The Macroeconomic Effects of Excess Savings

Bence Bardóczy Jae Sim Andreas Tischbirek*

June 2024

Abstract

We study the consequences of shocks to the household wealth distribution in dynamic general equilibrium by characterizing the rate at which excess wealth is depleted. Analytical results link the aggregate decumulation rate to the distribution of the additional balances, micro intertemporal marginal propensities to consume, and general equilibrium feedback. A quantitative heterogeneous agent New Keynesian model matches the depletion path of the excess savings built up during the COVID-19 pandemic across the income distribution. The model predicts a substantial but steadily waning boost to consumption and explains about 40 percent of the surge in inflation observed in 2020 and 2021.

Keywords: Excess savings, heterogeneous agent New Keynesian (HANK) models, incomplete markets, household portfolios, inflation dynamics, COVID-19 pandemic

JEL Classification: E21, E31, E32, E52

*Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue NW, Washington, DC 20551
We thank Stephanie Aaronson, Hess Chung, Sebastian Graves, Matteo Iacoviello, Antoine Lepetit, Gaston Navarro, Matthias Paustian, Rodolfo Rigato (discussant), and participants at the Annual Meeting of the AEA in San Antonio, the North American Summer Meeting of the Econometric Society in Nashville, and various seminars for valuable comments and suggestions. We also thank Aditya Aladangady, David Cho, Laura Feiveson, and Eugenio Pinto for sharing updated estimates of excess savings with us. The analysis and conclusions set forth in this paper are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors.
Bardóczy: bence.a.bardoczy@frb.gov
Sim: jae.w.sim@frb.gov
Tischbirek: andreas.j.tischbirek@frb.gov

1 Introduction

Household wealth fluctuates substantially over the business cycle, with an annual growth rate that is more volatile than that of nominal GDP in the U.S. over the post-war period. While fluctuations in wealth are traditionally viewed as a side effect of business cycle fluctuations, they may, in fact, contribute to generate them. The propensity to build up excess savings—wealth in excess of its level in normal times—to smooth consumption differs substantially across households. The aggregate depletion path similarly masks a vast amount of heterogeneity. Our interest lies in characterizing and quantifying such a depletion of excess savings and its macroeconomic implications. The pandemic-era excess savings provide a good case study because estimates of the distribution of excess savings are readily available.[1]

During the COVID-19 pandemic, U.S. households drastically reduced spending and received fiscal transfers of unprecedented size. Despite earnings losses, they accumulated excess savings estimated to amount to about 10 percent of pre-crisis GDP at their peak. The size and liquidity of the excess balances raised concerns that inflation could surge if they were spent quickly, shedding light on the inconclusiveness of model predictions about their decumulation. The large contribution of fiscal transfers to the buildup of excess savings may prompt two contrasting assessments of the decumulation rate. First, repeatedly applying a sizable marginal propensity to consume (MPC) in the spirit of classical Keynesian economics yields the prediction that excess savings are spent down rapidly. Second, consulting standard macroeconomic models with a representative agent leads to the conclusion that excess savings may never be depleted, as excess savings are the counterpart of the excess debt used to finance fiscal transfers. In line with Ricardian equivalence, the household refrains from consuming out of excess savings, maintaining them until the government raises taxes to repay its debt.

We analyze the mechanisms underlying the depletion of excess savings by relying on a heterogeneous agent New Keynesian (HANK) model that encompasses both Keynesian and Ricardian household behavior. A key theme of our analysis is that the distribution of excess savings has a strong impact on their aggregate depletion path because households with less liquid wealth are more Keynesian and households with more liquid wealth are more Ricardian. We clarify that, under general conditions, the partial equilibrium impulse response to a shock to household wealth is fully determined by the households' intertemporal marginal propensities to consume (iMPCs) and the distribution of the excess wealth.[2] The iMPCs also affect the amplification of spending out of excess savings but cease being sufficient to characterize the response in general equilibrium. In addition, we construct a quantitative medium-scale HANK model that we calibrate using these insights and apply to evaluate the effects of the excess

[1] Aladangady et al. (2022) estimate excess savings by income quartile in the U.S. The estimate is constructed by cumulating the flow deviations of household saving in the National Income and Product Accounts from a pre-crisis trend. It reflects changes in income and spending and excludes valuation effects.

[2] The term "iMPC" was introduced by Auclert et al. (2018) to describe the sequence-space Jacobian of an aggregate consumption function with respect to income.

savings built up during the COVID-19 pandemic.[3]

We begin by analyzing excess savings depletion following an arbitrary shock to the wealth distribution in a small-scale HANK model.[4] Micro iMPCs reflect a household's expected consumption gain a given number of periods after receiving a marginal unit of cash on hand. We demonstrate analytically that in partial equilibrium—that is, conditional on a path of the real interest rate—the responses of aggregate consumption and saving are fully determined by the joint distribution of initial excess savings and micro iMPCs. As this result is independent from the sources of the initial excess wealth, empirical work concerned with measuring iMPCs out of unexpected income is informative for the depletion of excess savings.[5] In general equilibrium, the iMPCs also affect the multiplier on spending out of excess savings but they are not sufficient to characterize aggregate consumption, as in the case of government spending (Auclert et al., 2018). For example, spending is inflationary, which results in feedback through adjustments of the real interest rate. The multiplier therefore depends not only on the iMPCs but also on the slope of the Phillips curve, the sensitivity of the real interest rate to inflation, and the intertemporal consumption response to changes in the interest rate path.

Our quantitative analysis relies on a medium-scale HANK model of the U.S. economy. It builds on the analytical results regarding the importance of iMPCs for the depletion of excess savings in two ways. First, we discipline the partial equilibrium consumption-saving behavior in the model through a calibration procedure that directly ties the iMPCs to empirical estimates. Second, the model includes a rich set of components that gives rise to realistic general-equilibrium forces beyond those implied by the iMPCs alone. These components include involuntary unemployment resulting from Diamond–Mortensen–Pissarides searching and matching frictions, Nash wage bargaining with wage adjustment frictions, model-consistent pricing of claims to all profits generated in the economy and long-term government debt allowing for asset revaluation effects, monetary policy implemented through a Taylor-type rule, and the dominance of debt financing with slowly adjusting distortionary taxation.

With the calibrated quantitative model in hand, we estimate the isolated consequences of excess savings in the wake of the COVID-19 pandemic. The baseline simulation includes an accumulation period followed by a decumulation period. The former ranges from the beginning of the pandemic until the third quarter of 2021, when the peak of the aggregate excess savings stock is estimated to have occurred. During the accumulation period, the model households are confronted with shocks that allow the model to replicate realized aggregate consumption and estimates of the excess savings accumulated by each quartile of the income distribution. After the peak in excess savings is reached, we restrict all shocks to zero and study the model

[3] The model described in this paper lays the foundations for the Federal Reserve Board's HANK framework ("FR-HANK") used in quantitative policy analysis.

[4] A number of contributions extend the standard Bewley–Huggett–Aiyagari incomplete markets model with nominal price-setting frictions. Early examples include Oh and Reis (2012), McKay et al. (2016), McKay and Reis (2016), Guerrieri and Lorenzoni (2017), Kaplan et al. (2018), and Gornemann et al. (2021), among others.

[5] For example, Fagereng et al. (2021) estimate iMPCs based on income from lottery prizes.

predictions about the decumulation period.

Key findings from the quantitative exercise include the following. First, the model closely matches the path of excess savings, on aggregate and quartile by quartile, over the part of the decumulation period for which empirical estimates are available. Second, the bottom quartile exhausts its excess savings first and the top quartile last. However, even the top households deplete their excess savings within about three years, implying that they are not fully Ricardian. Third, the corresponding increase in demand explains about 40 percent of the surge in inflation that occurred between the first half of 2020 and the second half of 2021, showing that post-COVID inflation was not exclusively a result of supply constraints. Moreover, by historical standards, the Federal Reserve underreacted to the inflation pressure that resulted from spending out of excess savings. Fourth, the excess savings path over the decumulation period is well predicted by the determinants of the partial equilibrium forces, the iMPCs and the initial allocation of excess savings. This is in contrast to the responses to a monetary policy shock, for example, which are dominated by indirect effects (Kaplan et al., 2018).

Literature. Our analysis is related to work on the role of household wealth in business cycle fluctuations and the recent literature on quantitative HANK models.

The large drop in household net worth seen in the U.S. during the Global Financial Crisis of 2007 to 2009 sparked a series of contributions to the literature on the implications of changes in household wealth. Mian et al. (2013) use ZIP code–level data to estimate the elasticity of consumption to housing wealth for the crisis period. They find a sizable average MPC out of housing net worth that declines with household income and leverage. In addition, Mian and Sufi (2014) show that employment contracted more strongly in counties with a larger decline in housing wealth, which is indicative of general equilibrium effects set off by the spending response. Further results on the wealth effects of housing are contained in Kaplan et al. (2020a) and Guren et al. (2021). Evidence that a deterioration of not only housing but also financial wealth causes spending to adjust is presented by Christelis et al. (2015), and Heathcote and Perri (2018) link a low valuation of household assets to volatility from equilibrium multiplicity. In contrast to these papers and in line with the excess savings built up during the pandemic, we study spending out of highly liquid assets rather than revaluations of illiquid housing or financial wealth. In addition, our analysis is based on a HANK model, which allows us to evaluate the influence of distributional and general equilibrium effects along the path of wealth depletion.

The focus on excess savings is shared by Auclert et al. (2023b), who use a stylized model to argue that excess savings have prolonged effects on aggregate demand. A large part of a dollar initially held by a poorer household with a higher MPC becomes income for wealthier households with lower MPCs, who spend it again—a "trickling up" that is repeated until the dollar lands in the hands of the ultrarich. We contribute to this insight by giving an analytical characterization of excess savings depletion based on optimal consumption-saving behavior, constructing a medium-scale model, and conducting a quantitative analysis of the implications of excess savings in the aftermath of the pandemic, accounting for a rich set of general equilibrium forces.

Our quantitative HANK model of the U.S. economy is most similar to the models of Auclert et al. (2020) and Bayer et al. (2024). In contrast to both models, ours allows for adjustments at the extensive labor margin by including searching and matching frictions coupled with wage bargaining. Following the former, we model a financial intermediary that prices all profits generated in the economy, but we allow the intermediary to build up net worth. Distributions of retained earnings by the intermediary are comparable to distributions from the illiquid account in their framework—from which we abstract. In the latter, households can optimally adjust an illiquid account with a Calvo-type probability, but monopoly profits are collected by entrepreneurs, which precludes equity revaluation effects that are important for our results.

While they are not concerned with excess savings per se, Carroll et al. (2021) and Bayer et al. (2023) use quantitative models to study the effects of the CARES Act passed in March 2020, which contributed to the buildup of excess savings. We regard these analyses of fiscal measures put in place at the onset of the pandemic as complementary to ours. Finally, a large number of papers seek to answer questions that are specific to the COVID-19 pandemic—for example, by developing models with economic and epidemiological features.[6] Our work is also tangentially related to this stream of the literature, but our interest lies in the implications of excess wealth decumulation more broadly.

Overview. The remainder of the paper is organized as follows. Section 2 draws on a stylized HANK model to provide analytical results about the dynamics of excess savings in partial and general equilibrium. Section 3 lays out the quantitative model. Section 4 explains our calibration strategy, which is informed by the analytical results. Section 5 studies the macroeconomic effects of excess savings in the aftermath of the COVID-19 pandemic through the lens of our model. A final section concludes.

2 Inspecting Excess Savings Depletion

In this section, we put excess savings in the context of workhorse macroeconomic models. We make two observations that will guide our quantitative exercise. First, in partial equilibrium, the joint distribution of iMPCs and initial excess savings is sufficient to characterize the speed of depletion. This result holds irrespective of the original cause of excess savings in a broad and relevant class of consumption-savings models. Second, in general equilibrium, iMPCs also affect the multiplier but are not sufficient anymore. Excess savings decumulation is an aggregate demand shock that puts upward pressure on prices. The slope of the Phillips curve

[6] See Kaplan et al. (2020b) and Eichenbaum et al. (2021), among others.

and accommodation on behalf of fiscal and monetary policy makers thus become relevant to determining the ultimate impact on the macroeconomy.

2.1 Partial Equilibrium

We start by considering the standard incomplete markets (SIM) model, the workhorse model of heterogeneous agent macroeconomics. We use the SIM model to give a formal definition of iMPCs at the micro level and argue that they fully characterize the excess savings decumulation process, conditional on the path of real interest rates.[7] This result holds in a broader class of models that includes the standard representative agent and spender-saver models.

SIM Model. There is a unit mass of households indexed by $i \in [0, 1]$ who face idiosyncratic income risk and trade in a single, non-state contingent asset. The Bellman equation is

$$V_t(y_{i,t}, a_{i,t-1}) = \max_{c_{i,t}, a_{i,t}} u(c_{i,t}) + \beta \mathbb{E}_t \left[V_{t+1}(y_{i,t+1}, a_{i,t}) \right], \tag{1}$$

$$\text{s.t. } c_{i,t} + a_{i,t} = (1 + r_{t-1})a_{i,t-1} + y_{i,t}, \tag{2}$$

$$a_{i,t} \geq \underline{a}, \tag{3}$$

where $y_{i,t}$ is idiosyncratic income that follows an exogenous Markov process, $a_{i,t-1}$ are assets at the end of the last period, $u : \mathbb{R} \to \mathbb{R}$ is a standard period utility function that satisfies the Inada conditions, $\beta \in (0, 1)$ is the discount factor, and r_t is the real return on assets.

Although distinguishing income and assets as separate state variables is often useful, it masks a well-known property of the SIM model that is crucial for thinking about excess savings. Conditional on cash-on-hand $x_{i,t} = (1 + r_{t-1})a_{i,t-1} + y_{i,t}$ and its forecasts $\mathbb{E}_t[x_{i,t+h}]$ for all $h > 1$, fluctuations in the components of cash on hand are irrelevant for households' decisions. This implies that entering the period with excess savings $a_{i,t-1} + \Delta$ has the same implications as receiving income $y_{i,t} + (1 + r_{t-1})\Delta$. This is a key observation that means that the extensive empirical evidence on spending responses to lump-sum transfers is directly applicable to the depletion of excess savings, irrespective of the initial cause of excess savings.

Excess Savings and iMPCs. Consider household i with initial cash-on-hand $x_{i,0}$ who receives a one-time transfer in period 0. Our goal is to compare its consumption-saving choices for all $t \geq 0$ with the counterfactual with no transfer. The use of a one-time transfer to generate excess savings is without loss of generality. It is immediate from the Bellman equation (1)–(3) that the past only matters through assets $a_{i,t-1}$. Therefore, any shocks that generate the same initial cash-on-hand $x_{i,0}$ will have the same implications for $t \geq 0$.

[7] Auclert et al. (2018) define iMPCs as derivatives of the aggregate consumption function. These macro iMPCs are equal to the population average of micro iMPCs as we define them in this section.

Let us fix a path of real interest rates $\{r_t\}_{t=0}^{\infty}$ of which agents have perfect foresight. Solving the Bellman equation (1)–(3) yields policy functions $c_t(x_{i,t})$ and $a_t(x_{i,t})$, where the time subscript t marks the dependence on the interest rate path.

Definition. *The micro iMPCs of a household with initial state $x_{i,0}$ are defined as the change in the expected consumption path for all $t \geq 0$ in response to an infinitesimal increase in cash on hand in period 0. Formally,*

$$m_t(x_{i,0}) \equiv \lim_{\Delta \to 0} \frac{\mathbb{E}\left[c_t(x_{i,t}(\Delta))|y_{i,0}\right] - \mathbb{E}\left[c_t(x_{i,t})|y_{i,0}\right]}{\Delta}, \tag{4}$$

where $x_{i,t}(\Delta)$ is cash on hand in period t following a transfer Δ in period 0.

The following proposition shows that iMPCs are effectively a direct measure of excess savings depletion over time, taking into account the return on savings as well as the idiosyncratic income shocks that may lead households to tap into their excess savings.

Proposition. *The micro iMPCs of household i with initial state $x_{i,0}$ satisfy*

$$m_t(x_{i,0}) = \mathbb{E}\left[c_t'(x_{i,t}) \prod_{s=0}^{t-1}(1+r_s)\Big(1 - c_s'(x_{i,s})\Big)\Big|y_{i,0}\right]. \tag{5}$$

See Appendix A.1 for a formal proof.

The impact iMPC, $m_0(x_{i,0}) = c_0'(x_{i,0})$, is equal to the slope of the consumption function at the initial state. This is the standard *static MPC*, which shows the fraction that household i spends of an incoming transfer. Subsequent iMPCs are the product of two terms. The first term, $c_t'(x_{i,t})$, is the static MPC at time t. The second term, $\prod_{s=0}^{t-1}(1+r_s)\,[1 - c_s'(x_{i,s})]$, is the cumulative return on the unspent fraction of the initial transfer—i.e., *excess savings*. That is, excess savings are determined by the histories of interest rates and of the static MPCs of household i.

The proposition shows that one can trace out the paths of consumption and excess savings at the household level by repeatedly applying the static micro MPC. This approach is reminiscent of traditional Keynesian economics. In contrast to traditional Keynesian analysis, however, the static MPC is not a primitive of the aggregate consumption function but an endogenous object that reflects utility maximization by individual households. Micro MPCs may evolve over time because of income shocks and planned saving. For example, households who are hit by a large negative income shock may plan to run down their savings to the borrowing limit. Conversely, households who are hit by a large positive shock may embark on a period of wealth accumulation. The result is that MPCs are heterogeneous across households at any given time.

The key takeaway from the proposition is that MPC heterogeneity carries over into heterogeneity in excess savings decumulation. To consider specific examples, a hand-to-mouth household with $m_0 = 1$ does not accumulate excess savings; at the other extreme, rich households with small MPCs are expected to deplete excess savings slowly. In the representative agent limit with $y_{i,t} \equiv Y_t$ and $\underline{a} = \infty$, it can be shown that $m_0 = 1 - \beta$, a very small number in typical calibrations. The SIM model is useful because it captures a range of MPCs between these two extremes.

Figure 1 visualizes the proposition by plotting $m_t(x_{i,0})$, aggregated to cash-on-hand quartiles, from our quantitative model. Panel A shows the iMPCs, while panel B shows the implied excess savings paths. When interpreting the figure, it is useful to keep in mind that the cumulative iMPC is 1 for everybody. All households have one dollar to spend—the question is how quickly they spend it. As expected, cash on hand is a strong predictor of the spending path. The bottom 75 percent of households exhaust most of their excess savings within three years, and virtually all in five years. In contrast, the top 25 percent of households display a much more gradual spending profile, retaining more than a third of excess savings after five years. In sum, the distribution of excess savings is crucial for the speed of decumulation and the implied increase in aggregate demand.

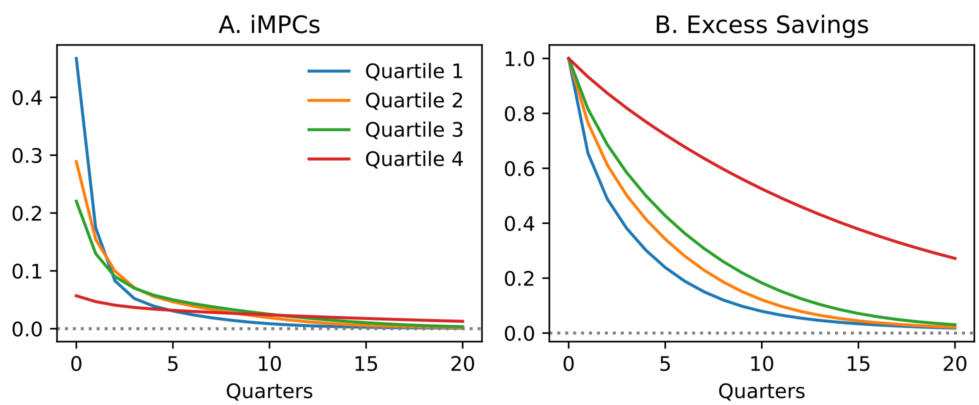

Figure 1: Decumulation of Excess Savings in Partial Equilibrium

Beyond the SIM Model. The insight that iMPCs capture excess savings depletion relies on two assumptions that are implicit in the SIM model.

Assumption 1. *Households make two choices in every period: consumption $c > 0$ and assets $a \in \mathcal{A}$, where $\mathcal{A} \subseteq \mathbb{R}$ is a compact subset of the real numbers that may depend on other exogenous states.*

Assumption 2. *Consumption is not a state variable in individual households' dynamic program.*

Prominent models that also satisfy these assumptions include the standard representative agent model, in which $y_{i,t}$ is the same for all households and $\underline{a} = -\infty$, and the spender-saver model (Campbell and Mankiw, 1989), in which a fraction μ of households are hand-to-mouth, with an asset choice set $\mathcal{A} = \{0\}$, while $1 - \mu$ households behave just like in the representative agent model. When it comes to excess savings, the predictions of these two tractable models are the same. Excess savings have to be concentrated in the hands of permanent income households who, in the absence of changes in prices, use their excess savings to finance a small, permanent increase in their consumption and never exhaust their excess savings. This prediction differs sharply from the gradual decumulation implied by the SIM model.

How restrictive are Assumptions 1 and 2? Assumption 1 rules out models in which households make labor supply choices (e.g., Aiyagari and McGrattan 1998) or portfolio choices (e.g., Kaplan et al. 2018). If households can spend their excess savings on both goods and leisure, income effects on labor supply become relevant for shaping the excess savings decumulation process in addition to iMPCs. We shut down labor supply because there is a consensus that income effects on labor supply are small for most households.[8] If households can save in multiple assets, the distribution of excess savings across asset classes and the marginal propensities to save in those asset classes become relevant. We consider portfolio choice an interesting generalization but abstract from it because during the COVID-19 pandemic U.S households accumulated excess savings almost exclusively in liquid assets.[9]

Assumption 2 ensures that we do not have to keep track of the indirect effect of excess savings through lagged consumption. This rules out, for example, models of habit formation. The assumption simplifies the exposition of the proposition but does not affect the sufficiency of iMPCs.

2.2 General Equilibrium

Next, we embed the SIM model in a small-scale New Keynesian model to highlight the indirect effects of excess savings decumulation in general equilibrium. We do so by analyzing the intertemporal Keynesian cross (IKC) implied by the model, which allows for a sharp decomposition between the direct and indirect effects of excess savings depletion. The main point is that iMPCs are an important determinant of the general equilibrium channels as well, though not sufficient as they are for the direct effect. We provide a brief summary of the model in the main text. The derivation of the IKC is relegated to Appendix A.2.[10]

Model Summary. Households solve (1)–(3) with income being $y_{i,t} = (1 - \tau_t) Y_t z_{i,t}$, where τ_t is the tax rate, Y_t is aggregate income, and $z_{i,t}$ is idiosyncratic productivity. The household block can be represented by an aggregate consumption function $\{C_t\}_{t=0}^{\infty} = \mathcal{C}(\{\tau_t, r_t, Y_t\}_{t=0}^{\infty}, \Gamma_0)$ that maps the sequences of tax rates, real interest rates, income, and the initial distribution of households Γ_0 into a sequence of aggregate consumption. In this setting, excess savings are equivalent to perturbations of the initial (wealth) distribution Γ_0. The counterpart of excess savings held by households is government debt. We assume that the fiscal authority implements a path for the tax rate $\{\tau_t\}_{t=0}^{\infty}$ such that government debt B_t matches aggregate savings

[8] In addition, Auclert et al. (2023a) point out that matching small but positive income effects on labor supply in conjunction with high MPCs is problematic in New Keynesian models. Solving these problems require introducing labor supply frictions, exploring which is beyond the scope of this paper.

[9] We review empirical evidence on excess savings during the COVID-19 pandemic in Section 5.1.

[10] For a comprehensive introduction to macroeconomic analysis in sequence space, see the seminal papers Auclert et al. (2018) and Auclert et al. (2021).

in period 0 but eventually returns to its steady-state value. Inflation follows a standard Phillips curve represented by $\{\pi_t\}_{t=0}^{\infty} = \mathcal{K}(\{Y_t\}_{t=0}^{\infty})$. The central bank sets the nominal interest rate i_t according to an inflation-targeting rule. Given the Fisher equation $r_t = i_t - \mathbb{E}_t[\log(\Pi_{t+1})]$, we can represent the Taylor rule as $\{r_t\}_{t=0}^{\infty} = \mathcal{R}(\{\pi_t\}_{t=0}^{\infty})$. In equilibrium, $Y_t = C_t$, and the asset market clears by Walras' law.

Intertemporal Keynesian Cross. For the purposes of this section, it is convenient to abbreviate sequences as bold-faced letters. Conditional on the initial distribution Γ_0, the equilibrium sequence of output solves

$$\mathbf{Y} - \mathcal{C}(\boldsymbol{\tau}, \mathcal{R}(\mathcal{K}(\mathbf{Y})), \mathbf{Y}, \Gamma_0) = 0, \tag{6}$$

where we expressed the sequence of real rates entering the household block as a function of output using the Phillips curve and the monetary policy rule. Differentiating (6), we obtain

$$d\mathbf{Y} = \underbrace{(I - \mathcal{C}_Y - \mathcal{C}_R \mathcal{R}_\pi \mathcal{K}_Y)^{-1}}_{\text{multiplier}} \Big(\underbrace{\mathcal{C}_\Gamma d\Gamma_0}_{\text{direct effect of ES}} + \underbrace{\mathcal{C}_\tau d\boldsymbol{\tau}}_{\text{fiscal reaction to ES}} \Big). \tag{7}$$

As we discussed in the previous section, the direct effect of excess savings on aggregate consumption depends on their distribution $d\Gamma_0$. The Jacobian $\mathcal{C}_\Gamma$ is an aggregator of micro iMPCs with an arbitrary distribution. Indirect effects come in two forms. First, the path of taxes that support the initial stock of excess savings and ensure that government bonds (and thus aggregate savings) eventually return to steady state impacts households via the Jacobian $\mathcal{C}_\tau$. Taxes eventually have to rise to bring down government debt. Households understand this and may respond by lowering their consumption when, or even before, higher taxes materialize. Second, there is a multiplier that captures the fact that one household's spending is another household's income via $\mathcal{C}_Y$ and the effect of the endogenous response of the real interest rate via $\mathcal{C}_R$.

Notably, $\mathcal{C}_Y$ is also an aggregator of micro iMPCs. It differs from $\mathcal{C}_\Gamma$ in that it weighs individuals specifically by their productivity (reflecting $y_{i,t} \propto Y_t z_{i,t}$) and in that it includes responses to anticipated rises in future income.[11] If prices are completely rigid ($\mathcal{K}_Y = 0$), the multiplier simplifies to $(I - \mathcal{C}_Y)^{-1}$, the dynamic, heterogeneous agent equivalent of $1/(1-m)$. In this case, the iMPCs remain a sufficient statistic even in general equilibrium. However, this is no longer the case when prices adjust and monetary policy is active. A rise in demand raises inflation and triggers a monetary tightening, which tends to lower consumption initially.[12]

[11] In fact, C_Y is the aggregate iMPC matrix as defined by Auclert et al. (2018).

[12] This is assuming that the negative substitution effects dominate the positive income effects of higher real interest rates, which is the case in most calibrated models.

3 Quantitative Model

In this section, we present a medium-scale HANK model that generates realistic iMPCs and accounts for a rich set of general equilibrium effects.

The economy includes a continuum of households subject to uninsurable idiosyncratic risk, a financial institution that intermediates funds between savers and the productive sector, a capital producer faced with investment adjustment frictions, a labor agency that hires workers under searching and matching frictions and bargains the wage with a representative labor union, monopolistically competitive producers of intermediate goods, a representative final goods producer, and fiscal and monetary policy authorities. Time is discrete and one period corresponds to a quarter.

3.1 Households

A continuum of ex ante identical households, indexed by $i \in [0, 1]$, are subject to idiosyncratic risk through shocks to their productivity and their employment status, implying that they differ ex post. Idiosyncratic productivity in period $t = 0, 1, \ldots$ is given by $z_{i,t}$. Productivity follows a Markov chain with a constant transition matrix Π^z and cross-sectional distribution π^z. Mean productivity $\sum \pi^z(z)z$ is normalized to one. Each household is either employed, $e_{i,t} = 1$, or unemployed, $e_{i,t} = 0$. The time-varying mass of households in each employment state $\pi_t^e(e)$ is determined in general equilibrium. All employed agents spend their entire available time endowment for work, supplying labor. With the work time endowment normalized to one, their nonfinancial income is given by $(1 - \tau_t)w_t z_{i,t}$, where τ_t is the tax rate and w_t is the real wage. Unemployed workers search for labor and receive unemployment benefits $(1 - \tau_t)\omega^{UI} w z_{i,t}$, with gross replacement rate $\omega^{UI} \in [0, 1]$ and steady-state wage w.[13]

The households are limited to a noncontingent short-term asset to insure against idiosyncratic risk, and borrowing is ruled out. Their optimal consumption plan satisfies

$$V_t(e_{i,t}, z_{i,t}, a_{i,t-1}) = \max_{c_{i,t}, a_{i,t}} \left\{ u(c_{i,t}) + \beta \mathbb{E}_t [V_{t+1}(e_{i,t+1}, z_{i,t+1}, a_{i,t})] \right\}, \tag{8}$$

$$c_{i,t} + a_{i,t} = (1 + r_{t-1}^A)a_{i,t-1} + (1 - \tau_t)\left[y_t(e_{i,t}, z_{i,t}) + \mathcal{D}_t^{FI}(a_{i,t-1})\right] + T_{i,t}, \tag{9}$$

$$a_{i,t} \geq 0, \tag{10}$$

where $u(\cdot)$ is the period utility function and $y_t(e_{i,t}, z_{i,t})$ is idiosyncratic labor income given by

$$y_t(e_{i,t}, z_{i,t}) = z_{i,t} \Big[\mathbb{1}(e_{i,t} = 1) w_t + \mathbb{1}(e_{i,t} = 0) \omega^{UI} w \Big]. \tag{11}$$

Additionally, $c_{i,t}$ is consumption, β is the discount factor, and $a_{i,t-1}$ are short-term deposits at a financial intermediary made in period $t-1$ that pay a safe return r_{t-1}^A in the following period.

[13] Below, all variables without time subscripts take on their respective steady-state value.

The households receive additional financial income through dividends paid by the financial intermediary $\mathcal{D}_t^{FI}(\cdot)$.[14] $T_{i,t}$ is a lump-sum government transfer (or tax) that is taken as given by the households.[15]

The first-order optimality condition is

$$u'(c_{i,t}) \geq \beta(1 + r_t^A)\mathbb{E}_t \left[u'(c_{i,t+1})\right] \tag{12}$$

along with equations (9) and (10), where (12) holds with equality if the borrowing constraint is not binding.

3.2 Financial Intermediary

The representative financial intermediary holds equity shares S_t and government bonds B_t financed by net worth N_t^{FI} and deposits from the household sector A_t. Its balance sheet constraint is

$$p_t^S S_t + q_t^B B_t = N_t^{FI} + A_t, \tag{13}$$

where p_t^S is the price of shares issued by the firms S_t and q_t^B is the price of long-term bonds issued by the government B_t. Government bonds take the form of perpetuities that pay off a geometrically declining coupon $(\delta_B)^s$ with $\delta_B \in (0, 1]$ and $s = 0, 1, 2, \ldots$ starting in period $t + 1$, following Woodford (2001). The deposits that are intermediated on behalf of the households are subject to a unit intermediation cost ξ.

Intermediary net worth is given by

$$N_t^{FI} = (D_t + p_t^S)S_{t-1} + (1 + \delta_B q_t^B)B_{t-1} - (1 + r_{t-1}^A + \xi)A_{t-1} - \mathcal{D}_t^{FI}. \tag{14}$$

Equity shares acquired in $t - 1$ earn dividends D_t in addition to the ex-dividend price p_t^S, bond holdings in $t - 1$ yield the coupon payment $\delta_B q_t^B$ plus the principal, and a unit of deposits is associated with costs of the amount of $1 + r_{t-1}^A + \xi$. Net worth is further reduced by the distributions paid to the households $\mathcal{D}_t^{FI}$.

The dividend payouts to the households are assumed to follow an ad hoc distribution rule:

$$\mathcal{D}_t^{FI} = \mathcal{D}^{FI} + \phi(N_{t-1}^{FI} - N^{FI}). \tag{15}$$

ϕ parametrizes the rate at which intermediary net worth returns to steady state.[16] The smaller

[14] We adopt the common assumption that ownership of the financial intermediary is nontradable but allow the dividend payments to differ along the liquid wealth distribution.

[15] Transfers are treated as exogenous even if we allow the government to relate them to variables such as income and asset holdings, as will become clear below. Hence, in our simulation of the COVID-19 period, the households do not anticipate the structure of the discretionary support measures paid by the government beyond standard unemployment benefits. Nonetheless, they fully internalize that the tax rate τ_t must ultimately adjust to satisfy the government budget constraint.

[16] Stability of the intermediary's balance sheet requires $\mathcal{D}^{FI} = (r^A + \xi)N^{FI}$ and $\phi > r^A + \xi$. We ensure that these

ϕ is, the stronger household income is insulated from return shocks and the less reactive consumption is to swings in equity prices, for example.

The financial intermediary maximizes its expected return on net worth after distributions $\mathbb{E}_t(1+r^N_{t+1}) = \mathbb{E}_t(N^{FI}_{t+1}/N^{FI}_t)$ subject to the balance sheet constraint (13) and the law of motion of net worth (14), which yields standard no-arbitrage relationships:

$$\mathbb{E}_t[(D_{t+1}+p^S_{t+1})/p^S_t] = \mathbb{E}_t[(1+\delta_B q^B_{t+1})/q^B_t] = 1+r^A_t+\xi \equiv 1+r_t. \tag{16}$$

In equilibrium, the expected return is equated across all financial assets.

3.3 Capital Producer

A representative firm maintains the economy's capital stock K_t and rents it out to the goods-producing sector. Investment I_t is subject to convex adjustment costs $\Phi_I(\cdot)$, giving rise to real rigidity. The capital stock evolves according to the following law of motion:

$$K_t = (1-\delta_K)K_{t-1} + \left[1-\Phi_I\left(\frac{I_t}{I_{t-1}}\right)\right]I_t, \tag{17}$$

where $\delta_K \in (0,1)$ is the depreciation rate and $\Phi_I(I_t/I_{t-1}) = (\psi_I/2)(I_t/I_{t-1}-1)^2$.

The capital producer chooses investment to maximize the expected present value of the expected dividend stream. Formally, it solves

$$p^K_t(K_{t-1},I_{t-1}) = r^K_t K_{t-1} + \max_{I_t,K_t}\left\{-I_t + \frac{1}{1+r_t}\mathbb{E}_t\left[p^K_{t+1}(K_t,I_t)\right]\right\} \tag{18}$$

subject to equation (17).[17] Optimality necessitates

$$1 = Q_t\left[1-\Phi_I\left(\frac{I_t}{I_{t-1}}\right) - \Phi_I'\left(\frac{I_t}{I_{t-1}}\right)\frac{I_t}{I_{t-1}}\right] + \frac{1}{1+r_t}\mathbb{E}_t\left[Q_{t+1}\Phi_I'\left(\frac{I_{t+1}}{I_t}\right)\left(\frac{I_{t+1}}{I_t}\right)^2\right], \tag{19}$$

where Q_t satisfies the recursion

$$Q_t = \frac{1}{1+r_t}\mathbb{E}_t\left[r^K_{t+1} + (1-\delta_K)Q_{t+1}\right]. \tag{20}$$

Note that the shadow value of capital Q_t can be interpreted as Tobin's marginal Q.

3.4 Labor Agency and Union

There is involuntary unemployment resulting from searching and matching frictions in the Diamond–Mortensen–Pissarides tradition.

conditions hold in our model calibration. The details are shown in Appendix B.1.

[17] See Appendix B.2 for the derivations.

A labor agency hires workers by posting vacancies, unemployed workers search for employment, and matches are formed stochastically. The agency sells the labor services supplied by successfully matched workers to the goods-producing sector. The wage is determined by Nash bargaining between the agency and a risk-neutral union that represents all households.[18] We introduce wage rigidity, as the Diamond–Mortensen–Pissarides model generates insufficient volatility in vacancy creation and unemployment with flexible wages (Shimer, 2004; Hall, 2005; Shimer, 2005). Rather than modeling the deep sources of wage rigidity, our tractable approach is to assume that the labor agency incurs wage adjustment costs in the spirit of Rotemberg (1982). These costs diminish the bargaining surplus and thus provide an incentive for both sides to reduce fluctuations in the wage bargained.[19]

Timing and Employment Flows. In each period, the labor market operates as follows.

1. The agency inherits a stock of employed workers N_{t-1} from the previous period. $U_{t-1} = 1 - N_{t-1}$ workers were not matched in $t-1$ and still search for employment. The realizations of all aggregate shocks become known.

2. Existing matches are destroyed with an exogenous probability $s \in (0, 1)$. The newly separated workers become searchers, implying that the mass of active searchers now is $u_t = U_{t-1} + sN_{t-1}$. The agency creates v_t new vacancies at cost κ_v for each vacancy.

3. New matches are formed according to a Cobb–Douglas matching technology, $m(u_t, v_t) = \Theta_m u_t^{\alpha_m} v_t^{1-\alpha_m}$ with $\Theta_m > 0$ and $\alpha_m \in (0, 1)$, and the labor agency pays a hiring cost κ_h for each new match. The job-finding rate f_t and the vacancy-filling rate q_t are, respectively, $f_t \equiv m(u_t, v_t)/u_t = \Theta_m \theta_t^{1-\alpha_m}$ and $q_t \equiv m(u_t, v_t)/v_t = \Theta_m \theta_t^{-\alpha_m}$, where $\theta_t = v_t/u_t$ is labor market tightness. Unfilled vacancies are destroyed. From the households' perspective, the law of motion of the labor market status can be summarized as

$$\begin{bmatrix} N_t \\ U_t \end{bmatrix} = \begin{bmatrix} 1 - s(1 - f_t) & f_t \\ s(1 - f_t) & 1 - f_t \end{bmatrix} \begin{bmatrix} N_{t-1} \\ U_{t-1} \end{bmatrix}. \tag{21}$$

   From the labor agency's perspective, the evolution of employment is

$$N_t = (1 - s)N_{t-1} + q_t v_t. \tag{22}$$

4. Wage bargaining and production take place. The agency receives a fee h_t and pays a wage w_t per efficiency unit of labor.

[18] The assumption that a union bargains on behalf of the households allows us to abstract from wage dispersion, which would arise under incomplete markets if each household bargained individually with the labor agency.

[19] Different approaches have been used to achieve wage rigidity in the context of searching and matching. Hall (2005) analyzes simple wage rules, Gertler and Trigari (2009) apply Nash bargaining with staggered multi-period wage setting over a fixed horizon, and Christiano et al. (2016) consider alternating offer bargaining.

Value of a Match. The labor agency faces convex real wage adjustment costs $\Phi_w(w_t, w_{t-1}) = \psi_w/2(w_t/w_{t-1} - 1)^2$, which gives rise to real wage rigidity. The profit-maximization problem of the labor agency is

$$\mathcal{J}_t(N_{t-1}) = \max_{N_t, v_t} \left\{ (h_t - w_t)N_t - (\kappa_v + \kappa_h q_t)v_t - \Phi_w(w_t, w_{t-1}) + \frac{1}{1+r_t}\mathbb{E}_t[\mathcal{J}_{t+1}(N_t)] \right\} \tag{23}$$

subject to (22). An optimum is characterized by the following:

$$J_t = h_t - w_t - \Phi_w(w_t, w_{t-1}) + \frac{1-s}{1+r_t}\mathbb{E}_t[J_{t+1}], \tag{24}$$

$$J_t = \frac{\kappa_v}{q_t} + \kappa_h, \tag{25}$$

where J_t is the labor agency's shadow value of a match—that is, its bargaining surplus.[20] Equation (24) shows that the value of a match is the fee received from the goods producers net of the wage paid and the adjustment cost incurred plus the continuation value obtained if the match is not destroyed. Equation (25) is a zero profit condition for the labor agency equating the value of the marginal worker to the sum of the costs associated with vacancy posting and hiring.

We assume that workers are represented by a risk-neutral union, for whom the value of the marginal match is

$$H_t = w_t - w^{UI}w + \frac{1-s}{1+r_t}\mathbb{E}[(1-f_{t+1})H_{t+1}]. \tag{26}$$

It is the wage net of the unemployment benefit—a worker's outside option in case no match is formed—plus the continuation value.

Wage Bargaining. Under Nash bargaining, the equilibrium wage maximizes the joint surplus, $H_t^{\eta} J_t^{1-\eta}$, where η parametrizes the union's bargaining power. The surplus is split such that $\Omega_t J_t = (1-\Omega_t)H_t$ with

$$\Omega_t \equiv \frac{\eta}{\eta + (1-\eta)(-J_{w,t}/H_{w,t})}, \tag{27}$$

governing the share received by the union.

3.5 Goods Producers

The goods-production sector includes two types of firms: a representative final goods producer and a continuum of differentiated input producers, which are the source of nominal rigidity.

[20] See Appendix B.3 for the details.

Final Goods Producer. A perfectly competitive firm produces Y_t units of a homogenous good using a constant elasticity of substitution (CES) technology with elasticity of substitution ϵ_p:

$$Y_t = \left(\int Y_{j,t}^{\frac{\epsilon_p - 1}{\epsilon_p}} dj \right)^{\frac{\epsilon_p}{\epsilon_p - 1}}. \tag{28}$$

Cost minimization implies that the demand for intermediate producer j's input $Y_{j,t}$ with price $P_{j,t}$ is given by

$$Y_{j,t} = \left(\frac{P_{j,t}}{P_t} \right)^{-\epsilon_p} Y_t, \tag{29}$$

where $P_t = \left(\int P_{j,t}^{1-\epsilon_p} dj \right)^{1/(1-\epsilon_p)}$ is the corresponding aggregate price index.

Intermediate Goods Producers. Monopolistically competitive firms indexed by $j \in [0,1]$ produce differentiated inputs using a Cobb–Douglas technology, $Y_{j,t} = \Theta K_{j,t-1}^{\alpha} N_{j,t}^{1-\alpha}$, where Θ is total factor productivity. Following Rotemberg (1982), they face price adjustment costs

$$\Phi_p(P_{j,t}, P_{j,t-1}, \Pi_{t-1}) = \frac{\chi_p}{2} \left[\log\left(\frac{P_{j,t}}{P_{j,t-1}} \right) - \log\left(\Pi_{t-1}^{\iota_p} \Pi^{1-\iota_p} \right) \right]^2 \tag{30}$$

with $\Pi_t = P_t / P_{t-1}$ denoting inflation. In an optimum, the parameters $\chi_p > 0$ and $\iota_p \in [0,1]$ determine the strength of the price adjustment friction and the degree of indexation, respectively.[21] Profit maximization yields a standard Phillips curve with a backward- and a forward-looking component:

$$\begin{aligned} \log\left(\Pi_t\right) - \log\left(\Pi_{t-1}^{\iota_p} \Pi^{1-\iota_p}\right) &= \kappa_p \left(\frac{\epsilon_p}{\epsilon_p - 1} mc_t - 1 \right) \\ &\quad + \frac{1}{1+r_t} \mathbb{E}_t \left[\log\left(\Pi_{t+1}\right) - \log\left(\Pi_t^{\iota_p} \Pi^{1-\iota_p}\right) \right] \frac{Y_{t+1}}{Y_t}. \end{aligned} \tag{31}$$

where the slope $\kappa_p = (\epsilon_p - 1)/\chi_p$ is the same as in the loglinearized version of (31) and real marginal cost is given by

$$mc_t = \frac{1}{\Theta} \left(\frac{r_t^K}{\alpha} \right)^{\alpha} \left(\frac{h_t}{1-\alpha} \right)^{1-\alpha} \tag{32}$$

and the factor prices satisfy $r_t^K = \alpha mc_t Y_t / K_{t-1}$ and $h_t = (1-\alpha) mc_t Y_t / N_t$.[22]

Aggregate Profits. In addition to the monopolistic intermediate goods producers, the capital producer and the labor agency generate profits that they distribute as dividends. The dividends

[21] If $\iota_p = 0$ and $\Pi = 1$, the second log term in equation (30) vanishes, and the adjustment cost is minimized if the price remains fixed. If $\iota_p = 1$, the cost is minimized for full indexation to inflation in the previous period.

[22] See Appendix B.4 for all derivations.

paid are, respectively,

$$D_t^F = Y_t - r_t^K K_{t-1} - h_t N_t - \Phi_p(P_t, P_{t-1}, \Pi_{t-1}) - \Psi \tag{33}$$
$$D_t^K = r_t^K K_{t-1} - I_t \tag{34}$$
$$D_t^L = (h_t - w_t)N_t - (\kappa_v + \kappa_h q_t)\tilde{V}_t - \Phi_w(w_t, w_{t-1})N_t, \tag{35}$$

where Ψ is a fixed cost of operation in goods production. The equity shares S_t traded in the economy are claims to the stream of aggregate dividends $D_t = D_t^F + D_t^K + D_t^L$.

3.6 Government

Government policies are implemented by a fiscal authority and a central bank.

Fiscal Authority. The government's budget constraint is

$$G_t + T_t + (1 + \delta_B q_t^B)B_{t-1}^S + \omega^{UI} w U_t = q_t^B B_t^S + \tau_t \Big(w_t N_t + \mathcal{D}_t^{FI} + \omega^{UI} w U_t \Big). \tag{36}$$

It spends on, in order, output goods, aggregate transfers, debt repayment, and unemployment benefits. Revenue is generated through debt issuance and proportional taxation. Fiscal policy sets a path for transfers, taxes, and spending, implying that debt supply B_t^S is determined by the budget constraint. The path for transfers will be an important input to our simulation—we return to it below. Following Auclert et al. (2020), the tax rate is set according to

$$\tau_t = \tau + \phi_B q^B \frac{B_{t-1}^S - B^S}{Y} \tag{37}$$

with $\phi_B > 0$. It rises with the debt level, ensuring that government debt is stationary. Finally, we let spending be constant, $G_t = G$.

Monetary Policy. The central bank sets the nominal short-term interest rate according to an inertial Taylor-type rule,

$$i_t = \phi_i i_{t-1} + (1 - \phi_i) \left[i + \phi_\pi \log \left(\frac{\Pi_t}{\Pi} \right) \right] \tag{38}$$

and the real rate satisfies the standard Fisher equation $r_t = i_t - \mathbb{E}_t[\log(\Pi_{t+1})]$.

3.7 Equilibrium

In an equilibrium, all agents follow their optimal decision rules, and markets clear.

The asset market clearing conditions are the following:

$$A_t = \sum_e \sum_z \int a_t(e, z, a_{-1})\Gamma_t(e, z, a_{-1})da_{-1}, \tag{39}$$

$$S_t = 1, \tag{40}$$

$$B_t = B_t^S, \tag{41}$$

where $\Gamma_t(e, z, a_{-1})$ is the mass of households in any given employment, productivity, and asset state, and $a_t(e, z, a_{-1})$ is the optimal saving policy. Labor market clearing requires

$$\sum_e \pi_t^e(e)e = N_t, \tag{42}$$

$$N_t + U_t = 1. \tag{43}$$

If the equations (39) to (43) are satisfied, Walras' law ensures that the goods market clearing condition holds, which can be expressed as

$$\begin{aligned} Y_t = C_t + I_t + G_t + \xi A_{t-1} + (\kappa_v + \kappa_h q_t)\tilde{V}_t \\ + \Phi_w(w_t, w_{t-1})N_t + \Phi_p(P_t, P_{t-1}, \Pi_{t-1})Y_t + \Psi, \end{aligned} \tag{44}$$

where $C_t \equiv \sum_e \sum_z \int c(e, z, a_{-1})\Gamma_t(e, z, a_{-1})da_{-1}$ is aggregate consumption. An equilibrium can then be characterized as follows.

A recursive equilibrium is a sequence of prices $\{\Pi_t, p_t^S, q_t^B, r_t, r_t^K, Q_t, h_t, w_t\}$, policy functions $\{c_t(e, z, a_{-1}), a_t(e, z, a_{-1})\}$, the distribution $\{\Gamma_t(e, z, a_{-1})\}$, aggregates $\{C_t, A_t, Y_t, N_t, B_t, B_t^S, K_t, I_t, \mathcal{D}_t^{FI}, U_t, V_t, \tilde{U}_t, \tilde{V}_t, v_t, \theta_t, J_t, H_t, \Omega_t, f_t, q_t, \tau_t, mc_t, \Phi_{p,t}, \Phi_{I,t}, \Phi_{w,t}, D_t^F, D_t^K, D_t^L\}$, and policy $\{i_t, G_t, T_t\}$ such that (1) the evolution of the wealth distribution is consistent with labor market outcomes, the productivity process, and the households' policy functions; (2) the households, the financial intermediary, the capital producer, the labor agency, the union, the final goods producer, and the intermediate good producers attain their respective constrained optimum; (3) monetary and fiscal policy satisfy their respective rules as well as the government budget constraint; and (4) all markets clear—that is, equations (39) to (43) hold.

4 Calibration

In line with the analytical results presented in Section 2, our calibration strategy is particularly attentive to two sets of moments: iMPCs and fiscal multipliers. The calibration procedure can be summarized as follows. First, we set a large set of parameters based on procedures that are standard in the literature. Second, we use a subset of parameters to target iMPC estimates from micro data. Third, we validate that the iMPCs, in combination with our choice of one additional parameter, give rise to realistic general equilibrium forces by comparing the government spending multipliers of the model with reduced-form estimates. We now fill in the details, starting with the last two points.

4.1 Disciplining Consumption-Saving Behavior in Partial Equilibrium

The mapping between the partial equilibrium iMPCs and excess savings decumulation described in Section 2.1 implies that we can discipline the latter in our model by targeting the iMPCs, statistics that have been estimated empirically.

iMPCs in Data. We opt to follow Auclert et al. (2020) in basing our calibration on the iMPCs estimated by Fagereng et al. (2021), which are calculated using Norwegian lottery winnings.[23] Lottery prizes provide an ideal way of measuring iMPCs in partial equilibrium because they are drawn at random and each recipient household is infinitesimal relative to macro aggregates. The empirical evidence gives rise to three stylized facts. First, the average annual MPC on impact is sizable—about 0.5 according to Fagereng et al. (2021). Second, the average iMPCs are declining with the response horizon. However, they are still around 0.1 in the second year after the impact year. Third, the impact MPCs are strongly negatively correlated with the liquid asset position of households. While Fagereng et al. (2021) estimate a decline in the annual MPC from 0.62 to 0.46 from the bottom to the top quartile of the distribution, Holm et al. (2021) find a somewhat stronger falloff at the top of the distribution with their method of extracting MPCs from a decomposition of the consumption response to monetary policy shocks in Norway.[24]

Targeting Procedure. Our incomplete markets set-up allows us to generate a high average impact MPC together with a gradually declining average iMPC path (Auclert et al., 2020). We target these two data moments using the discount factor and a simple transfer function, leaving the distribution of iMPCs untargeted. The structural form of the transfer function assumed is

$$T_{i,t} = \tau^{LS} + \tau^a a_{i,t-1} + \varepsilon_{i,t}. \tag{45}$$

With $\beta = 0.95$, $\tau^{LS} = 0.05$, and $\tau^a = -0.07$, we obtain an annual impact MPC of 0.56 and a close model fit of the entire iMPC path, as illustrated in the left panel of Figure 2, which compares the annual data with the corresponding model values. Intuitively, a lower discount factor yields less precautionary savings in steady state and, hence, a higher impact MPC and steeper iMPC path. Similarly, a more redistributive transfer system allows for higher consumption when the

[23] To our knowledge, iMPC estimates from U.S. households are unavailable to this date. The benefits of well identified moments that are estimated from high-quality administrative tax data from Norway come at the cost of some uncertainty about cross-country differences in the consumption-saving behavior of households. While quantitative differences are possible, the qualitative features highlighted below are unlikely to differ.

[24] Holm et al. (2021) estimate a similar impact MPC at the bottom of the liquid asset distribution but the point estimate for the top 20 percent of households is about 0.35 (See Figure E.4 in their Supplemental Material).

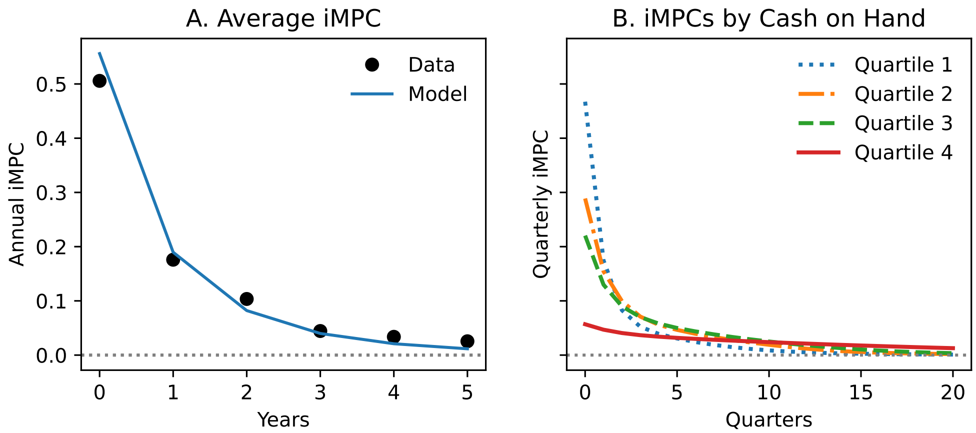

Figure 2: iMPCs in the Data and in the Model

Notes: The left panel shows the annual iMPC estimates by Fagereng et al. (2021) (dots) and the corresponding model values (line). The right panel shows the quarterly iMPC path averaged across the households in each quartile of the cash-on-hand distribution in the model.

borrowing constraint is binding and therefore deters private insurance, raising the iMPCs in the first quarters. While there is no wealth tax in the U.S., the negative value for τ^a that we obtain can be interpreted as capturing redistributive features of the tax system that we do not model explicitly.

The right panel of Figure 2 shows the iMPCs by quartile of the cash-on-hand distribution.[25] In line with the third stylized fact described above, the model successfully generates a sizable falloff in the annual impact MPCs from the first to the fourth quartile (0.78 to 0.18), which is, however, somewhat larger than what the empirical point estimates suggest.

4.2 Validating General Equilibrium Effects

Given the model's sizable MPCs, it is capable of generating substantial multipliers, as shown in Section 2.2. We validate the general equilibrium feedback in the model by comparing the government spending multipliers implied by it with the estimates by Ramey and Zubairy (2018). Specifically, we consider the cumulative government spending multiplier $\sum_{t=0}^{h} dY_t / \sum_{t=0}^{h} dG_t$ for $h = 1, 2, \ldots, 20$, where $\{dY_t\}$ and $\{dG_t\}$ are the impulse responses of output and government spending, respectively. Their results, depicted in Figure 3, are obtained using local projections with long time series from the U.S., in which government spending is instrumented with exogenously identified military news shocks.

In a first step, we estimated two model parameters—the parameter of the fiscal rule ϕ_B and the persistence parameter ρ_G of government spending specified as an AR(1) process—using the estimates by Ramey and Zubairy (2018) as targets.[26] Both parameters influence the gov-

[25] Cash on hand is the sum of all earnings, transfers, and financial income (including the principal) net of taxes.

[26] The estimation was done via the simulated method of moments and included the standard errors of the empirical targets in the weighting matrix.

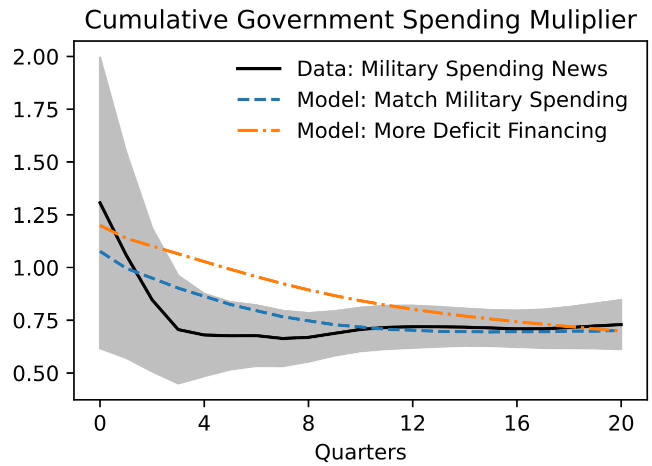

Figure 3: Cumulative Government Spending Multipliers in the Data and in the Model

Notes: The figure shows the cumulative government spending multiplier estimates by Ramey and Zubairy (2018) ("Data: Military Spending News") as well as the corresponding sequences from the model for $\phi_B = 1.93$ ("Model: Match Military Spending") and $\phi_B = 0.025$ ("Model: More Deficit Financing"). The shaded region is the 95% confidence interval.

ernment spending multipliers in the model: The former governs the sensitivity of the tax rate to additional government debt, and the latter governs the allocation of government spending over time. The dashed blue line in Figure 3 shows that the model closely matches the empirical targets with the estimated parameter values $\rho_G = 0.62$ and $\phi_B = 1.93$. The relatively high value of ϕ_B implies that a significant part of spending is tax financed. While this may be a plausible description of the fiscal response to military news shocks, the large fiscal programs that contributed to the buildup of excess savings during the COVID-19 period were almost exclusively debt financed. Therefore, in a second step, we lowered ϕ_B to the smaller value of 0.025 from Auclert et al. (2020), which implies a high degree of debt financing while ensuring that debt remains stationary. The corresponding dash-dotted orange line shows that the resulting multipliers are somewhat higher in the first quarters, as expected. However, they only lie outside the 95-percent confidence region of the estimates from military news in a few intermediate quarters. Given the strong dominance of debt financing during the pandemic and the acceptable match with the empirical estimates, we opt for the smaller of the two values in our application.

4.3 Remaining Parameters

The above calibration steps are conditional on a large set of parameters that we select based on data from the U.S. as well as findings from the literature. For example, the idiosyncratic income process with $n_z = 33$ states is taken from Kaplan et al. (2018); the parametrization of the searching and matching frictions closely follows Christiano et al. (2016); and our target for the ratio of household wealth to annual output $(A + N^{FI})/4Y$ is 3.82, as in Auclert et al. (2020), giving rise to a deposit ratio that is in line with the 2023 wave of the Survey of Consumer

Finances. Other values, such as an elasticity of intertemporal substitution of 0.5 or an average U.S. government debt maturity of five years, are standard. Section C in the Appendix contains a detailed discussion of the remaining parameter choices, which are summarized in Table C.1.

5 Excess Savings and the COVID-19 Pandemic

During the COVID-19 pandemic, U.S. households accumulated a large amount of liquid assets. We use our quantitative model to study the macroeconomic effects of the changes in household wealth that took place in this historical episode. We deliberately abstract from other pandemic-specific influences such as inflationary pressure from international supply constraints, as our interest lies in the effects that are directly attributable to the dynamics of household savings.

5.1 Historical Background

Two factors played dominant roles for the accumulation of funds in household balance sheets during the COVID-19 pandemic: large fiscal support packages and social-distancing measures. The panels A and B of Figure 4 plot aggregate transfer receipts and consumption expenditures between 2019 and 2022, respectively. In the second quarter of 2020 and the first quarter of 2021, transfers were nearly twice as high as before the pandemic. The two spikes reflect the "Economic Impact Payments" under the CARES Act and the Tax Relief Act, in combination with the American Rescue Plan.[27] Moreover, social-distancing measures contributed to a collapse in consumption in mid-2020. While these two channels increased household savings, declines in labor earnings due to business closures and layoffs reduced them. Aladangady et al. (2022) estimate that changes in fiscal support, outlays, and income resulted in a combined buildup of about 2.2 trillion dollars of excess savings at their peak in the third quarter of 2021—about 10 percent of the pre-crisis GDP in 2019.[28]

The bottom panel of Figure 4 shows a decomposition of the peak holdings into the three channels outlined above for each quartile of the income distribution based on estimates by Aladangady et al. (2022).[29] The figure reveals several distributional patterns. First, the fiscal support measures increased savings across the entire distribution, but their contribution declined with income. Second, changes in outlays reduced savings at the bottom and increased

[27] Under the Coronavirus Aid, Relief, and Economic Security (CARES) Act, the government paid eligible adults up to 1,200 dollars and children up to 500 dollars as a lump sum. These payments were phased out for individuals with a gross annual income above 75,000 dollars. The Tax Relief Act provided funds of up to 600 dollars for both adults and children subject to the same income thresholds. In a third round of stimulus payments, the American Rescue Plan Act authorized up to an additional 1,400 dollars subject to income thresholds for each adult and dependent, among other emergency relief measures.

[28] Outlays are the sum of personal consumption expenditures and interest payments. All values are in 2019 U.S. dollars.

[29] Each bar represents the deviation of a variable from its pre-crisis trend cumulated between the onset of the pandemic and the third quarter of 2021.

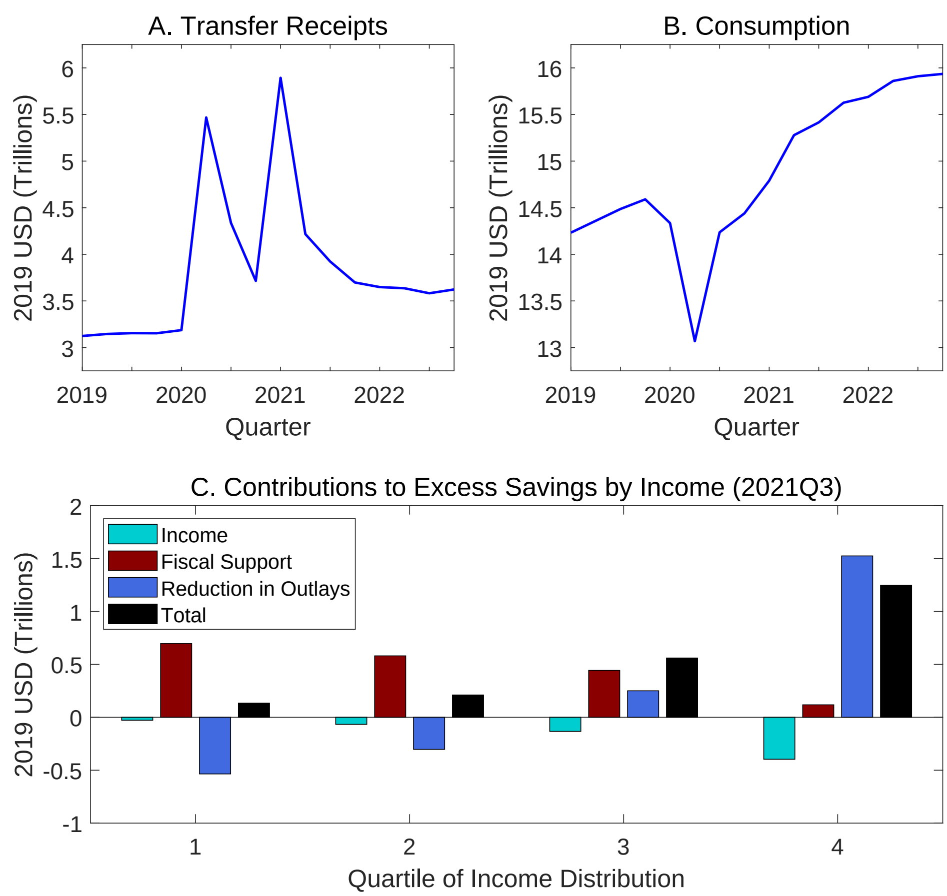

Figure 4: Aggregate Transfers and Consumption

Notes: Panels A and B show real personal current transfer receipts and real personal consumption expenditures, respectively (seasonally-adjusted annual rates). Panel C shows contributions to the stock of excess savings at the peak in 2021:Q3 by quartile of the income distribution.

savings at the top of the distribution. The negative effect at the bottom is consistent with significant spending out of the fiscal stimulus payments received. Third, income losses reduced savings in all quartiles. The total income losses were largest for the highest-income households. Fourth, excess savings were concentrated at the top of the income distribution. More than half of the stock at the peak was held by the top 25 percent.

Finally, a determinant of the rate at which excess savings are depleted is their allocation to different asset classes. Evidence from the Distributional Financial Accounts of the U.S. suggests that almost all excess savings were held as liquid assets (Batty et al., 2023). The vast majority was held in the form of bank deposits, with some rebalancing into money market funds and equities. Of course, there were also sizable revaluations of asset positions during the pandemic. While revaluation effects are not included in our definition of excess savings, they are captured by our model and contribute to the results.

5.2 Simulation Set-Up

Our experiment includes an accumulation and a decumulation period of excess savings. Prior to the pandemic, the economy is assumed to be in steady state. Starting with the first pandemic period, the first quarter of 2020, we feed exogenous shocks into the model to match data on aggregate consumption and the buildup of excess savings until their peak in the third quarter of 2021. From then onward, we restrict the shocks to zero and study the model predictions about the effects of excess savings on macroeconomic outcomes.

Shocks. We use series of targeted transfer shocks to replicate the contribution of the fiscal support programs. In addition, shocks to the discount factor approximate the contribution of the voluntary consumption restraints that resulted from social-distancing measures. Figure 4 shows the three main components of excess savings accumulation: fiscal support, reduction in outlays, and income. Of these three components, we only target the first two and let income adjust endogenously in general equilibrium. Since less affluent households received more fiscal support, we allow the transfer shocks to be specific to the quartiles of the cash-on-hand distribution at the onset of the pandemic.[30] In sum, in the buildup stage, there are four series of transfer shocks $\{\varepsilon_{q,2020Q1}, \varepsilon_{q,2020Q2}, \ldots, \varepsilon_{q,2021Q3}\}$, with $q \in \{1,2,3,4\}$ together with a sequence of common discount factors $\{\beta_{2020Q1}, \beta_{2020Q2}, \ldots, \beta_{2021Q3}\}$.

Targets. The data targets are the stock of excess savings held in all four quartiles of the income distribution and aggregate consumption over the seven quarters of the buildup period. We choose the shocks such that they minimize the squared distance between the data targets and their model counterparts. Since the four transfer shock series and the discount factor shock series are effective determinants of household savings by income quartile and aggregate consumption, respectively, we obtain a near exact match of the data targets in the model.

Validation. Our simulation set-up captures the main contributors to the buildup of excess savings in a parsimonious way. By targeting excess savings directly, we leave transfer payments untargeted. To validate our set-up, we compare the aggregate transfers implied by our estimation with realized data in Figure 5. The fact that the transfers required by the model in the buildup phase are of the right magnitude is a success of the model.[31]

In contrast to Section 2, we now model the buildup of excess savings. This is not strictly necessary. One could also start the simulation in the third quarter of 2021 and confront the model with a set of one-off shocks to the initial conditions that allow the asset distribution to

[30] In our model with only liquid assets, there is an equivalence between household income and cash on hand, as all components of cash on hand are income in each quarter. We exploit this equivalence by proxying for household income using cash on hand, the relevant state variable in our model.

[31] A sizable discrepancy appears in the first quarter of the simulation, in which the importance of transfers relative to consumption restraints is overstated by the model. However, experimenting with constraints on transfers in the first quarter of 2020 showed that their effect on the decumulation phase is negligible.

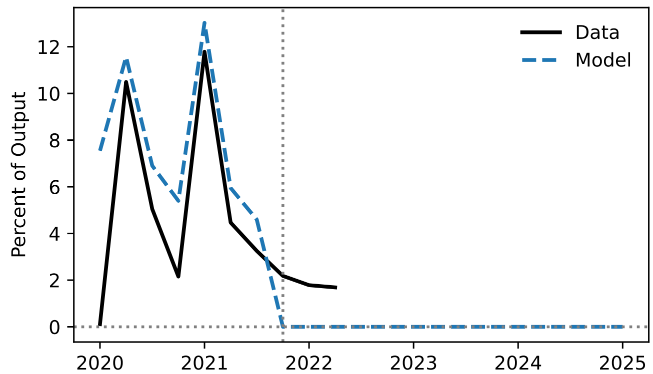

Figure 5: Aggregate Transfers in the Buildup of Excess Savings

Notes: The figure shows the ratio of aggregate transfers to steady-state output in the model (black solid line) and the ratio of realized transfer income in deviations from its pre-crisis trend relative to the pre-crisis trend in GDP in the U.S. (blue dashed line). The vertical line marks the third quarter of 2021, the end of the buildup period.

line up with the data at that point in time. Including the buildup in the simulation has several advantages, though. During the accumulation phase, general equilibrium forces are set off that interact with the excess savings depletion. In our view, these interactions are an important aspect of the excess savings dynamics and should not be excluded from the analysis. In addition, as we discuss in the context of the inflation response below, the forward-looking nature of the model implies that some of the macroeconomic effects of the excess savings depletion occur before the peak stock is reached. Starting the simulation at the onset of the buildup allows us to capture these effects.

5.3 Excess Savings Dynamics

Figure 6 compares simulated total excess savings with the corresponding empirical counterpart, and Figure 7 contains a disaggregation into income quartiles, with the empirical estimates on the left and the simulation output on the right.

Aggregate Excess Savings. While the near exact data match until the third quarter of 2021 is achieved by construction, an important test is how the model performs in replicating the behavior of realized excess savings between the end of 2021 and the end of 2022, the part of the decumulation phase for which the empirical estimates are available. Figure 6 shows that the model does well in predicting the realized decumulation rate. The model-implied path is slightly steeper initially, which may reflect a more gradual resolution of consumption restraints than our experiment assumes. In the model and in the data, the households have spent a bit more than half of their excess savings by the end of 2022. Almost the entire excess savings stock is depleted at the beginning of 2025. A standard representative agent or spender-saver framework would predict that households hold on to excess savings until taxes rise. Given the slow pace of fiscal consolidation, these models would imply almost no decline in excess

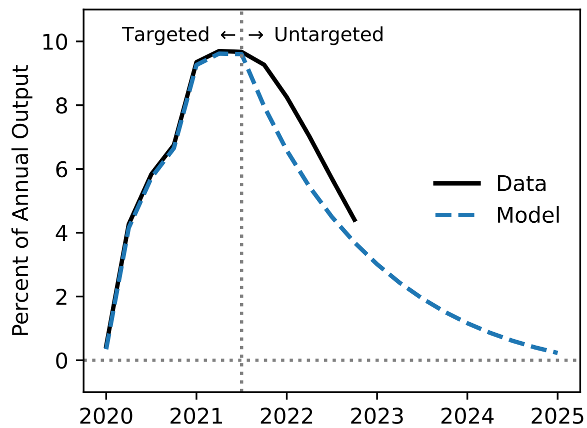

Figure 6: Aggregate Excess Savings

Notes: The figure compares aggregate excess savings in the model with the empirical estimates of Aladangady et al. (2022). In the model, excess savings are the deviation of aggregate assets from the steady state, which we normalize by steady-state output. In the data, excess savings are normalized by a pre-pandemic trend in nominal GDP. The vertical line marks the third quarter of 2021, the end of the buildup period.

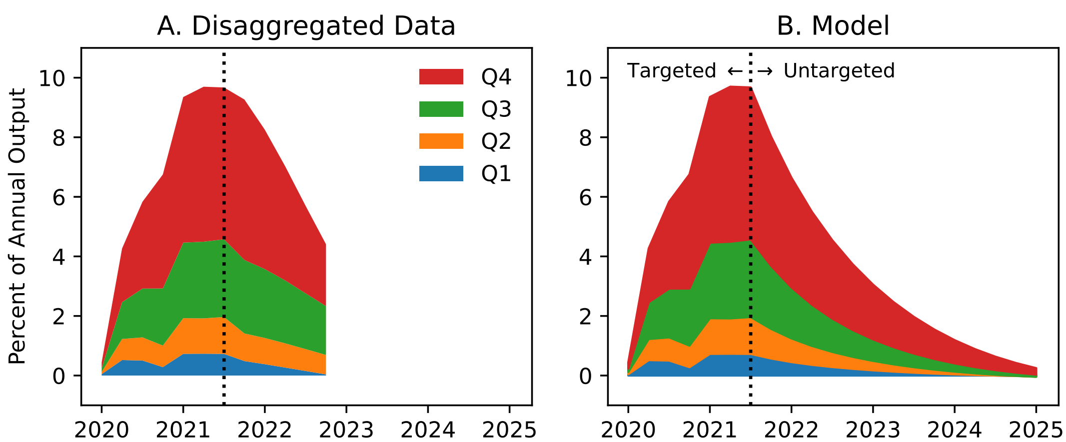

Figure 7: Distribution of Excess Savings: Data vs. Model

Notes: The figure compares excess savings across income quartiles in the model with the empirical estimates of Aladangady et al. (2022). In the model, excess savings are the deviation of aggregate assets from steady state, which we normalize by steady-state output. In the data, excess savings are normalized by a pre-pandemic trend in nominal GDP. The vertical line marks the third quarter of 2021, the end of the buildup period.

savings over the period considered. In contrast, our HANK model successfully predicts the rapid depletion of excess savings.

Distribution. Figure 7 demonstrates that the depletion rate is matched well by the model not only in the aggregate but also across the income distribution. The households in the highest quartile maintain the largest share of excess savings and the ones in the lowest quartile run out of excess savings the fastest. These dynamics may be a result of the concentration of the peak excess savings stock at the top of the distribution in conjunction with initially lower iMPCs of the top quartile. They may also result from "trickling up," the process by which savings flow from high-MPC to low-MPC households in general equilibrium. We return to a separation of partial and general equilibrium effects in Section 5.5. Note also that even the households in the top quartile run down their excess savings in less than four years, which is consistent with the still nonnegligible MPCs of top households in our model and at odds with the behavior of permanent income consumers in the absence of strong general equilibrium forces.

5.4 Macroeconomic Implications

We now turn to the broader macroeconomic consequences of excess savings. Figure 8 portrays the impact of excess savings on output, aggregate consumption, and investment on the left as well as the nominal short-term interest rate, inflation, and the real rate, all expressed as annualized percentage rates, on the right. Aggregate consumption and hence output initially collapse and then expand, overshooting their respective steady-state values at the end of the accumulation phase.[32] Because inflation depends on the expected discounted sum of real marginal cost and marginal cost follows the dynamics of output, inflation first declines but rises above its target level as soon as mid-2020. The monetary authority adjusts the nominal interest rate accordingly. Depressed demand during the buildup stage and the anticipation of heightened real rates during the decumulation stage weigh on investment, which partly crowds out the positive effect of spending out of excess savings on aggregate demand. The rise in the real rate also weakens aggregate consumption. All in all, the implications of the excess savings dynamics are akin to those of standard fiscal shocks, as expected from the combination of disturbances that underlie their buildup.

Inflation and Monetary Policy. Our simulation further indicates that the decumulation of excess savings contributed to the surge in inflation seen in the U.S. in the wake of the pandemic and that the Federal Reserve reacted less restrictively to the inflation pressure created by the depletion of excess savings than a standard Taylor rule prescribes.

Realized inflation and the simulated inflation path are compared in Figure 9. Inflation in the model is now recentered around a target value of 2 percent for comparability. Over the first two quarters of 2020, average realized inflation dropped to about 0.5 percent. Subsequently, realized inflation rose steeply, averaging 4.9 percent over the second half of 2021. According

[32] The model implies that aggregate consumption jumps up at the onset of the decumulation phase and then decays gradually. Personal consumption expenditure (PCE) in the U.S. has indeed been remarkably strong since the middle of 2021, as Figure D.1 in the Appendix shows. Although PCE did not jump as abruptly as in the model, its cumulative strength over 2022–2024 aligns well with that in the model, and the model-implied path for excess savings matches the empirical estimates well. The abrupt consumption change in the model reflects that we shut down the exogenous drivers of excess savings accumulation for all households at once starting in 2021Q4. We interpret the smoother consumption path in the data as a result of a more gradual "return to normalcy" but still prefer a clean separation between excess savings accumulation and depletion for transparency. With a strongly forward-looking Phillips curve, the timing of the rise in aggregate demand matters much less for inflation than its overall magnitude.

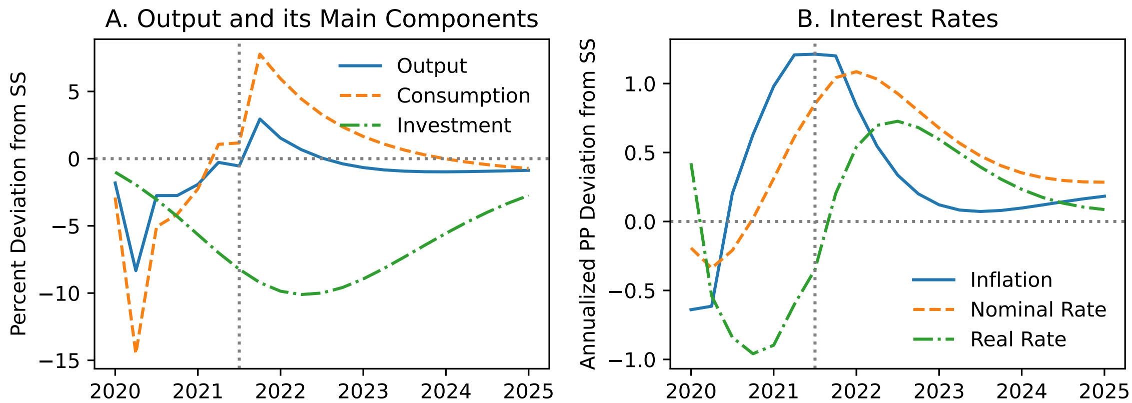

Figure 8: Effect of Excess Savings Depletion on Macroeconomic Aggregates

Notes: The figure shows model-based simulations of macroeconomic aggregates on the left and interest rates and inflation on the right. Shown are either percent- or annualized percentage point-deviations from steady state (SS). The vertical lines mark the third quarter of 2021, the end of the buildup period.

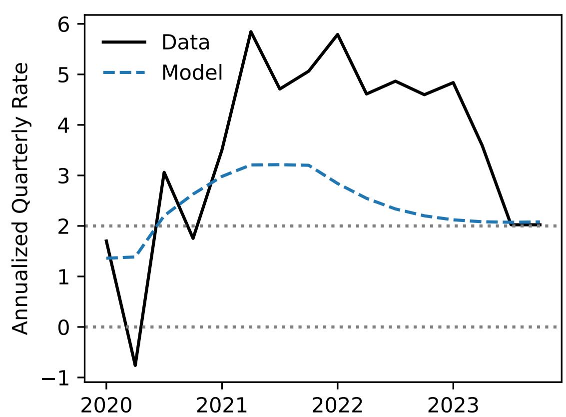

Figure 9: Simulated Inflation Due to Excess Savings and Realized Inflation

Notes: The figure shows simulated inflation in the model and realized core personal consumption expenditure price inflation (annualized quarterly rates). The simulated series is re-centered, reflecting an inflation target of 2 percent for comparability. The vertical line marks the third quarter of 2021, the end of the buildup period.

to our model, excess savings are associated with an increase in the inflation rate of about 1.8 percentage points over the same period. We therefore conclude that about 40 percent of the inflation built up by late 2021 may have originated in excess savings dynamics.

In the model, the policy rate falls about 0.25 percentage points below its long-run value in the first half of 2020. Thereafter, the model's policy rate increases about 1.2 percentage points over the next two years because of the inflation created by the rundown of excess savings. The Federal Reserve similarly entered into a sequence of rate hikes but reacted more slowly than the model's Taylor rule implies. While, in the model, the policy rate peaks close to 3 percent, about a percentage point above its long-run value of 2 percent, in the beginning of 2022, the effective federal funds rate stood at only 0.3 percent at the end of the first quarter of

2022.[33] Thus, compared to a standard Taylor rule that is seen to capture monetary policy well in recent decades, the Federal Reserve initially under-reacted to the inflation pressure generated by excess savings.[34] Clearly, the Federal Reserve reacted to more influences on inflation than the ones related to excess savings isolated by our analysis. However, the figure shows that these additional forces jointly resulted in even more inflation so that the monetary-policy response that follows the Taylor rule is even stronger and the discrepancy with realized policy is even larger if these additional factors are taken into account. Conversely, if we fixed monetary policy at the more expansionary path in the data, our model would attribute an even larger share of the realized inflation surge in 2021 to excess-savings dynamics.

Sensitivity. The inflation path in our model simulation is sensitive to the value assumed for the parameter governing the slope of the Philips curve κ_p. It is well known that several identification issues arise in the estimation of the slope parameter based on macro data alone (Mavroeidis et al., 2014; McLeay and Tenreyro, 2020). Our baseline calibration, $\kappa_p = 0.05$, follows Gagliardone et al. (2023) who estimate the slope of the Philips curve using detailed micro panel data at the firm-product level that mitigate these concerns. Their estimate of 0.05 to 0.06 for the slope of the marginal cost-based Phillips curve lies within the range of values that can be found in the DSGE literature. For example, the estimation by Smets and Wouters (2007) yields a point estimate of 0.01, while Kaplan et al. (2018) opt for a value of 0.1 referring to Schorfheide (2008). Recent estimates based on regional data suggest that the Phillips curve with a measure of economic slack as the forcing variable was very flat in the years leading up to the COVID-19 pandemic (e.g., Hazell et al., 2022). Gagliardone et al. (2023) demonstrate that this finding is consistent with their estimates when the low elasticity of marginal cost to the output gap or the unemployment gap is taken into account.[35] Additionally, Harding et al. (2023) argue that the Phillips curve may have steepened in the wake of the pandemic.

Table 1 shows how the simulated inflation dynamics change when the slope parameter κ_p is varied away from the baseline value of 0.05. As the the Phillips curve becomes flatter and inflation becomes less responsive to marginal cost and hence aggregate demand, the rise in inflation between the first half of 2020 and the second half of 2021 declines. A moderately high slope of $\kappa_p \in (0.075, 0.1)$ enables the model to match the initial drop in realized inflation to 0.47 percent at the onset of the pandemic. With a low value of $\kappa_p = 0.025$, the models still associates about 20 percent of the realized surge in inflation with excess savings dynamics.

[33] The parameters of the Taylor rule are $\phi_i = 0.8$ and $\phi_\pi = 1.5$ reflecting a realistic degree of inertia and a standard amount of inflation responsiveness.

[34] It should be stressed that our analysis is not an assessment of optimal or constrained optimal policy.

[35] Micro-founded formulations of the Phillips curve in terms of the output gap rely on assumptions that make the output gap proportional to marginal cost and hence a good proxy thereof.

Table 1: Inflation and the Slope of the Phillips Curve

	2020H1	2021H2	Δ	Fraction of Data
Data	0.47	4.89	4.42	
Model				
$\kappa_p = 0.01$	2.36	2.70	0.34	0.08
$\kappa_p = 0.025$	2.00	2.91	0.92	0.21
$\kappa_p = 0.05$	1.37	3.21	1.83	0.42
$\kappa_p = 0.075$	0.79	3.47	2.68	0.61
$\kappa_p = 0.10$	0.24	3.70	3.46	0.78

Notes: Shown is average quarterly inflation in the first two quarters of 2020 (2020H1) and in the last two quarters of 2021 (2021H2) in the data and in the model, expressed as annual percentage rates. The model results are shown for different values of the slope of the Phillips curve κ_p and are re-centered around an inflation target of 2 percent. Δ is the change between 2021H2 and 2020H1. The last column gives the ratio of Δ in the model and in the data.

5.5 Isolated Decumulation or Equilibrium Feedback

The depletion rate of aggregate excess savings depends on the joint distribution of the initial excess savings and the iMPCs, as well as general equilibrium forces, as demonstrated in Section 2. We now investigate the relative importance of these two factors, showing that the direct effect of excess savings is a strong predictor of the depletion path.

Formal Decomposition. A decomposition of the excess savings dynamics into partial and general equilibrium components is facilitated by our solution procedure, the sequence-space Jacobian method laid out in Auclert et al. (2021). According to this method, the model is divided into blocks, subsets of model equations, for which Jacobians are calculated in sequence space. The equilibrium conditions pertaining to the households are collected in a block that determines paths for aggregate savings and consumption if it is supplied with all block inputs—sequences for the exogenous variables, $\{d\beta_t\}$ and $\{d\varepsilon_{q,t}\}$, and sequences for the endogenous variables, $\{dr_t^A\}$, $\{dw_t\}$, $\{d\tau_t\}$, $\{d\mathcal{D}_t^{FI}\}$, $\{df_t\}$, $\{dU_t\}$, and $\{dN_t\}$. Let $dA = (dA_0, dA_1, \dots)'$ be a column vector giving the general equilibrium response of aggregate excess savings to the combination of shocks fed into our simulation. Then, to a first-order approximation,

$$dA = \underbrace{\sum_{\varsigma \in \mathcal{I}^{ex}} \mathcal{J}_\varsigma^A d\varsigma}_{PE} + \underbrace{\sum_{X \in \mathcal{I}^{en}} \mathcal{J}_X^A dX}_{GE}, \tag{46}$$

where $\mathcal{I}^{ex} = \{\beta, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\}$ is the set of all exogenous inputs to the household block and $\mathcal{I}^{en} = \{r^A, w, \tau, D^{FI}, f, U, N\}$ collects all endogenous inputs. Further, $\mathcal{J}_v^A$ is a square matrix with elements $[\mathcal{J}_v^A]_{t,s} = \partial A_t / \partial v_s$ for any $v \in \mathcal{I}^{ex} \cup \mathcal{I}^{en}$, $d\varsigma$ is a vector specifying the value of shock ς in each period, and dX is the evolution of the endogenous input X in general equilibrium. Thus, dA can be decomposed into two components labeled "PE" and "GE," respectively: the partial response to the exogenous shocks holding all other model variables fixed and the response to

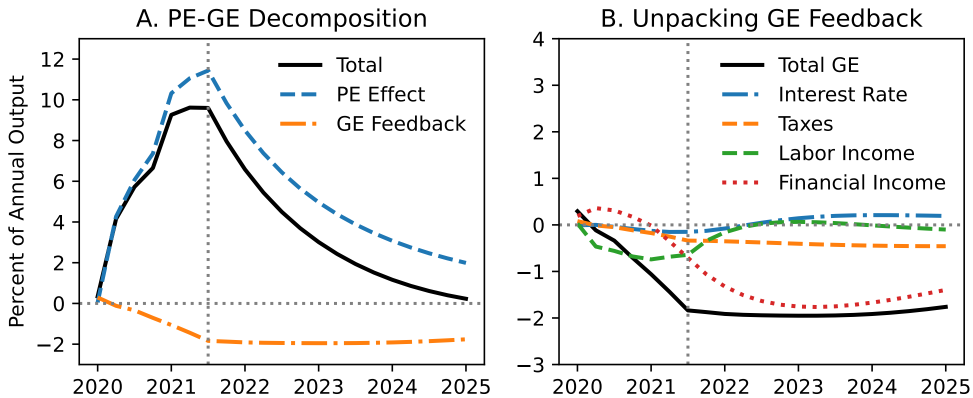

Figure 10: Partial and General Equilibrium Contributions to Excess Savings
Notes: The figure shows total excess savings and their decomposition given by equation (46). The vertical lines mark the third quarter of 2021, the end of the buildup period.

the equilibrium dynamics of the endogenous variables.

Excess Savings in PE and GE. Figure 10 provides a visual representation of the decomposition. The left panel shows that the partial equilibrium forces dominate the excess savings response. By equation (46), besides the shock series, the PE component depends on $\mathcal{J}^A_\beta$, the response of savings to changes in the discount factor, and $\mathcal{J}^A_{\varepsilon_q}$, which is determined by the iMPCs of the households in quartile q. Figure D.2 in the Appendix breaks down the PE component into the contributions of the discount factor shocks and all transfer shocks. It reveals that the former contribute somewhat more to the buildup of excess savings in partial equilibrium than the latter. Both, the excess savings generated through the fiscal transfers and those resulting from voluntary consumption restraints are spent out rapidly. This is the case because the transfers targeted bottom households with high MPCs and social distancing predominantly affected the spending of top households, who are responsive to variation in the time preference rate.

The model's general equilibrium forces reduce excess savings through four channels, as can be seen from the right panel of Figure 10. First, the contribution of labor income is negative initially as weaker aggregate demand lowers the wage rate and the job-finding rate before it gradually recovers. Second, the initial drop in the real interest rate discourages saving and the subsequent rise yields a small positive effect in later quarters. Third, with a higher expected real interest rate path over the decumulation period, government bond and equity prices fall. The decline in asset prices results in smaller distributions from the financial intermediary and a negative total contribution of financial income despite the higher anticipated return on deposits.[36] Fourth, rising government debt is met with increases in the income tax rate and, hence, losses in disposable income.

[36] While the response of distributions $\mathcal{D}^{FI}_t$ is negative at all times, the distributions' contribution to excess savings is small but positive in the first quarters, as households with high cash-on-hand initially cut into consumption to maintain their target stock of savings.

Our analysis indicates that the general equilibrium feedback, on net, leads households to accumulate less excess savings and to spend them down faster. This result is in contrast with Auclert et al. (2023b), who emphasize that the multiplier process—one household's spending from excess savings is another household's income—prolongs the duration of excess savings. The reason for this discrepancy is that the aggregate demand multiplier in our model is dampened by several additional factors. Higher spending generates inflation, prompting the central bank to raise interest rates. Higher real rates reduce not only the consumption of unconstrained households but also investment and vacancy creation. As we explain in Section 4, our model's cumulative multiplier fits the evidence from military spending shocks well, assuming a low degree of deficit financing. In our main excess-savings experiment, we allow for a much higher degree of deficit financing, which raises the multiplier in the short run. Eventually, taxes rise to consolidate real government debt, reducing the cumulative multiplier in the long run. In contrast, Auclert et al. (2023b) let government debt rise permanently.

6 Conclusion

The permanent income hypothesis (PIH) underpinning standard representative agent models implies that wealth shocks are essentially irrelevant. The representative household is willing to hold any amount of assets indefinitely as long as $\beta = (1+r)^{-1}$. Two-agent models of the standard spender-saver type share this stark implication. By definition, excess savings are held by savers, who behave according to the PIH. As Bilbiie et al. (2021) point out, savings can never be excessive according to these models.

HANK models permit a richer theory of excess savings, allowing us to account for the fact that excess savings are distributed across households with different iMPCs. The distribution matters because the iMPCs govern the spend-out rate of excess savings. Moreover, micro-level iMPCs are informative for the consumption responses to changes in taxes, labor earnings, and financial income, which are associated with excess savings depletion in general equilibrium. Building on these analytical insights, we construct a quantitative HANK model with tightly disciplined iMPCs and general equilibrium multipliers to study the macroeconomic consequences of the excess savings built up during the COVID-19 pandemic. A parsimonious set of transfer and discount factor shocks allows us to replicate the buildup of excess savings between the beginning of the pandemic and their peak in the third quarter of 2021. While they are not targeted, the model closely matches both aggregate and distributional data from the part of the rundown period for which estimates of excess savings are available.

We find that the historical spike in inflation that occurred in the U.S. in 2021 had a significant demand component, which contradicts the prevalent view that the inflation surge was largely driven by supply-side constraints. According to conventional wisdom, while central banks may face trade-offs following supply shocks that induce them to remain inactive, a monetary contraction is warranted in response to expansionary demand shocks. Our analysis suggests that the Federal Reserve may have underestimated the demand sources underlying the inflation surge in 2021 and therefore responded with a considerable delay.

Finally, we find that the macroeconomic effects of excess savings can be well predicted from the joint distribution of initial excess savings and iMPCs. Our analysis demonstrates that models that abstract from distributions cannot generally substitute for models that accommodate distributional data directly.

References

Aiyagari, S. R. and McGrattan, E. R. (1998). The optimum quantity of debt. *Journal of Monetary Economics*, 42(3):447–469.

Aladangady, A., Cho, D., Feiveson, L., and Pinto, E. (2022). Excess Savings during the COVID-19 Pandemic. FEDS Notes 2022-10-21, Board of Governors of the Federal Reserve System.

Auclert, A., Bardóczy, B., and Rognlie, M. (2023a). MPCs, MPEs, and Multipliers: A Trilemma for New Keynesian Models. *The Review of Economics and Statistics*, 105(3):700–712.

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

Auclert, A., Rognlie, M., and Straub, L. (2018). The Intertemporal Keynesian Cross. Working Paper 25020, National Bureau of Economic Research.

Auclert, A., Rognlie, M., and Straub, L. (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model. CESifo Working Paper Series 8051, CESifo.

Auclert, A., Rognlie, M., and Straub, L. (2023b). The trickling up of excess savings. *AEA Papers and Proceedings*, 113:70–75.

Batty, M., Deeken, E., Holmquist, E., and Volz, A. H. (2023). Wealth inequality and COVID-19 in the U.S.: evidence from the distributional financial accounts. In *Post-pandemic landscape for central bank statistics*, volume 58 of *IFC Bulletin*. Bank for International Settlements.

Bayer, C., Born, B., and Luetticke, R. (2024). Shocks, Frictions, and Inequality in US Business Cycles. *American Economic Review*, 114(5):1211–47.

Bayer, C., Born, B., Luetticke, R., and Müller, G. J. (2023). The Coronavirus Stimulus Package: How Large is the Transfer Multiplier. *The Economic Journal*, 133(652):1318–1347.

Bilbiie, F., Eggertsson, G., Primiceri, G., and Tambalotti, A. (2021). "Excess Savings" Are Not Excessive. Liberty Street Economics 20210405a, Federal Reserve Bank of New York.

Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER Macroeconomics Annual*, 4:185–216.

Carroll, C. D. (2004). Theoretical Foundations of Buffer Stock Saving. NBER Working Papers 10867, National Bureau of Economic Research, Inc.

Carroll, C. D., Crawley, E., Slacalek, J., and White, M. N. (2021). Modeling the Consumption Response to the CARES Act. *International Journal of Central Banking*, 17(1):107–141.

Christelis, D., Georgarakos, D., and Jappelli, T. (2015). Wealth shocks, unemployment shocks and consumption in the wake of the Great Recession. *Journal of Monetary Economics*, 72:21–41.

Christiano, L., Eichenbaum, M., and Trabandt, M. (2016). Unemployment and business cycles. *Econometrica*, 84:1523–1569.

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2021). The Macroeconomics of Epidemics. *The Review of Financial Studies*, 34(11):5149–5187.

Fagereng, A., Holm, M. B., and Natvik, G. J. (2021). MPC Heterogeneity and Household Balance Sheets. *American Economic Journal: Macroeconomics*, 13(4):1–54.

Gagliardone, L., Gertler, M., Lenzu, S., and Tielens, J. (2023). Anatomy of the Phillips Curve: Micro Evidence and Macro Implications. Working Paper 31382, National Bureau of Economic Research.

Gertler, M. and Trigari, A. (2009). Unemployment fluctuations with staggered nash wage bargaining. *Journal of Political Economy*, 117(1):38–86.

Gornemann, N., Kuester, K., and Nakajima, M. (2021). Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy. Unpublished manuscript.

Guerrieri, V. and Lorenzoni, G. (2017). Credit Crises, Precautionary Savings, and the Liquidity Trap. *The Quarterly Journal of Economics*, 132(3):1427–1467.

Guren, A. M., McKay, A., Nakamura, E., and Steinsson, J. (2021). Housing Wealth Effects: The Long View. *The Review of Economic Studies*, 88(2):669–707.

Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. *American Economic Review*, 95(1):50–65.

Harding, M., Lindé, J., and Trabandt, M. (2023). Understanding post-COVID inflation dynamics. *Journal of Monetary Economics*, 140:S101–S118.

Hazell, J., Herreño, J., Nakamura, E., and Steinsson, J. (2022). The Slope of the Phillips Curve: Evidence from U.S. States. *The Quarterly Journal of Economics*, 137(3):1299–1344.

He, Z., Liao, G., and Wang, B. (2022). What gets measured gets managed: Investment and the cost of capital. Working Paper 29775, National Bureau of Economic Research.

Heathcote, J. and Perri, F. (2018). Wealth and Volatility. *The Review of Economic Studies*, 85(4):2173–2213.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal Tax Progressivity: An Analytical Framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Holm, M. B., Paul, P., and Tischbirek, A. (2021). The Transmission of Monetary Policy under the Microscope. *Journal of Political Economy*, 129(10):2861–2904.

Kaplan, G., Mitman, K., and Violante, G. L. (2020a). The Housing Boom and Bust: Model Meets Evidence. *Journal of Political Economy*, 128(9):3285–3345.

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

Kaplan, G., Moll, B., and Violante, G. L. (2020b). The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S. Working Paper 27794, National Bureau of Economic Research.

Koby, Y. and Wolf, C. (2020). Aggregation in Heterogeneous-Firm models: Theory and measurement. Unpublished Manuscript.

Mavroeidis, S., Plagborg-Møller, M., and Stock, J. H. (2014). Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature*, 52(1):124–88.

McKay, A., Nakamura, E., and Steinsson, J. (2016). The Power of Forward Guidance Revisited. *American Economic Review*, 106(10):3133–3158.

McKay, A. and Reis, R. (2016). The Role of Automatic Stabilizers in the U.S. Business Cycle. *Econometrica*, 84(1):141–194.

McLeay, M. and Tenreyro, S. (2020). Optimal Inflation and the Identification of the Phillips Curve. *NBER Macroeconomics Annual*, 34:199–255.

Mian, A., Rao, K., and Sufi, A. (2013). Household Balance Sheets, Consumption, and the Economic Slump. *The Quarterly Journal of Economics*, 128(4):1687–1726.

Mian, A. and Sufi, A. (2014). What Explains the 2007-2009 Drop in Employment? *Econometrica*, 82(6):2197–2223.

Oh, H. and Reis, R. (2012). Targeted transfers and the fiscal response to the great recession. *Journal of Monetary Economics*, 59(S):50–64.

Ramey, V. and Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from us historical data. *Journal of Political Economy*, 126(2):850 – 901.

Rotemberg, J. (1982). Monopolistic price adjustment and aggregate output. *Review of Economic Studies*, 49(4):517–531.

Schorfheide, F. (2008). DSGE model-based estimation of the New Keynesian Phillips curve. *Economic Quarterly*, 94(Fall):397–433.

Shimer, R. (2004). The consequences of rigid wages in search models. *Journal of the European Economic Association*, 2(2-3):469–479.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, 95(1):25–49.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606.

Woodford, M. (2001). Fiscal requirements for price stability. *Journal of Money, Credit and Banking*, 33(3):669–728.

A Illustrative Model

A.1 Proof of the Proposition

Two comments. First, we will use the fact that the consumption function is differentiable. This can be established along the lines of Carroll (2004). Second, we will show the result conditional on a particular sequence of productivity shocks that, starting from an initial x_0, give rise to a sequence of cash-on-hand values $x_1, x_2, \ldots, x_t$. This is sufficient to prove the result in the main text, which averages these idiosyncratic shocks. Finally, we omit i subscripts for simplicity.

Consider $t = 0$. Let $x_0 = (1 + r_{-1})a_{-1} + y_0$ denote the default cash-on-hand in this period. The MPC from lump-sum transfers received in period 0 is

$$m_0(x_0) = \lim_{\Delta \to 0} \frac{c_0(x_0 + \Delta) - c_0(x_0)}{\Delta} = c_0'(x_0). \tag{A.1}$$

That is, $m_0(x_0)$ is simply the slope of the time-0 consumption function at the original state x_0.

Consider $t = 1$. Let $x_1 = (1 + r_0)a_0(x_0) + y_1$ denote the default cash-on-hand in period 1. If there was lump-sum transfer Δ in period 0, then cash-on-hand in period 1 is

$$\begin{aligned}
x_1(\Delta) &= y_1 + (1 + r_0)a_0(x_0 + \Delta) && \text{(A.2)}\\
&= y_1 + (1 + r_0)\left[x_0 + \Delta - c_0(x_0 + \Delta)\right] && \text{(A.3)}\\
&= x_1 + (1 + r_0)\left[\Delta - c_0(x_0 + \Delta) + c_0(x_0)\right] && \text{(A.4)}\\
&= x_1 + (1 + r_0)\Delta\left[1 - \frac{c_0(x_0 + \Delta) - c_0(x_0)}{\Delta}\right] && \text{(A.5)}
\end{aligned}$$

Substitute this into the definition of m_1:

$$m_1(x_0) = \lim_{\Delta \to 0} \frac{c_1\left(x_1 + (1 + r_0)\Delta\left[1 - \frac{c_0(x_0+\Delta) - c_0(x_0)}{\Delta}\right]\right) - c_1(x_1)}{\Delta} \tag{A.6}$$

Since $c_t(\bullet)$ is a differentiable function for all t, we may take the limit inside and write

$$\begin{aligned}
m_1(x_0) &= \lim_{\Delta \to 0} \frac{c_1\Big(x_1 + (1 + r_0)\Delta\left[1 - c_0'(x_0)\right]\Big) - c_1(x_1)}{\Delta} && \text{(A.7)}\\
&= \lim_{\Delta' \to 0} \frac{c_1\left(x_1 + \Delta'\right) - c_1(x_1)}{\Delta'}(1 + r_0)\left[1 - c_0'(x_0)\right] && \text{(A.8)}\\
&= c_1'(x_1) \cdot (1 + r_0)\left[1 - c_0'(x_0)\right] && \text{(A.9)}
\end{aligned}$$

That is, $m_1(x_0)$ is the slope of the time-1 consumption function at state x_1 times the *excess savings* that remains from the transfer in period 0.

Consider $t = 2$. Let $x_2 = (1 + r_1)a_1(x_1) + y_2$ denote the default cash-on-hand in period 2. If there was lump-sum transfer Δ in period 0, then cash-on-hand in period 2 is

$$\begin{aligned} x_2(\Delta) &= y_2 + (1 + r_1)a_1(x_1(\Delta)) \\ &= y_2 + (1 + r_1)\left[x_1(\Delta) - c_1(x_1(\Delta))\right] \\ &= x_2 + (1 + r_1)\Delta\left[\frac{x_1(\Delta) - x_1}{\Delta} - \frac{c_1\big(x_1(\Delta)\big) - c_1(x_1)}{\Delta}\right] \end{aligned}$$

Using the results for $t = 1$, we have

$$\lim_{\Delta\to 0}\frac{x_1(\Delta) - x_1}{\Delta} = \lim_{\Delta\to 0}(1 + r_0)\left[1 - \frac{c_0(x_0 + \Delta) - c_0(x_0)}{\Delta}\right] = (1 + r_0)\left[1 - c_0'(x_0)\right] \tag{A.10}$$

$$\lim_{\Delta\to 0}\frac{c_1\big(x_1(\Delta)\big) - c_1(x_1)}{\Delta} = (1 + r_0)\left[1 - c_0'(x_0)\right]c_1'(x_1) \tag{A.11}$$

As before, we substitute these limits into the definition of $m_2(x_0)$, invoking the differentiability of the consumption functions, to get

$$m_2(x_0) = \lim_{\Delta\to 0}\frac{c_2\big(x_2(\Delta)\big) - c_2(x_2)}{\Delta} = c_2'(x_2)\cdot(1 + r_1)(1 + r_0)\Big[1 - c_0'(x_0)\Big]\Big[1 - c_1'(x_1)\Big]. \tag{A.12}$$

From here, it's easy to see the pattern for general $t \geq 1$:

$$m_t(x_0) = c_t'(x_t)\prod_{s=0}^{t-1}(1 + r_s)\Big[1 - c_s'(x_s)\Big] \tag{A.13}$$

The first term is the *static MPC* in period t. The second term is the *excess savings* left from the initial transfer.

A.2 Intertemporal Keynesian Cross

In this appendix, we specify the details of a small-scale HANK model that gives rise to the intertemporal Keynesian cross (IKC)

$$\mathbf{Y} = \mathcal{C}(\boldsymbol{\tau}, \mathcal{R}(\mathcal{K}(\mathbf{Y})), \mathbf{Y}, \Gamma_0). \tag{A.14}$$

The economy is populated by a unit mass of households, a representative firm, a government, and a central bank.

Phillips Curve $\mathcal{K}(\mathbf{Y})$. We assume that the final good price P_t is flexible and the nominal wage W_t is sticky. Specifically, wage inflation $\pi_t^w \equiv W_t/W_{t-1} - 1$ follows the textbook New Keynesian wage Phillips curve

$$\pi_t^w(1+\pi_t^w) = \kappa_w \left(\mu_w \frac{v'(N_t)}{u'(C_t)} - 1 \right) + \mathbb{E}_t \left[\pi_{t+1}^w (1+\pi_{t+1}^w) \right], \tag{A.15}$$

where $\kappa_w > 0$ is the slope of the Phillips curve, and $\mu_w > 1$ is the desired markup of the seller of labor services, $v'(N_t)$ is the marginal disutility of labor, and $u'(C_t)$ is the marginal utility of consumption for a hypothetical representative agent. That is, the term in parentheses is the percentage deviation of the marginal rate of substitution between hours and consumption from its steady state. Using the MRS of a hypothetical representative agent instead of the average MRS of the heterogeneous agents simplifies the derivation of the intertemporal Keynesian cross.

Writing (A.15) as $\mathcal{K}(\mathbf{Y})$ requires a few more steps. First, let the final good be produced by a competitive firm with linear technology $Y_t = N_t$. Since the nominal profit of the firm is $P_tY_t - W_tN_t$, the equilibrium price level must be $P_t = W_t$ for all t. This implies that wage inflation equals price inflation $\pi_t^w = P_t/P_{t-1} - 1 = \pi_t$. Using these results and goods market clearing $Y_t = C_t$, the Phillips curve (A.15) under perfect foresight can indeed be written as $\boldsymbol{\pi} = \mathcal{K}(\mathbf{Y})$.

Consumption Function $\mathcal{C}(\boldsymbol{\tau}, \mathbf{r}, \mathbf{Y}, \Gamma_0)$. The household sector is an instance of the SIM model (1)-(3) in which income is after-tax labor income $y_{i,t} = (1-\tau_t)\frac{W_t}{P_t}N_t z_{i,t}$ with real wage W_t/P_t, hours N_t, and idiosyncratic productivity $z_{i,t}$ that follows an exogenous Markov process and has a mean of 1. Given the production function $Y_t = N_t$ and the result $\frac{W_t}{P_t} = 1$, the only endogenous variables entering the household block are $\{\boldsymbol{\tau}, \mathbf{r}, \mathbf{Y}\}$. This proves that the aggregate consumption function can be written as $\mathcal{C}(\boldsymbol{\tau}, \mathbf{r}, \mathbf{Y}, \Gamma_0)$.

Real Rate Function $\mathcal{R}(\boldsymbol{\pi})$. Follows immediately from combining the Taylor rule $i_t = r_{ss} + \phi_\pi \pi_t$ with the Fisher equation $r_t = i_t - \mathbb{E}_t[\pi_{t+1}]$.

B Quantitative Model

B.1 Financial Intermediary

The financial intermediary chooses S_t, B_t, A_t and N_t^{FI} to maximize $\mathbb{E}_t(1 + r_{t+1}^N)$ subject to the constraints (13) to (15).

Asset Returns. By combining the constraints with the definition of the return on net worth after distributions and taking the conditional expectation, one obtains:

$$\mathbb{E}_t(1 + r_{t+1}^N) = \mathbb{E}_t \frac{(D_{t+1} + p_{t+1}^S)S_t + (1 + \delta_B q_{t+1}^B)B_t - (1 + r_t^A + \xi)A_t - \mathcal{D}_t^{FI}}{p_t^S S_t + q_t^B B_t - A_t} \tag{B.1}$$

The first-order conditions are

$$0 = \mathbb{E}_t(D_{t+1} + p_{t+1}^S)N_t^{FI} - \mathbb{E}_t N_{t+1}^{FI} p_t^S, \tag{B.2}$$
$$0 = \mathbb{E}_t(1 + \delta_B q_{t+1}^B)N_t^{FI} - \mathbb{E} N_{t+1}^{FI} q_t^B, \tag{B.3}$$
$$0 = -(1 + r_t^A + \xi)N_t^{FI} + \mathbb{E}_t N_{t+1}^{FI}. \tag{B.4}$$

Together with the definition of $1 + r_{t+1}^N$, the above conditions imply

$$\mathbb{E}_t(1 + r_{t+1}^N) = \mathbb{E}_t \frac{D_{t+1} + p_{t+1}^S}{p_t^S} \tag{B.5}$$
$$= \mathbb{E}_t \frac{1 + \delta_B q_{t+1}^B}{q_t^B} \tag{B.6}$$
$$= 1 + r_t^A + \xi, \tag{B.7}$$

which are the no-arbitrage relationships shown in Section 3.2.

Stability of Balance Sheet. Equation (13) to (15) imply that the law of motion of net worth is

$$\begin{aligned} N_{t+1}^{FI} = &\left[D_{t+1} + p_{t+1}^S - (1 + r_t^A + \xi)p_t^S\right]S_t \\ &+ \left[1 + \delta_B q_{t+1}^B - (1 + r_t^A + \xi)q_t^B\right]B_t \\ &+ (1 + r_t^A + \xi)N_t^{FI} - \mathcal{D}^{FI} - \phi(N_t^{FI} - N^{FI}). \end{aligned} \tag{B.8}$$

Taking conditional expectations and using the asset pricing relationships (B.5) to (B.7) gives

$$\mathbb{E}_t N_{t+1}^{FI} = (1 + r_t^A + \xi)N_t^{FI} - \mathcal{D}^{FI} - \phi(N_t^{FI} - N^{FI}), \tag{B.9}$$

or equivalently

$$\mathbb{E}_t \Delta N_{t+1}^{FI} = (r_t^A + \xi) N_t^{FI} - \mathcal{D}^{FI} - \phi(N_t^{FI} - N^{FI}). \tag{B.10}$$

Thus, a steady state exists only if

$$\mathcal{D}^{FI} = (r^A + \xi) N^{FI}.$$

Slightly expanding equation (B.10), plugging in for $\mathcal{D}^{FI}$, and collecting terms yields

$$\mathbb{E}_t \Delta N_{t+1}^{FI} = (r_t^A - r^A) N_t^{FI} + (r^A + \xi - \phi)(N_t^{FI} - N^{FI}),$$

which shows that, whenever $r_t^A = r^A$, net worth converges to its steady state value from above or below if $\phi > r^A + \xi$.

B.2 Capital Producer

The capital producer's problem can be stated as

$$\begin{aligned} p_t^K(K_{t-1}, I_{t-1}) = r_t^K K_{t-1} + \min_{Q_t} \max_{I_t, K_t} \Bigg\{ -I_t + \frac{1}{1+r_t} \mathbb{E}_t \Big[p_{t+1}^K(K_t, I_t) \Big] \\ + Q_t \Bigg[(1-\delta_K) K_{t-1} + \left[1 - \Phi_I\left(\frac{I_t}{I_{t-1}} \right) \right] I_t - K_t \Bigg] \Bigg\}, \end{aligned} \tag{B.11}$$

where Q_t is the Lagrangian multiplier attached to the law of motion of the capital stock. The first-order condition for investment is given by

$$0 = -1 + \mathbb{E}_t \frac{p_{I,t+1}^K(K_t, I_t)}{1+r_t} + Q_t \left[1 - \Phi_I\left(\frac{I_t}{I_{t-1}} \right) - \Phi_I'\left(\frac{I_t}{I_{t-1}} \right) \frac{I_t}{I_{t-1}} \right], \tag{B.12}$$

where

$$p_{I,t}^K(K_{t-1}, I_{t-1}) = Q_t \Phi_I'\left(\frac{I_t}{I_{t-1}} \right) \left(\frac{I_t}{I_{t-1}} \right)^2. \tag{B.13}$$

Hence combining (B.12) and (B.13) yields equation (19) in the main text.

The first-order condition of capital is

$$Q_t = \frac{1}{1+r_t} \mathbb{E}_t \Big[p_{K,t+1}^K(K_t, I_t) \Big], \tag{B.14}$$

with

$$p_{K,t}^K(K_{t-1}, I_{t-1}) = r_t^K + (1-\delta_K) Q_t. \tag{B.15}$$

Combining (B.14) and (B.13) yields equation (20).

B.3 Labor Market

This section derives the bargaining surplus of the labor agency and the union and provides details on the Nash bargaining solution.

Labor Agency's Problem. The profit maximization problem solved by the labor agency is

$$\mathcal{J}_t(N_{t-1}) = \max_{N_t, v_t} \left\{ (h_t - w_t)N_t - (\kappa_v + \kappa_h q_t)v_t - \Phi_w(w_t, w_{t-1})N_t + \frac{1}{1+r_t}\mathbb{E}_t[\mathcal{J}_{t+1}(N_t)] \right\} \tag{B.16}$$

$$\text{s.t. } N_t = (1-s)N_{t-1} + q_t v_t, \tag{B.17}$$

$$\Phi_w(w_t, w_{t-1}) = \frac{\psi_w}{2}\left(\frac{w_t}{w_{t-1}} - 1\right)^2, \tag{B.18}$$

Let J_t be the Lagrange multiplier on the first constraint—the shadow value of an additional worker with average productivity. Then, the first-order and envelope conditions are

$$0 = h_t - w_t - \Phi_w(w_t, w_{t-1}) - J_t + \frac{1}{1+r_t}\mathbb{E}_t \mathcal{J}_{N,t+1}(N_t), \tag{B.19}$$

$$0 = -(\kappa_v + \kappa_h q_t) + J_t q_t, \tag{B.20}$$

$$\mathcal{J}_{N,t}(N_{t-1}) = (1-s)J_t. \tag{B.21}$$

Combining the equations (B.19) to (B.21) yields equation (24). Equation (25) then follows from equation (B.20).

Union Surplus. The union's valuation of the marginal match depends on the value of an employed worker $\mathcal{W}_t$ and the value of an unemployed worker $\mathcal{U}_t$,

$$\mathcal{W}_t = w_t + \frac{1}{1+r_t}\mathbb{E}_t\{[1 - s(1-f_{t+1})]\mathcal{W}_{t+1} + s(1-f_{t+1})\mathcal{U}_{t+1}\}; \tag{B.22}$$

$$\mathcal{U}_t = \omega^{UI} w + \frac{1}{1+r_t}\mathbb{E}_t[f_{t+1}\mathcal{W}_{t+1} + (1-f_{t+1})\mathcal{U}_{t+1}]. \tag{B.23}$$

Hence, H_t is given by

$$\begin{aligned} H_t &= \mathcal{W}_t - \mathcal{U}_t \\ &= w_t - \omega^{UI} w + \frac{1-s}{1+r_t}\mathbb{E}_t(1-f_{t+1})H_{t+1}. \end{aligned} \tag{B.24}$$

which is equation (26).

Nash Bargaining. The Nash bargaining solution satisfies $\Omega_t J_t = (1-\Omega_t)H_t$, where

$$\Omega_t \equiv \frac{\eta}{\eta + (1-\eta)(-J_{w,t}/H_{w,t})}, \tag{B.25}$$

$$J_{w,t} = -1 - \Phi_1^w(w_t, w_{t-1}) + \frac{1-s}{1+r_t}\mathbb{E}_t J_{w,t+1}, \tag{B.26}$$

$$J_{w,t+1} = -\Phi_2^w(w_{t+1}, w_t), \tag{B.27}$$

$$H_{w,t} = 1. \tag{B.28}$$

The wage adjustment cost function implies

$$\Phi_1^w(w_t, w_{t-1}) = \psi_w\left(\frac{w_t}{w_{t-1}} - 1\right)\frac{1}{w_{t-1}}; \tag{B.29}$$

$$\Phi_2^w(w_{t+1}, w_t) = -\psi_w\left(\frac{w_{t+1}}{w_t} - 1\right)\frac{w_{t+1}}{w_t^2}. \tag{B.30}$$

In a steady state, $-J_w = H_w = 1$ and therefore $\Omega = \eta$.

B.4 Intermediate Good Producers

This section derives the Phillips curve under Rotemberg-pricing allowing for price indexation. The problem can be written recursively as

$$\begin{aligned} X_t(P_{jt-1}) = \max_{P_{jt}, K_{jt-1}, N_{jt}} \Bigg\{ & \left(\frac{P_{jt}}{P_t}\right)^{1-\epsilon_p} Y_t - r_t^K K_{jt-1} - h_t N_{jt} - \Psi \\ & - \frac{\chi_p}{2}\left[\log(P_{jt}/P_{jt-1}) - \log\left(\Pi_{t-1}^{\iota_p}\Pi^{1-\iota_p}\right)\right]^2 Y_t + \frac{1}{1+r_t}\mathbb{E}_t X_{t+1}(P_{jt}) \Bigg\} \end{aligned} \tag{B.31}$$

$$\text{s.t.} \quad \left(\frac{P_{jt}}{P_t}\right)^{-\epsilon_p} Y_t = \Theta K_{jt-1}^{\alpha} N_{jt}^{1-\alpha}. \tag{B.32}$$

The corresponding first-order conditions are:

$$\begin{aligned} 0 = {} & (1-\epsilon_p)\left(\frac{P_{jt}}{P_t}\right)^{-\epsilon_p}\frac{Y_t}{P_t} - \chi_p\left[\log(P_{jt}/P_{jt-1}) - \log\left(\Pi_{t-1}^{\iota_p}\Pi^{1-\iota_p}\right)\right]\frac{Y_t}{P_{jt}} \\ & + \lambda_{jt}\epsilon_p\left(\frac{P_{jt}}{P_t}\right)^{-\epsilon_p-1}\frac{Y_t}{P_t} + \frac{1}{1+r_t}\mathbb{E}_t X_{t+1}'(P_{jt}); \end{aligned} \tag{B.33}$$

$$0 = -r_t^K + \lambda_{jt}\alpha\Theta K_{jt-1}^{\alpha-1} N_{jt}^{1-\alpha}; \tag{B.34}$$

$$0 = -h_t + \lambda_{jt}(1-\alpha)\Theta K_{jt-1}^{\alpha} N_{jt}^{-\alpha}. \tag{B.35}$$

In the above equations, λ_{jt} is the Lagrange multiplier associated with the constraint. The envelope condition is

$$X_t'(P_{jt-1}) = \chi_p \left[\log(P_{jt}/P_{jt-1}) - \log\left(\Pi_{t-1}^{\iota_p}\Pi^{1-\iota_p}\right)\right] \frac{Y_t}{P_{jt-1}}. \tag{B.36}$$

The first-order conditions for capital and labor imply that all firms have the same capital-to-labor ratio and hence the same multiplier λ_t, which can be interpreted as real marginal cost,

$$mc_t \equiv \lambda_t = \frac{(r_t^K)^\alpha (h_t)^{1-\alpha}}{\Theta} \left(\frac{1}{\alpha}\right)^\alpha \left(\frac{1}{1-\alpha}\right)^{1-\alpha}. \tag{B.37}$$

In a symmetric equilibrium with $P_{jt} = P_t \; \forall j$, combining the first-order condition (B.33) with the envelope condition (B.36) and rearranging terms yields the Phillips curve

$$\begin{aligned} \log\left(\Pi_t\right) - \log\left(\Pi_{t-1}^{\iota_p}\Pi^{1-\iota_p}\right) &= \frac{\epsilon_p}{\chi_p}\left(mc_t - \frac{\epsilon_p - 1}{\epsilon_p}\right) \\ &\quad + \frac{1}{1+r_t}\mathbb{E}_t\left[\log\left(\Pi_{t+1}\right) - \log\left(\Pi_t^{\iota_p}\Pi^{1-\iota_p}\right)\right]\frac{Y_{t+1}}{Y_t}. \end{aligned} \tag{B.38}$$

The last equation involves level deviations of real marginal cost from steady state. It's useful to rewrite it in terms of percentage deviations for the slope κ_p to be directly comparable to standard loglinearized Phillips curves.

$$\frac{\epsilon_p}{\chi_p}\left(mc_t - \frac{\epsilon_p - 1}{\epsilon_p}\right) = \underbrace{\frac{\epsilon_p}{\chi_p}\frac{\epsilon_p - 1}{\epsilon_p}}_{\kappa_p}\left(\frac{\epsilon_p}{\epsilon_p - 1}mc_t - 1\right) \tag{B.39}$$

C Calibration

Our calibration of the search and matching block follows Christiano et al. (2016). The calibration of the rest of the model is close to Auclert et al. (2020). Although we do not estimate the transition-specific parameters, we show that our model implies cumulative government spending multipliers that match well the estimates of Ramey and Zubairy (2018).

Households. We assume a constant relative risk aversion (CRRA) period utility function, $u(c) = c^{1-1/\sigma}/(1-1/\sigma)$ with an elasticity of intertemporal substitution of $\sigma = 0.5$. We set the quarterly steady-state real interest rate r to 0.5 percent, implying an annual rate of 2 percent. This corresponds to the real return on capital and government bonds. The real return on deposits is $r^A = r - \xi$ and we set the quarterly intermediation spread ξ to 1 percent. That is, the real return on deposits is -2 percent per annum. The time preference rate β is 0.95 which implies a liquid wealth to annual GDP ratio of $A/(4Y) = 0.24$, very close to the value of 0.26 targeted by Kaplan et al. (2018). For the productivity process, we use the quarterly, discrete-time version of the leptokurtic process estimated by Kaplan et al. (2018), with one modification: we scale down the variance of innovations by $(1 - 0.181)^2$, where 0.181 is the degree of tax progressivity in Heathcote et al. (2017).

Search and Matching. The calibration of this block follows Christiano et al. (2016). The steady-state unemployment rate is 5 percent. We set the job finding rate to $f = 0.6$ and the vacancy filling rate to $q = 0.7$, informed by CPS (for men aged 25–54) and JOLTS data, respectively. These choices imply a quarterly separation rate of $s = 0.09$ which is also consistent with the separation rate of prime-age men in the CPS data. We set the elasticity of the matching function to $\alpha_m = 0.5$ and back out the matching efficiency as a residual. The unemployment insurance replacement rate is $\omega^{UI} = 0.5$, the most common value across U.S. states. We calibrate the bargaining power of the union such that the total cost of filling a job $(\kappa_v + \kappa_h q)v$ is 7 percent of the quarterly wage of the average worker. As in Christiano et al. (2016), the vacancy filling cost κ_h accounts for 94 percent of the total cost.

Supply. The calibration of this block follows Auclert et al. (2020). Total factor productivity Θ is chosen such that quarterly output Y is normalized to 1. We set B such that government debt is 46 percent of annual output, corresponding to domestic holdings. We choose the coupon rate δ_B to match the average duration of U.S. government debt of 5 years. Government spending G is set to 16 percent of output, which implies a tax rate of $\tau = 0.24$. We choose a depreciation rate of $\delta_K = 0.083$ annually and calibrate the capital share α to match a quarterly capital to output ratio of 8.92. These choices imply a steady-state labor share of 62 percent in line with U.S. data. The fixed cost Ψ is calibrated to let total wealth $p^S + q^B B$ be equal to 382 percent of annual output given a standard value for the elasticity of substitution of $\epsilon_p = 7$.

Transition-Specific Parameters. As dicussed in the main text, we set the slope of the Phillips curve to $\kappa_p = 0.05$ based on Gagliardone et al. (2023) and allow for a moderate amount of indexation, $\iota_p = 0.2$. We set the parameters of the Taylor rule to conventional values, with an inflation coefficient of $\phi_\pi = 1.5$ and an inertia parameter of $\phi_r = 0.8$. We calibrate the investment adjustment cost to $\psi_I = 1.8$ which implies that the semi-elasticity of investment to the real rate is -5% at an annual frequency, in line with the estimates of Koby and Wolf (2020) and He et al. (2022). The real wage adjustment cost ψ_w is 100 which implies that wages are moderately stickier than wages. The steady state dividend yield from the perspective of the households $\mathcal{D}^{FI}/p$ is about 2.1 percent annually. We set the parameter of the intermediary's distribution rule ϕ to 0.01, which ensures balance sheet stability and implies that the annualized dividend yield rises by about 0.4 percentage points for each 10 percentage points that intermediary net worth exceeds its steady state. These values are consistent with the average dividend yield of the S&P 500 that lay in the range of about 1 to 3 percent over the last decades. Table C.1 summarizes the calibration.

Table C.1: Calibration

Description	Parameter/Target	Value
Elasticity of intertemporal substitution	σ	0.5
Real interest rate	r	0.005
Intermediation cost	ξ	0.01
Time preference rate	β	0.95
Unemployment rate	U	0.05
Job finding rate	f	0.6
Vacancy filling rate	q	0.7
Searchers' share of matching function	α_m	0.5
Replacement ratio	ω^{UI}	0.5
Search cost-to-wage ratio	$(\kappa_v + \kappa_h q)v/w$	0.07
Vacancy cost share of search cost	$\kappa_v/(\kappa_v + \kappa_h q)$	0.06
Output	Y	1
Government debt-to-GDP ratio	$q_B B/4Y$	0.46
Maturity of government bond	$1/\delta_B$	20
Share of government spending	G/Y	0.16
Depreciation rate	δ_K	0.083/4
Capital-to-output ratio	$K/4Y$	8.92/4
Wealth-to-GDP ratio	$(p^S + q^B B)/4Y$	3.82
Elasticity of substitution	ϵ_p	7
Slope of the Phillips curve	κ_p	0.1
Indexation in price setting	ι_p	0.2
Taylor rule coefficient on inflation	ϕ_π	1.5
Taylor rule coefficient on inertia	ϕ_i	0.8
Investment adjustment cost	ψ_I	1.8
Real wage adjustment cost	ψ_w	100
Payout rate of retained earnings	ϕ	0.01

D Application

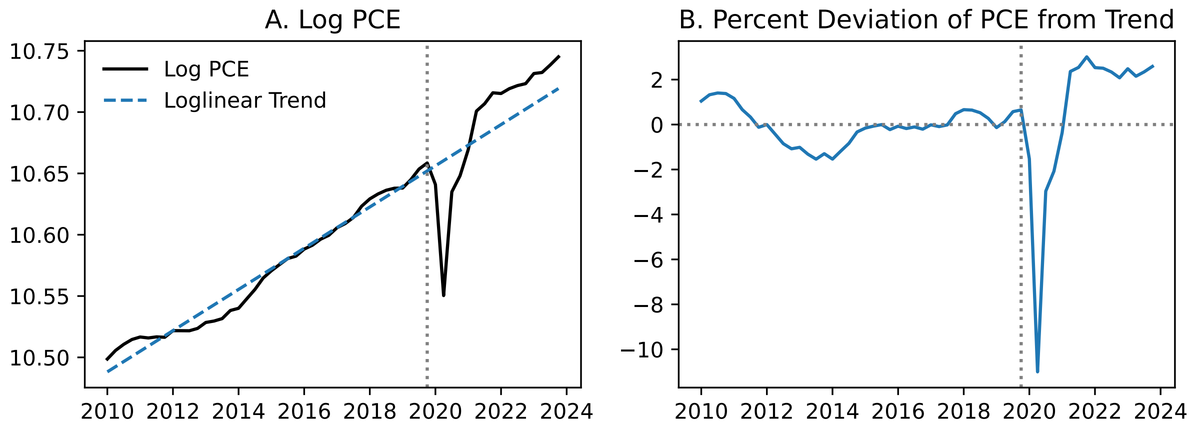

Figure D.1: Personal Consumption Expenditure

Notes: Panel A shows log per-capita PCE and its loglinear trend estimated over the 2010-2019 period. Panel B shows the percent deviation of per-capita PCE from trend. The vertical line indicates the end of the period used to estimate the trend.

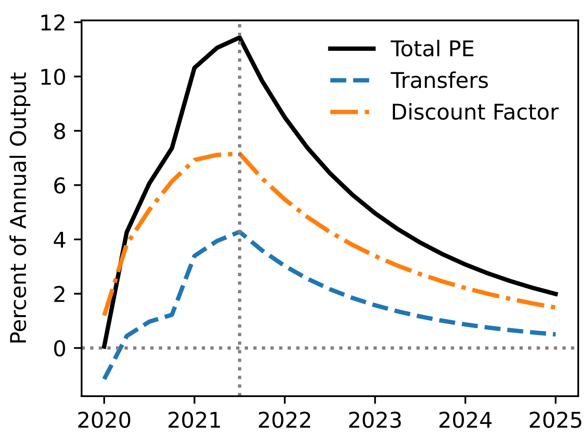

Figure D.2: Contributions to Excess Savings in Partial Equilibrium

Notes: The figure shows the partial-equilibrium components of excess savings based on equation (46). The vertical line marks the third quarter of 2021, the end of the buildup period.