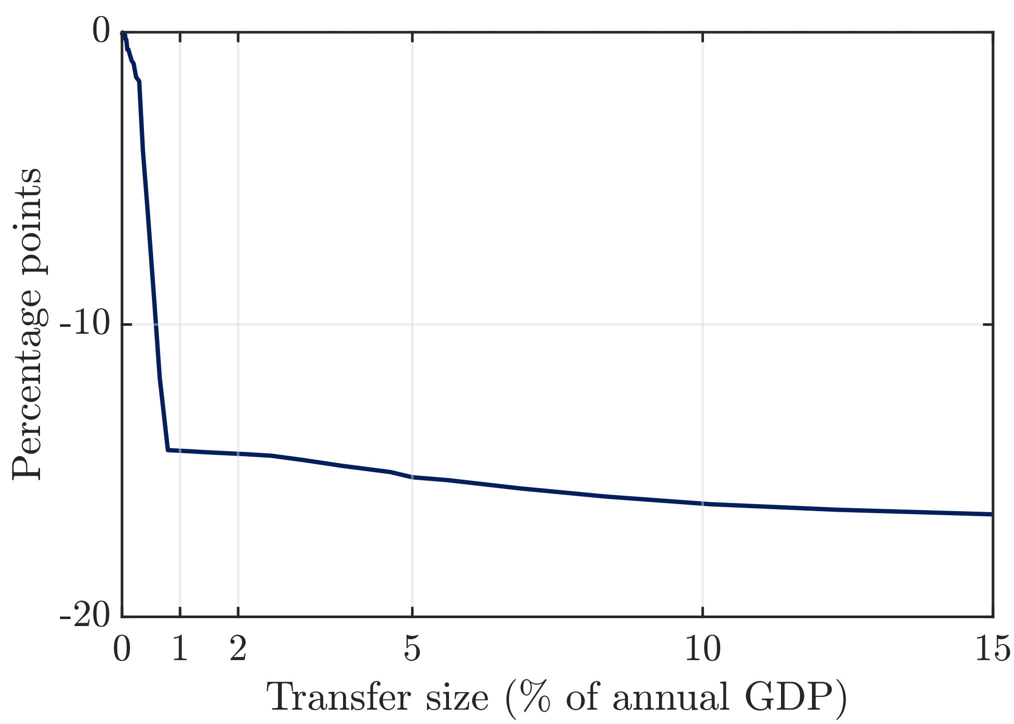
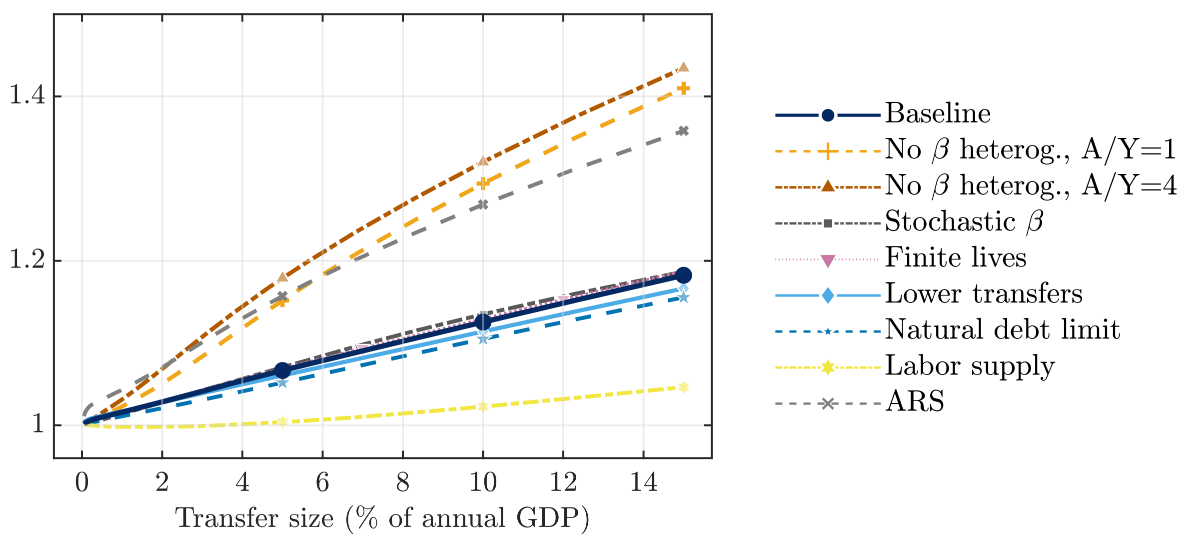
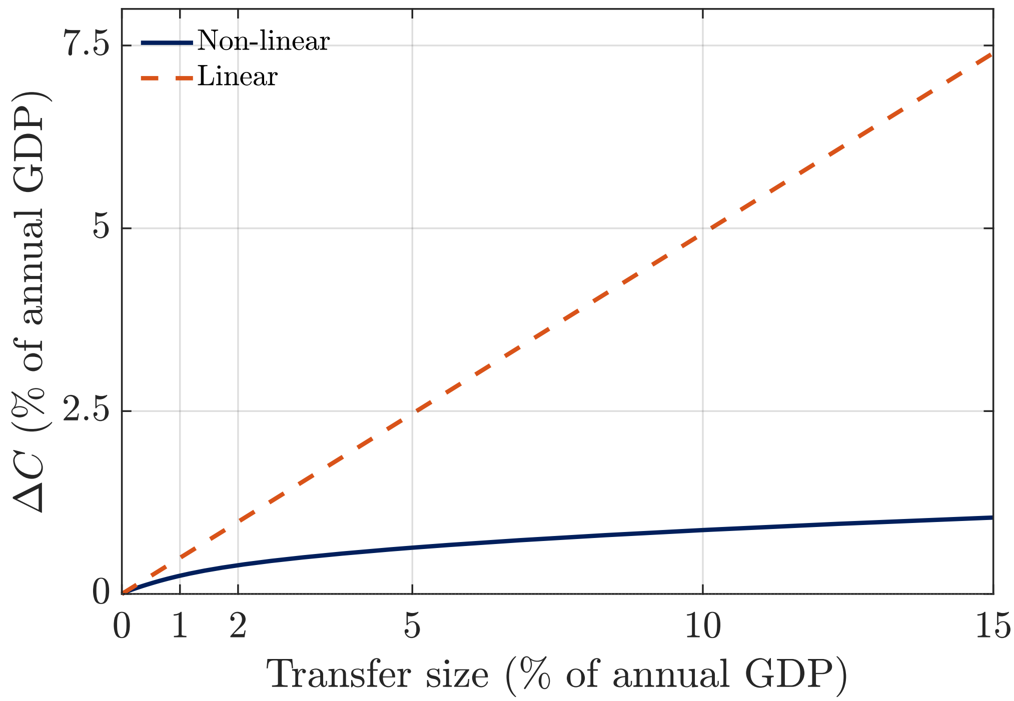


Figure F.6: First-year response: Linear to non-linear aggregate consumption response



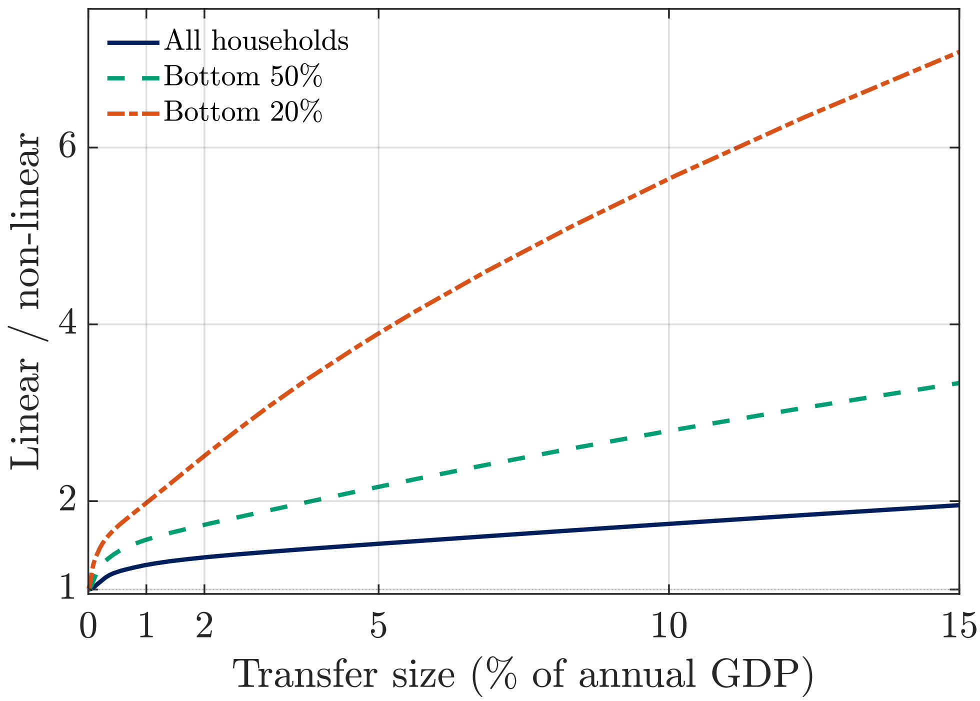
Note: The figure presents the ratio of the 12-month linear to non-linear consumption response in the partial equilibrium model.

Figure F.7: Targeted Transfers (Partial Equilibrium)



Note: partial-equilibrium consumption response in the baseline HANK model where transfers in period 0 are directed toward the 20% lowest-income households. Partial-equilibrium counterpart of Figure G.1.

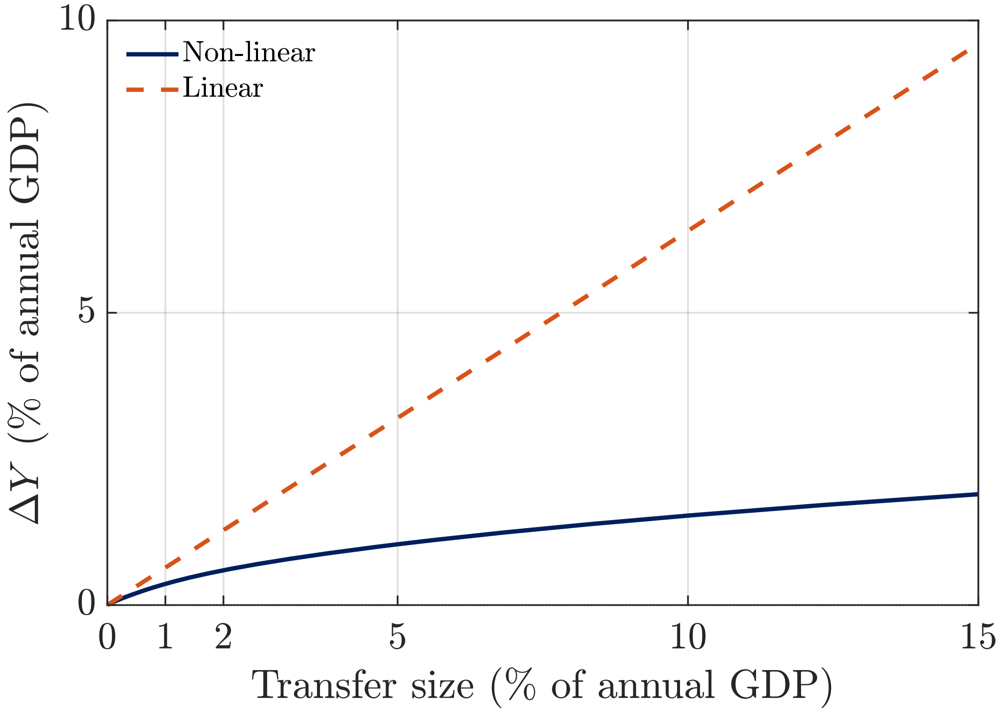
Figure F.8: Targeted Transfers



Note: ratio of the linear to the non-linear impact partial-equilibrium consumption response to a one-time transfer, as a function of transfer size, when the period-0 transfer is directed to all households, the bottom 50%, or the bottom 20% of the income distribution. This is the partial-equilibrium counterpart of Figure G.2.

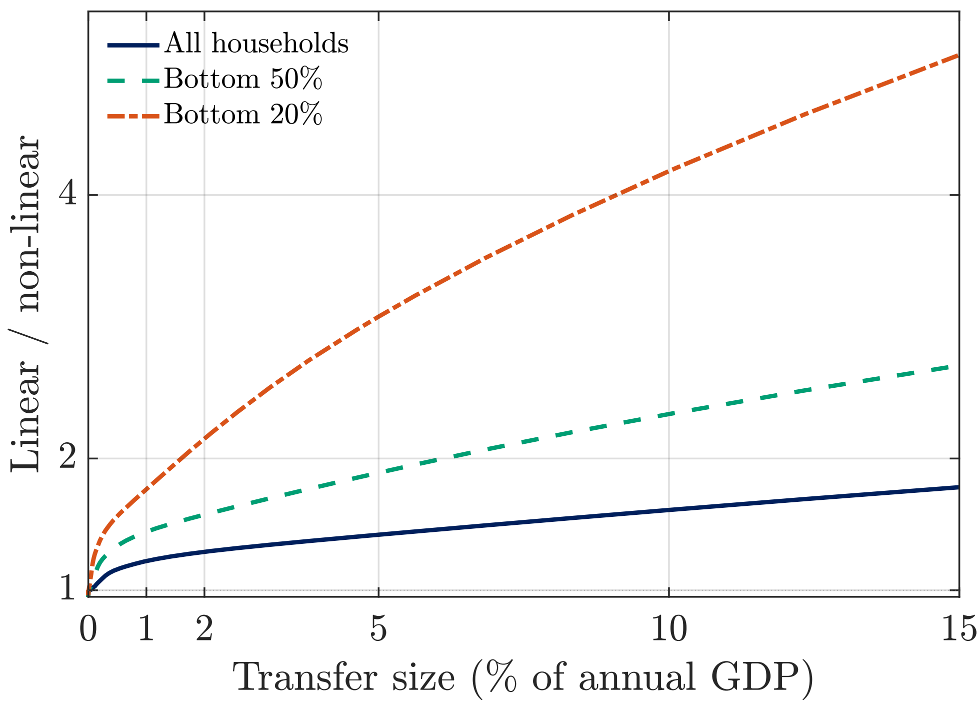
G Additional Figures: General Equilibrium

Figure G.1: Targeted Transfers



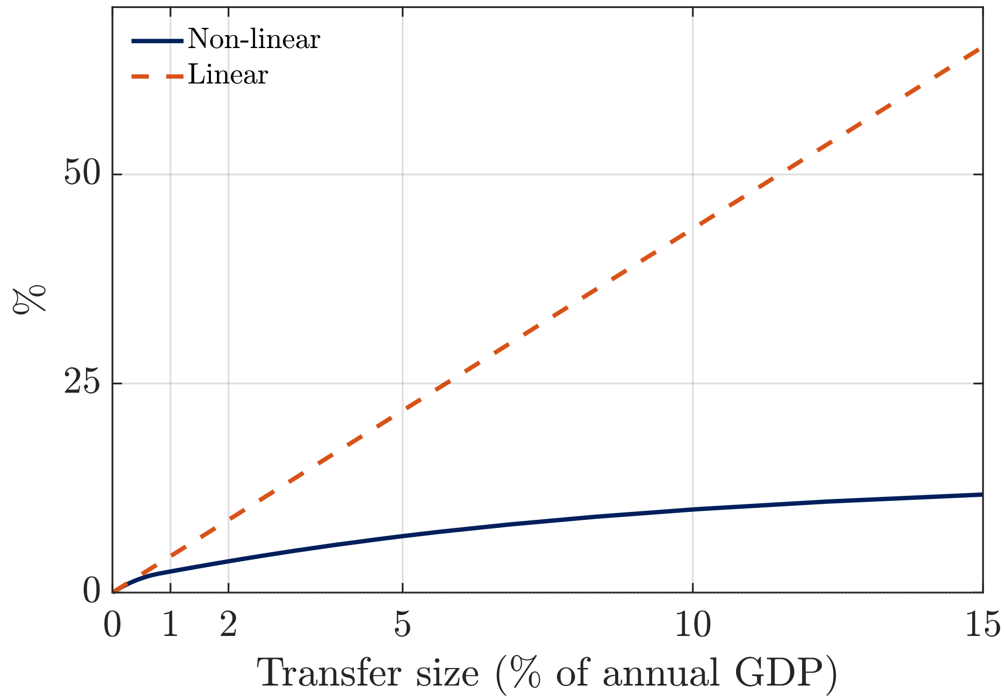
Note: output response in the baseline HANK model where transfers in period 0 are directed toward the 20% lowest-income households.

Figure G.2: Targeted vs Non-Targeted Transfers



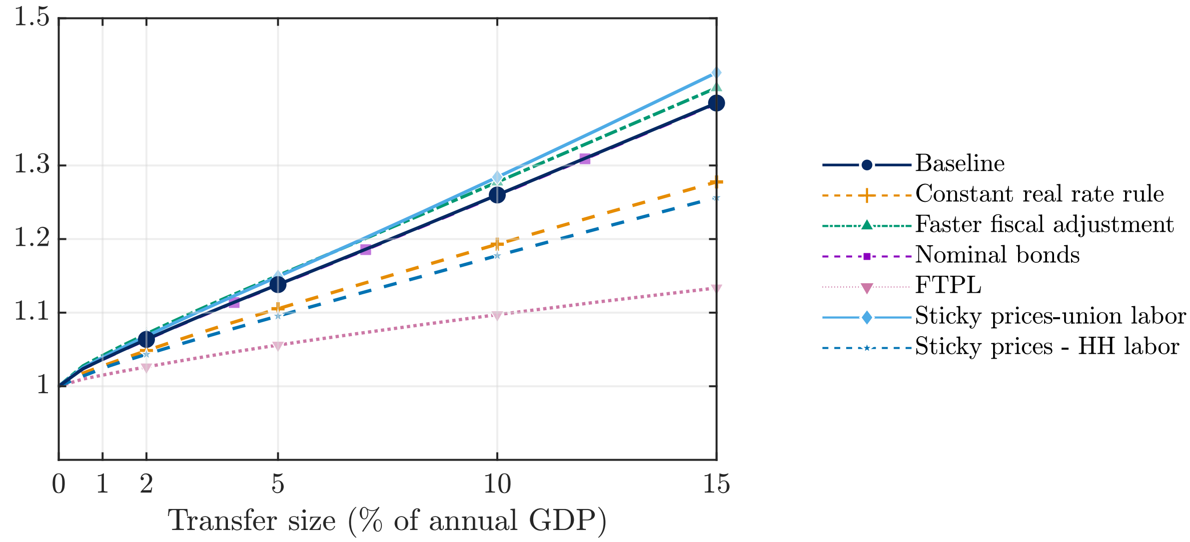
Note: The figure presents the ratio of the linear to the non-linear impact GE output response to a one-time transfer, as a function of transfer size, when the period-0 transfer is directed to all households, the bottom 50%, or the bottom 20% of the income distribution. Directing the transfer toward low-income (high-MPC) households sharpens the concavity of the response, so the linear approximation overstates the effect by more.

Figure G.3: Real Interest Rate Response in Flex-Price Economy



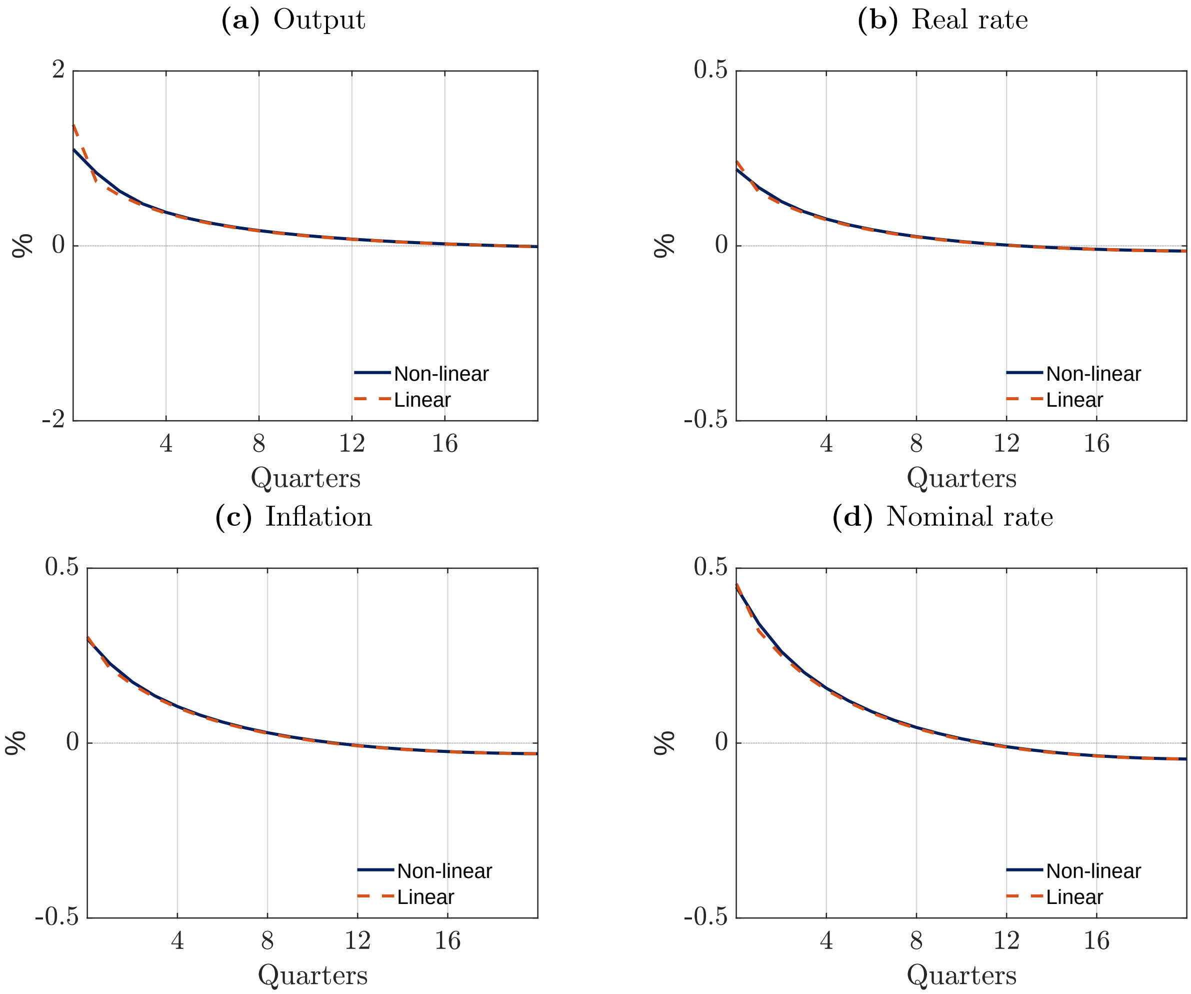
Note: real interest rate in economy with flexible prices and wages in which households choose labor.

Figure G.4: Sensitivity in HANK: Annual Responses



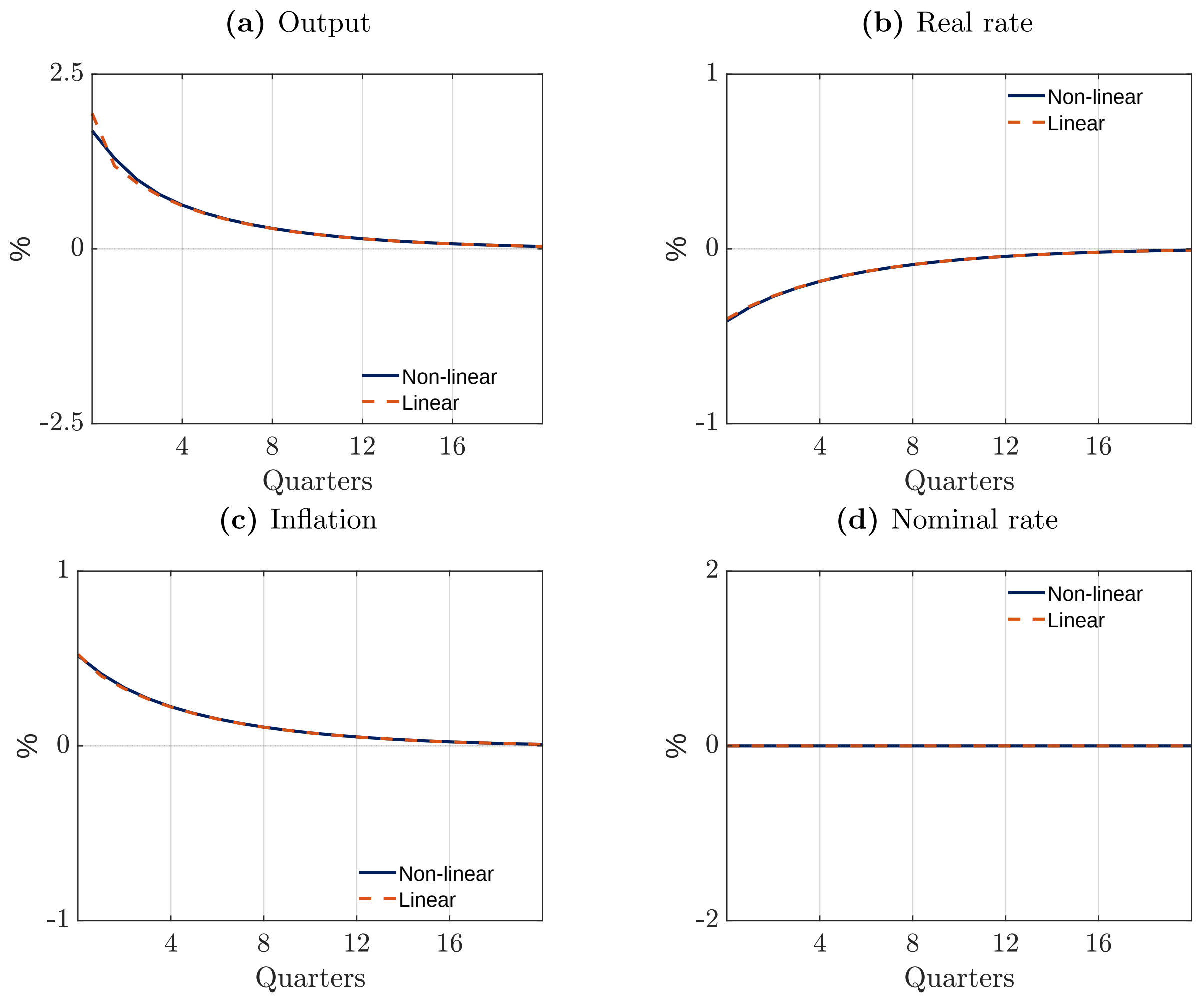
Note: The figure presents the ratio of the 12-month linear to non-linear output response in the general equilibrium model.

Figure G.5: Impulse responses baseline HANK



Note: output, real rate, inflation, and nominal rate for a 1% transfer. All variables are expressed as deviations from steady state. Output is expressed in percent. The real rate, inflation, and nominal rate are annualized and expressed in percent.

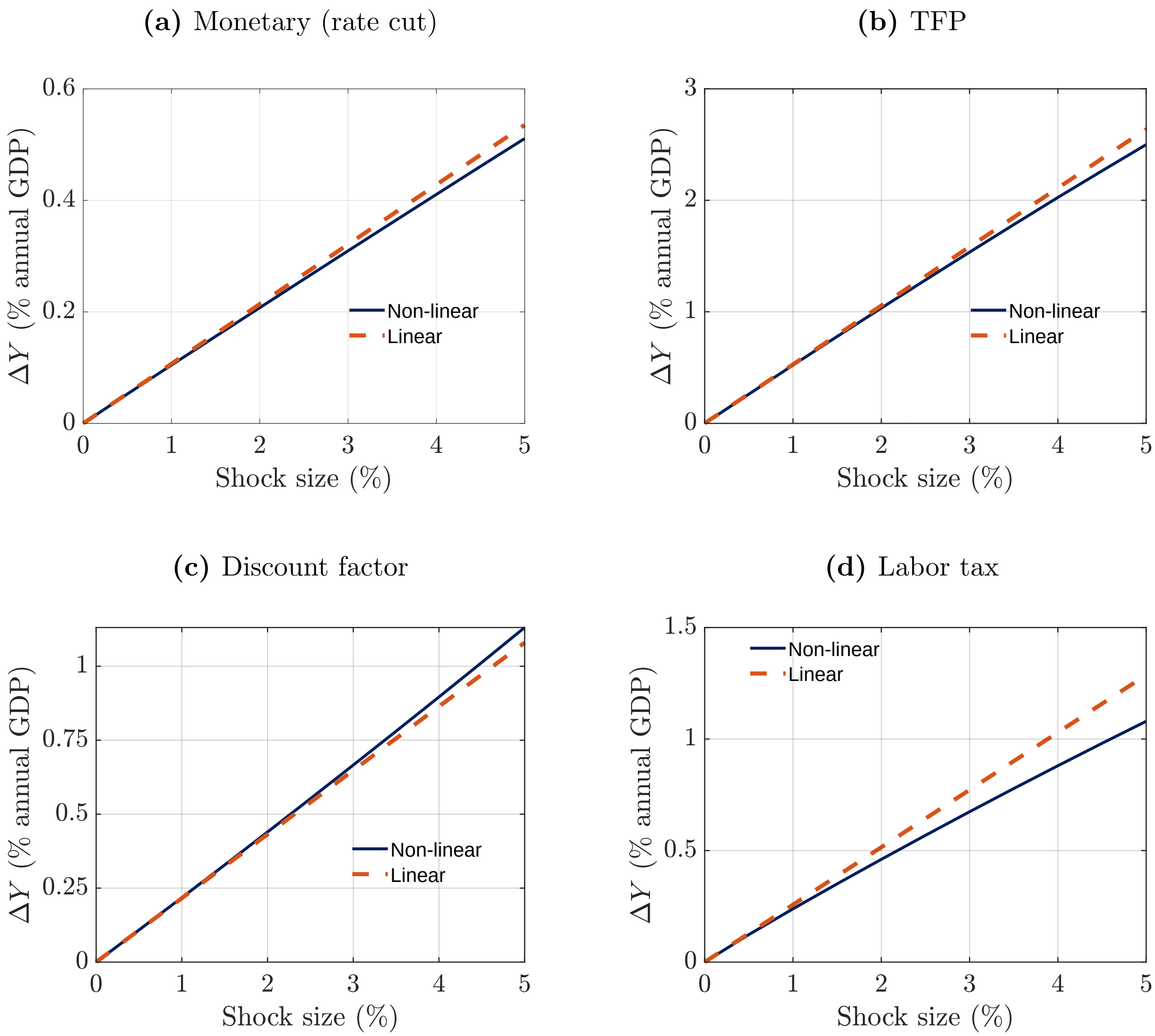
Figure G.6: Impulse responses: FTPL (unfunded fiscal stimulus)



Note: output, real rate, inflation, and nominal rate for a 1% transfer. All variables are expressed as deviations from steady state. Output is expressed in percent. The real rate, inflation, and nominal rate are annualized and expressed in percent.

H Other shocks

Figure H.1: Non-linearities in the output response across shocks.



Note: the figures present the output response as a function of shock size. Panel (a): monetary shock ε_i (pp annualized) such that $i_t = \bar{i} + \phi_\pi(\pi_t - \bar{\pi}) - \varepsilon_i$. Panel (b): productivity shock Z_0. Panel (c): discount factor shock $\tilde{\beta}_{i0} = \beta_{it}(1 - \varepsilon_\beta)$. Panel (d): labor tax shock $\tau_0 = \tau(1 - \varepsilon_\tau)$. Here ε_β and ε_τ represent shock size.