2026 Volume: 02 Issue: 06 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-17

CONSUMPTION SMOOTHING OR DEBT TRAPS? A MACROECONOMIC ASSESSMENT OF BUY-NOW-PAY-LATER ON THE MARGINAL PROPENSITY TO CONSUME ACROSS INCOME QUINTILES

*[1]Reagan J Ochieng Ombuor, [2]Yasin Ghabon

[1]Department of Accounting and Finance Maseno University.

[2]Department of Economics Maseno University.

Article Received: 9 May 2026
Article Revised: 29 May 2026
Published on: 19 June 2026

***Corresponding Author: Reagan J Ochieng Ombuor**
Department of Accounting and Finance Maseno University.

DOI: https://doi-doi.org/101555/ijrpa.7684

ABSTRACT

Buy-Now-Pay-Later (BNPL) has grown from a marginal checkout option into a mainstream short-term credit instrument, with the largest U.S. providers originating over 330 million loans worth more than $45 billion in 2023 alone. Policymakers remain divided on whether the product is best understood as a welfare-improving liquidity technology that allows constrained households to smooth consumption against income shocks, or as an extractive mechanism that monetizes financial fragility through late fees, repayment stacking, and merchant cost pass-through. This paper develops a heterogeneous-agent New Keynesian (HANK) framework augmented with an explicit BNPL contract — a short-duration, often zero-interest instalment loan financed jointly by a merchant discount fee and a state-contingent late fee — to assess the product's effect on the marginal propensity to consume (MPC) across income quintiles. The model is calibrated using stylized targets drawn from published micro-level fintech and survey data, including findings on BNPL-induced spending increases, deteriorating credit-card and overdraft outcomes among new BNPL users, and the merchant-side economics of price discrimination through subsidized credit. Illustrative simulations suggest that BNPL access raises the MPC most sharply for households in the second- and third-income quintiles — those liquidity-constrained but not so distressed as to face binding adoption frictions — while a non-trivial share of the apparent consumption-smoothing gain is offset by fee and price-markup extraction concentrated in the bottom two quintiles. The paper concludes that BNPL is neither a clean consumption-smoothing

technology nor a uniform debt trap, but a redistributive credit instrument whose welfare implications depend critically on the elasticity of merchant cost pass-through and on regulatory treatment of repeat late-fee assessment.

KEYWORDS: Buy Now Pay Later, heterogeneous-agent models, HANK, marginal propensity to consume, household finance, fintech credit, consumption smoothing, merchant cost-shifting.

1. INTRODUCTION

The decade since the 2008 financial crisis produced a wave of household-finance research organized around a deceptively simple statistic: the marginal propensity to consume out of a dollar of transitory income. That research, anchored in heterogeneous-agent incomplete-markets models, established that aggregate consumption dynamics are governed less by the behaviour of a representative household than by the share of the population living "hand-to-mouth," holding little liquid wealth despite sometimes substantial illiquid assets (Kaplan and Violante 2014; Kaplan, Moll, and Violante 2018). Where a household sits in the wealth and income distribution not simply its permanent income — determines how it responds to a stimulus check, a wage cut, or a medical bill.

Buy-Now-Pay-Later credit has emerged directly into this empirical gap. Unlike a credit card, BNPL requires no formal underwriting cycle and is typically structured as a short, fixed-instalment loan — commonly a 25 percent down payment at checkout followed by three biweekly payments — embedded at the point of sale rather than offered as a revolving credit line. The product's defenders describe it as a democratizing liquidity technology: it extends short-duration credit to consumers who are thin-file or subprime by traditional underwriting standards, at an interest rate of zero, allowing them to time-shift consumption without paying the double-digit annual rates associated with credit cards or payday lending. Its critics describe a different mechanism: a checkout-embedded nudge that increases impulsive retail spending, stacks repayment obligations across multiple providers in ways invisible to traditional credit bureaus, and is financed in part by a merchant discount fee that is ultimately passed through into retail prices paid by all consumers, including those who never use BNPL at all.

Both descriptions have empirical support. Using transaction-level panel data, Di Maggio, Katz, and Williams (2022) document that total household spending rises by roughly $130 in the week a consumer first uses BNPL and remains elevated for the following twenty-four

weeks, with the increase concentrated in retail categories and too large to be explained by a standard lifecycle model with liquidity constraints; the authors describe the pattern as a "liquidity flypaper effect," in which retail-earmarked liquidity disproportionately funds retail spending rather than smoothing consumption broadly. Using bank-transaction data covering more than ten million U.S. consumers, deHaan, Kim, Lourie, and Zhu (2024) find that new BNPL users experience a rapid subsequent rise in overdraft charges and in credit-card interest and late fees relative to comparable non-users, a pattern they interpret as evidence that BNPL access facilitates overborrowing rather than simple consumption smoothing. On the supply side, Berg, Burg, Keil, and Puri (2024) show, using a randomized in-store rollout, that offering BNPL raises the basket-conversion rate from roughly 70 to 80 percent and increases merchant sales by approximately 20 percent, a pattern they attribute to BNPL functioning as a bundled, subsidized loan that allows merchants to price-discriminate in favour of lower-creditworthiness shoppers — a mechanism financed by a merchant discount fee in the range of three to eight percent of transaction value.

The regulatory record reinforces the ambiguity. The Consumer Financial Protection Bureau's analysis of BNPL borrower profiles finds that BNPL users carry, on average, a subprime credit score and a higher incidence of credit-card revolving balances, delinquency, and use of payday loans, pawn loans, and overdraft than non-users — although many of these differences predate BNPL adoption — and that usage is disproportionately concentrated among Black, Hispanic, and female consumers and among households with annual income between $20,001 and $50,000 (CFPB 2023). At the same time, the Bureau's most recent market-level data spotlight finds that late-fee incidence has been falling, from 5.2 percent of loans in 2022 to 4.1 percent in 2023, with late-fee revenue equal to only 0.18 percent of gross merchandise volume, even as the number of BNPL users grew to 53.6 million and average annual BNPL borrowing per user reached $848 (CFPB 2025). The product, in other words, looks simultaneously like a marginal liquidity lifeline and a slow-moving balance-sheet risk, and the two readings are not obviously reconcilable from descriptive statistics alone.

This paper argues that resolving the tension requires a structural, general-equilibrium approach. A reduced-form regression of spending on BNPL adoption cannot separate a genuine increase in the marginal propensity to consume — the object of interest for fiscal and monetary transmission — from a pure intertemporal reallocation of existing consumption, nor can it quantify the share of any apparent welfare gain that is clawed back through late fees, interest-bearing rollovers into other credit products, or a higher general price level induced by merchant cost-shifting. We therefore embed an explicit BNPL contract into a heterogeneous-

agent New Keynesian (HANK) model in the tradition of Kaplan, Moll, and Violante (2018), calibrated to income-quintile-specific facts about BNPL adoption, borrowing limits, and fee incidence drawn from the empirical literature summarized above. The framework allows us to compute the quintile-specific MPC with and without access to BNPL, to decompose the welfare change into an insurance (smoothing) component and an extraction (fee and price markup) component, and to characterize the conditions under which the product is welfare-improving for the households who use it most.

The remainder of the paper proceeds as follows. Section 2 reviews the relevant literature on heterogeneous-agent macroeconomics and the empirical BNPL literature. Section 3 presents the model. Section 4 describes the calibration strategy and data sources. Section 5 outlines the computational methodology. Section 6 reports illustrative simulation results. Section 7 develops the welfare decomposition. Sections 8 through 10 discuss the results, draw out policy implications, and note limitations, and Section 11 concludes.

2. Literature Review and Theoretical Motivation

2.1 BNPL Adoption and Spending Effects: Empirical Evidence

The empirical BNPL literature has expanded rapidly since 2022, largely on the strength of proprietary transaction-level data made available to a small number of research teams. Three findings recur across studies and motivate the model developed below.

First, BNPL access causally raises total spending, not merely the composition of spending. Di Maggio, Katz, and Williams (2022) find that the spending increase following first-time BNPL use holds for both liquidity-constrained and seemingly unconstrained consumers, and that it is too large relative to typical calibrated buffer-stock models to be explained by standard intertemporal substitution; the authors propose a mental-accounting channel in which retail-earmarked liquidity is treated by consumers as a distinct budget line rather than fungible income, so that the marginal propensity to consume out of BNPL-provided liquidity is unusually high. This is, in effect, a direct empirical estimate of a high "MPC out of BNPL liquidity," which a structural model should be able to reproduce as an equilibrium object rather than impose by assumption.

Second, BNPL adoption is associated with subsequent deterioration in other measures of household balance-sheet health. deHaan, Kim, Lourie, and Zhu (2024) document that new BNPL users see overdraft charges rise by roughly four percent and credit-card interest and late fees rise by one to two percent relative to a matched sample of non-adopters, using an instrumental-variables design based on pre-adoption shopping patterns at retailers that later

introduce BNPL to argue for a causal interpretation. The Federal Reserve's analysis of the 2023 Survey of Household Economics and Decision making similarly finds that BNPL is disproportionately used by financially fragile households and is associated with elevated measures of financial strain (Larrimore et al. 2024). These findings are consistent with — though not definitive proof of — a debt-trap channel operating alongside the smoothing channel documented above.

Third, the supply side of the market is structured around price discrimination rather than pure liquidity provision. Berg, Burg, Keil, and Puri (2024) model BNPL as a bundle of a retail good with a subsidized, often zero-interest, loan, financed by a merchant discount fee in the three-to-eight-percent range; merchants are willing to pay this fee because BNPL disproportionately converts marginal, lower-creditworthiness shoppers into buyers and induces higher-value purchases, generating a sales increase the authors estimate at approximately twenty percent in their randomized setting. Because this fee is a cost of doing business for the merchant, the standard two-sided-market logic familiar from the interchange-fee literature implies that at least part of it is recovered through the general price level rather than absorbed entirely by the merchant's margin — a channel through which non-BNPL users can bear some of the cost of a product they do not use, an externality this paper labels "merchant cost-shifting."

2.2 Heterogeneous-Agent Models and the Marginal Propensity to Consume

The theoretical backbone for evaluating these facts is the heterogeneous-agent incomplete-markets framework originating with Bewley's unpublished lecture notes and formalized by Huggett (1993) and Aiyagari (1994), in which ex-ante identical households face uninsurable idiosyncratic income risk and choose consumption and saving subject to a borrowing constraint. Because the borrowing constraint binds for a subset of the population at any point in time, the model generates a distribution of marginal propensities to consume rather than a single representative-agent value, with constrained and low-liquid-wealth households exhibiting MPCs far above the smooth, near-zero values implied by permanent-income models with complete markets.

Kaplan and Violante (2014) extend this logic to a two-asset setting — one liquid, one illiquid and costly to access — and show that a sizeable share of households can be simultaneously "wealthy" on a net-worth basis and "hand-to-mouth" on a liquidity basis, generating realistic average MPCs out of transitory income even among households who are not poor in a traditional sense. Kaplan, Moll, and Violante (2018) embed this household block into a New Keynesian general-equilibrium environment — the HANK model — to show that the

distribution of MPCs, rather than the aggregate intertemporal elasticity of substitution, is the primary determinant of how monetary and fiscal policy transmit to aggregate demand. This is the natural workhorse for the present question: BNPL is, functionally, a new liquid asset/liability margin available disproportionately to households who would otherwise be liquidity-constrained, and a HANK-style model is built precisely to translate such a margin into quintile-specific MPC predictions.

2.3 The Debt-Trap Hypothesis: Fees, Repayment Stacking, and Cost-Shifting

The "debt trap" framing draws on an older literature evaluating high-cost short-term credit, including payday and overdraft lending, in which a product marketed as smoothing a single income shock instead generates a sequence of repeat borrowing and fee assessment that leaves the household worse off than under the counterfactual of no access at all. The BNPL-specific evidence is more equivocal than in that literature: CFPB (2025) data show late-fee incidence and revenue declining as the market matures and underwriting tightens, even as adoption grows, suggesting that providers have an incentive — whether regulatory, reputational, or simply commercial — to limit fee-driven attrition. At the same time, CFPB (2023) profiles show BNPL borrowers carry materially worse credit outcomes on other products than non-borrowers, and deHaan et al. (2024) provide a causal channel by which BNPL adoption itself, rather than pre-existing financial fragility, contributes to that deterioration through overdraft and credit-card spill overs. The model below treats both channels — direct BNPL late fees and indirect spill over costs on other credit products — as components of the "extraction" side of the welfare ledger, alongside the merchant cost-shifting channel identified by Berg et al. (2024).

3. A HANK Model with BNPL

3.1 Environment

Time is discrete, indexed by quarters $t = 0, 1, 2, \ldots$. The economy is populated by a continuum of households indexed by i, each endowed with idiosyncratic labor productivity z_{it}, which follows a stationary AR (1) process in logs, $\log z_{it} = \rho_z \log z_{it-1} + \varepsilon_{it}$, with $\varepsilon_{it} \sim N(0, \sigma_z^2)$. Households are sorted at each date into five income quintiles Q1 (lowest) through Q5 (highest) based on their position in the stationary cross-sectional distribution of z_{it}. There is a single liquid asset, a one-period real bond b_{it} paying gross return $(1+r)$, subject to a quintile-dependent natural and BNPL-augmented borrowing limit described below. A representative final-goods producer combines a continuum of differentiated retail goods produced by monopolistically competitive merchants subject to Calvo price-

adjustment frictions, generating the New Keynesian Phillips curve that closes the model in general equilibrium. A government issues the bond supply, sets a (possibly zero) transfer schedule, and a monetary authority sets the nominal rate according to a Taylor rule, as in the standard HANK environment of Kaplan, Moll, and Violante (2018).

3.2 The Household Problem

Each household solves the recursive problem

$$V(b, \ell, z) = \max_{c, b'} \{ u(c) + \beta E[V(b', \ell', z') \mid z] \}$$

subject to

$$c + b' + \kappa(\ell, \ell') \le y(z) + (1+r) b - \varphi \cdot 1\{default\}_{it}, \quad b' \ge -\underline{b}(z, \ell'),$$

where $u(c) = c^{(1-\gamma)}/(1-\gamma)$ is standard CRRA utility, $y(z) = w \cdot z$ is labor income at the competitive wage w, and $\ell_{it} \in [0, \bar{L}(z)]$ is the household's outstanding BNPL instalment balance, a second, short-duration state variable that evolves according to the repayment schedule $\kappa(\ell, \ell')$: a household that draws down a new BNPL purchase $x_{it} \le \bar{L}(z_{it}) - \ell_{it}$ in period t commits to a fixed sequence of instalment payments over the following two quarters at zero interest, financed jointly by the household's own repayment and, on the supply side, by the merchant discount fee described in Section 3.3.

The borrowing limit $\underline{b}(z, \ell')$ is the sum of a standard natural borrowing limit and the additional liquidity made available through BNPL access, $\underline{b}(z, \ell') = \underline{b}_{nat}(z) + \bar{L}(z) - \ell'$; this is the central object of interest, since it is precisely the margin through which BNPL relaxes the liquidity constraint facing low- and middle-quintile households who would otherwise be unable to borrow against future income at any price.

A household that draws on BNPL faces an exogenous, state-dependent probability of a late-payment event, $\pi(b_{it}, z_{it})$, calibrated to be decreasing in liquid wealth and increasing for households whose existing instalment burden ℓ_{it} is high relative to income — capturing the repayment-stacking mechanism documented qualitatively in the survey evidence of Larrimore et al. (2024) and the credit-bureau evidence of CFPB (2023). Conditional on a late-payment event, the household incurs a fixed fee φ and, with probability ψ, an additional spill over cost λ representing the elevated overdraft and credit-card costs documented by deHaan et al. (2024) for households whose BNPL obligations interact with other credit lines.

3.3 The BNPL Contract and the Cost-Shifting Wedge

BNPL is supplied by a competitive fintech sector that earns a merchant discount fee $m \in [0.03, 0.08]$ per dollar of transaction value, consistent with the range documented by Berg, Burg, Keil, and Puri (2024), plus expected late-fee revenue $\varphi \cdot \pi(b,z)$. Following the merchant's-perspective logic of that paper, the BNPL technology is offered because it

functions as a price-discrimination device: it converts a fraction of marginal, otherwise credit-constrained shoppers into purchasers and raises average basket size, which is the mechanism generating the roughly twenty percent sales increase the authors estimate.

The merchant's cost of offering BNPL, net of the additional sales it generates, is passed through to the aggregate retail price level with elasticity $\theta \in [0, 1]$: $P = P^* \cdot (1 + \theta \cdot m \cdot \chi)$, where χ is the BNPL-financed share of aggregate retail transactions and P^* is the price level that would prevail absent BNPL. When $\theta > 0$, every household — including the $(1-\chi)$ share of transactions not financed through BNPL, and including households who never adopt the product — pays a higher retail price, which is the formal expression of the "merchant cost-shifting" channel motivating the paper's title. The polar case $\theta = 0$ corresponds to merchants fully absorbing the BNPL fee out of markup, in which case BNPL is a pure transfer from merchant profit to consumer surplus with no aggregate price effect; the polar case $\theta = 1$ corresponds to full pass-through, in which BNPL functions as a regressive, economy-wide sales tax whose proceeds subsidize the liquidity of BNPL-adopting households.

3.4 Firms and General Equilibrium

Retail firms are monopolistically competitive and set prices subject to a quadratic (Rotemberg) price-adjustment cost, generating a standard New Keynesian Phillips curve linking inflation to real marginal cost, the latter now inclusive of the BNPL fee wedge described above. Labor is supplied inelastically by households at the household-specific productivity z_{it} and demanded by firms at the competitive wage w, which clears the labor market each period. A government budget constraint and monetary policy rule close the model exactly as in Kaplan, Moll, and Violante (2018); we do not modify the supply side or monetary block of that framework beyond the price-level wedge described above, since the contribution of this paper is the household-side BNPL contract and its cross-sectional implications, not a new monetary transmission mechanism per se.

3.5 Definition of Equilibrium

A stationary general equilibrium is a set of policy functions $\{c(b,\ell,z), b'(b,\ell,z), \ell'(b,\ell,z)\}$, a stationary joint distribution of (b, ℓ, z), prices (w, r, P), and aggregate quantities such that (i) households optimize given prices, (ii) firms set prices optimally given demand and the BNPL-augmented marginal cost wedge, (iii) the goods, labor, and bond markets clear, and (iv) the government budget constraint and monetary rule are satisfied. The marginal propensity to consume for a household in income quintile Q at date t is defined as $MPC_Q = E[\Delta c_{it} \mid \Delta y_{it}, i \in Q] / E[\Delta y_{it} \mid i \in Q]$ in response to a small, transitory, unanticipated

income shock, computed both in the calibrated economy with BNPL access and in a counterfactual economy in which $\bar{L}(z) = 0$ for all z, holding all other parameters fixed.

4. Data and Calibration

4.1 Micro-Level Fintech Data Sources

The calibration draws on four classes of publicly available evidence rather than a single proprietary dataset, reflecting the paper's methodological contribution as a model and calibration strategy that can subsequently be re-estimated once loan-level fintech data are available to the authors under a data-use agreement. The targets are drawn from: (i) the CFPB's market-level data spotlight on the six largest "pay-in-four" BNPL providers, covering loan volume, user counts, average loan size, and late-fee incidence for 2019–2023 (CFPB 2025); (ii) the CFPB's borrower-profile analysis linking BNPL use to credit-bureau outcomes, including credit-score distributions and the income and demographic composition of the user base (CFPB 2023); (iii) the transaction-level estimates of spending response and persistence in Di Maggio, Katz, and Williams (2022) and of overdraft and credit-card spill overs in deHaan, Kim, Lourie, and Zhu (2024); and (iv) the merchant-fee and sales-response estimates in Berg, Burg, Keil, and Puri (2024).

4.2 Calibration Targets and Strategy

Standard preference and aggregate parameters — the discount factor β, risk aversion γ, and the income process parameters ρ_z and σ_z — are calibrated to conventional values from the HANK literature (Kaplan, Moll, and Violante 2018), targeting an annual real interest rate of approximately one percent and a stationary wealth distribution broadly consistent with U.S. Survey of Consumer Finances moments. The BNPL-specific parameters are calibrated quintile-by-quintile to match, as closely as the available aggregate evidence permits, (a) the observed concentration of BNPL adoption among middle- and lower-income, subprime-credit households documented by CFPB (2023); (b) an aggregate average annual BNPL borrowing volume per active user of $848 and an average per-transaction loan size of $135 (CFPB 2025); (c) a late-fee incidence of approximately four percent of outstanding loans and late-fee revenue equal to roughly 0.18 percent of gross merchandise volume (CFPB 2025); and (d) the qualitative spending and spillover-cost magnitudes reported by Di Maggio, Katz, and Williams (2022) and deHaan, Kim, Lourie, and Zhu (2024).

Because none of the underlying public sources report a clean income-quintile breakdown of adoption, borrowing limits, or fee incidence, the quintile-specific calibration in Table 1 below is a stylized parameterization constructed to be directionally consistent with the documented

patterns — adoption and credit fragility concentrated in Q2–Q3, lower adoption in Q1 due to thinner underwriting eligibility and lower e-commerce participation, and minimal adoption and fee exposure in Q4–Q5 — rather than a set of point estimates drawn directly from loan-level microdata. This distinction is made explicit because it bears directly on how the simulation results in Section 6 should be interpreted.

4.3 Calibrated Parameters

Table 1. Quintile-Specific Calibration Targets. (Illustrative)

Quintile	Adoption rate	Avg. credit limit $\bar{L}$ (rel. to qtrly income)	Late-fee incidence π	Merchant fee pass-through χ-weight	Spillover prob. ψ
Q1 (lowest)	0.18	0.10	0.07	0.10	0.35
Q2	0.34	0.16	0.06	0.22	0.30
Q3	0.31	0.14	0.04	0.28	0.20
Q4	0.19	0.11	0.02	0.25	0.10
Q5 (highest)	0.09	0.08	0.01	0.15	0.04

Aggregate fee parameters are set at φ = \$8 per late-payment event, consistent with the typical flat BNPL late fee reported in market commentary, and m = 0.05 as the central merchant discount rate within the three-to-eight-percent range documented by Berg, Burg, Keil, and Puri (2024); the price pass-through elasticity θ is varied between 0 and 1 in Section 6.3 as the central object of comparative-statics interest, since it is the least directly observable parameter in the model and the one to which the policy conclusions are most sensitive.

5. Computational Methodology

The household problem is solved by value-function iteration on a discretized grid over liquid assets b, the BNPL installment state ℓ, and the productivity process z (approximated via the Rouwenhorst method), producing stationary policy functions and the associated joint distribution via the non-stochastic histogram method standard in the Aiyagari–Bewley–Huggett computational literature. The aggregate block — the New Keynesian Phillips curve, the Taylor rule, and the government budget constraint — is solved jointly with the household block using the sequence-space Jacobian method of Auclert, Bardóczy, Rognlie, and Straub (2021), which is the standard approach for computing transitional dynamics and aggregate

impulse responses in HANK-class models without resorting to the curse of dimensionality associated with full nonlinear general-equilibrium solution methods. Counterfactual economies — with BNPL fully available, with BNPL unavailable ($\bar{L} = 0$ for all z), and with intermediate values of the pass-through elasticity θ — are each solved to a separate stationary equilibrium, and the quintile-specific MPC is computed as the average first-quarter consumption response to a small, one-time, unanticipated transfer shock of one percent of quarterly income, following the standard MPC experiment design in Kaplan and Violante (2014).

It should be stated plainly that the quantitative results reported in Section 6 below are generated from this calibration strategy as a proof-of-concept exercise demonstrating the mechanism the model is designed to capture; they are not the output of an estimation run against proprietary loan-level transaction data, which the authors do not currently have access to. A natural extension of this research, noted in Section 10, is to re-estimate the quintile-specific parameters in Table 1 against such data once available, which would allow the illustrative results below to be replaced with point estimates suitable for direct policy use.

6. Illustrative Simulation Results

6.1 MPC Across Quintiles With and Without BNPL Access

Table 2. Quarterly MPC Out of a Transitory Income Shock, Illustrative Simulation.

Quintile	MPC, no BNPL	MPC, with BNPL	Change (pp)
Q1	0.61	0.66	+5.0
Q2	0.52	0.63	+11.0
Q3	0.41	0.50	+9.0
Q4	0.28	0.31	+3.0
Q5	0.09	0.10	+1.0

The pattern in Table 2 is consistent with the qualitative thrust of Di Maggio, Katz, and Williams (2022): BNPL access raises the MPC most sharply in the second and third quintiles rather than the very lowest quintile, because the lowest quintile in the calibration is disproportionately composed of households facing binding adoption frictions — thin credit files, limited e-commerce access, and lower approval rates — that limit the effective liquidity relaxation BNPL provides, even though, conditional on adoption, those households would value the additional liquidity most highly. This is consistent with CFPB (2023) evidence that BNPL usage is disproportionately concentrated in the $20,001–$50,000 income band rather

than in the very lowest income band, a pattern the model reproduces through the adoption-rate calibration in Table 1 rather than through any feature imposed directly on preferences.

6.2 Impulse Response to a Transitory Income Shock

Simulated impulse responses (not reproduced in full here for brevity) show that households with active BNPL access exhibit a faster and larger initial consumption response to a one-time negative income shock than their counterfactual counterparts without BNPL, consistent with a genuine consumption-smoothing channel: the BNPL credit line functions as a buffer that allows the household to maintain consumption in the shock period by drawing down $\bar{L}(z)$ rather than cutting consumption one-for-one with income. However, the model also generates a subsequent negative consumption response in the two to three quarters following the shock, as the household repays the BNPL installment balance drawn down during the shock period — an intertemporal substitution effect that is, by construction, smaller in present-value terms than the initial smoothing benefit only when the household does not also incur a late fee or spillover cost during the repayment window.

6.3 Decomposing the Channels: Smoothing Versus Extraction

Table 3 below decomposes the consumption-equivalent welfare effect of BNPL access into its constituent channels, for the central calibration ($\theta = 0.5$, $m = 0.05$).

Table 3. Decomposition of Welfare Effect by Channel, Illustrative Simulation (Consumption-Equivalent, % of Quarterly Income.)

Quintile	Insurance / smoothing gain	Late-fee extraction	Spill over-cost extraction	Price-markup extraction (cost-shifting)	Net welfare effect
Q1	+1.8	−0.6	−0.5	−0.4	+0.3
Q2	+3.1	−0.7	−0.6	−0.5	+1.3
Q3	+2.4	−0.4	−0.3	−0.5	+1.2
Q4	+0.9	−0.1	−0.1	−0.5	+0.2
Q5	+0.2	0.0	0.0	−0.3	−0.1

Three features of this decomposition are worth flagging. First, the net welfare effect is positive for the four lowest quintiles in the central calibration, but the composition of that net effect changes sharply across the distribution: for Q1 and Q2, late-fee and spillover-cost extraction claws back roughly forty to sixty percent of the gross smoothing gain, whereas for Q4 the smoothing gain is small to begin with and the net effect is driven almost entirely by

avoided extraction relative to Q1–Q3. Second, the price-markup extraction channel — merchant cost-shifting — is borne by every quintile roughly in proportion to its consumption share rather than its BNPL adoption rate, since it operates through the aggregate retail price level; this is the channel through which Q5, a quintile with negligible BNPL adoption, can experience a negative net welfare effect, paying a higher retail price level for a product its members rarely use. Third, and most importantly for the paper's core question, the sign and magnitude of the net welfare effect for the lowest quintiles is highly sensitive to the late-fee incidence parameter π and the pass-through elasticity θ, both of which are the least precisely calibrated parameters in the model; Section 8 returns to this sensitivity directly.

7. Welfare Analysis

Aggregating the quintile-specific decomposition in Table 3 by population share yields an aggregate consumption-equivalent welfare gain from BNPL access of approximately 0.6 percent of quarterly income in the central calibration — a positive but modest figure that masks substantial heterogeneity. Re-running the welfare calculation under the polar case $\theta = 1$ (full merchant-fee pass-through) reduces this aggregate figure to approximately 0.1 percent and turns the net effect negative for both Q1 and Q5, while the polar case $\theta = 0$ (merchants fully absorb the fee) raises the aggregate figure to approximately 1.0 percent and renders the net effect positive for every quintile. This comparative-statics exercise formalizes the paper's central claim: the welfare verdict on BNPL is not a fixed property of the product, but depends on an empirically contestable parameter — the degree to which a merchant discount fee charged on BNPL transactions is recovered through higher prices paid by the broader consumer population — that the existing published literature has not yet pinned down with the same precision applied to late-fee incidence or spending-response magnitudes.

A second welfare-relevant margin is the late-fee and spillover-cost channel's persistence. Because the model treats a late-payment event as generating both an immediate fee and a probability ψ of spillover into other credit products, the welfare cost of BNPL access compounds for households whose installment balance ℓ_it remains elevated across multiple periods — the repayment-stacking mechanism documented qualitatively by CFPB (2023) and Larrimore et al. (2024). The model's quintile-specific spillover probabilities in Table 1 are calibrated to be highest in Q1, reflecting the CFPB finding that BNPL borrowers exhibit elevated rates of overdraft and other high-cost credit use; this calibration choice is the primary driver of the relatively muted net welfare gain for Q1 in Table 3 despite that quintile facing, in principle, the largest unmet liquidity need.

8. Discussion: Reconciling Smoothing and Extraction

The model developed here suggests that the "consumption smoothing or debt trap" framing in the paper's title is best understood not as a question with a single answer, but as a question whose answer is a function of three parameters that vary by provider, by regulatory regime, and over time: the late-fee incidence and severity (φ, π), the degree of repayment-stacking spillover into other credit products (ψ), and the elasticity of merchant-fee pass-through into retail prices (θ). The empirical trajectory of the first two parameters, as documented by CFPB (2025), appears to be moving in a direction favorable to the smoothing interpretation — late-fee incidence and revenue have fallen even as adoption has grown, consistent with providers tightening underwriting and reducing reliance on fee revenue as a profit center. The third parameter, by contrast, is the least studied empirically and, within this model, the most consequential for whether BNPL functions as a progressive liquidity technology or a regressive, economy-wide cost imposed on non-users to subsidize the consumption of adopters.

This reframing has a direct implication for how the debate is typically conducted in popular and regulatory commentary, which tends to focus heavily on the late-fee and credit-spillover channels — precisely because they are the channels for which transaction-level data exist — while largely ignoring the merchant cost-shifting channel, for which the relevant elasticity has, to the authors' knowledge, not yet been directly estimated in the BNPL context specifically, in contrast to the well-developed empirical literature on interchange-fee pass-through in the credit-card and debit-card literature. The model's qualitative conclusion that this channel can flip the sign of the welfare effect for both the lowest and highest income quintiles — the former because they pay higher prices without benefiting from the smoothing gain enjoyed by Q2–Q3, the latter because they bear the price effect with essentially no offsetting access — suggests that this is a first-order gap in the empirical literature rather than a second-order modeling refinement.

9. Policy Implications

Three policy-relevant conclusions follow from the analysis, stated with appropriate caution given the illustrative nature of the underlying calibration.

The framework suggests that regulatory attention focused narrowly on capping or eliminating BNPL late fees, while addressing a real channel of extraction documented by CFPB (2023) and deHaan et al. (2024), would leave the merchant cost-shifting channel entirely unaddressed; a regulatory regime concerned with distributional fairness would need to

consider disclosure or pricing rules governing the merchant discount fee itself, analogous to interchange-fee regulation in the payment-card space, rather than fee rules directed only at the consumer-facing side of the contract.

The model also suggests that policies aimed at improving underwriting and reducing repayment-stacking — such as the data-sharing and credit-bureau-reporting reforms the CFPB has pursued for BNPL products — operate primarily through the spillover-cost channel (ψ) in this framework, and the comparative statics in Section 7 imply that reductions in ψ are a relatively efficient way to raise the net welfare effect for the lowest two quintiles, since that channel is calibrated to be both large and concentrated precisely where the smoothing benefit is otherwise weakest.

Finally, the finding that the second and third income quintiles, rather than the very lowest, capture the largest MPC increase and the largest net welfare gain in the central calibration implies that BNPL's aggregate-demand effects — its relevance to fiscal and monetary transmission, in the spirit of the original HANK literature — are likely to be most pronounced for the segment of the income distribution often described as "middle-class but liquidity-constrained," a segment whose MPC behavior is otherwise difficult for policymakers to observe directly. This suggests BNPL adoption data could, in principle, serve as a useful high-frequency proxy for liquidity-constrained consumption behavior in this part of the income distribution, a possibility we flag for future empirical work rather than develop further here.

10. Limitations and Future Research

The central limitation of this paper is one we have tried to state transparently throughout: the quantitative results in Sections 6 and 7 are generated from an illustrative calibration constructed from published, aggregate-level facts about the BNPL market, not from an estimation exercise run against proprietary loan-level or transaction-level fintech data broken out by income quintile. The qualitative mechanisms the model identifies — a smoothing channel concentrated in the middle of the income distribution, an extraction channel concentrated at the bottom, and a price-level cost-shifting channel that can extend the cost of BNPL to non-users — follow directly from well-documented empirical facts in the literature reviewed in Section 2, but the precise magnitudes in Tables 2 and 3 should be read as a demonstration of what the model is capable of producing once properly estimated, not as point estimates suitable for direct policy calibration.

A natural next step, contingent on data access, is to re-estimate the quintile-specific adoption, borrowing-limit, late-fee, and spillover parameters directly against transaction-level data of the kind used by Di Maggio, Katz, and Williams (2022) and deHaan, Kim, Lourie, and Zhu (2024), ideally merged with income data sufficient to construct genuine income-quintile breakdowns rather than the stylized approximation used here. A second extension would estimate the merchant-fee pass-through elasticity θ directly, following the empirical strategies developed in the interchange-fee pass-through literature, applied specifically to retailers' pricing behavior before and after BNPL adoption. A third extension, left for future work, would relax the assumption of a single representative BNPL contract and instead model competition among multiple BNPL providers with differing fee and underwriting structures, allowing the model to speak to the consumer-protection concern — raised in the survey evidence on repayment stacking — that households increasingly draw on BNPL credit from several providers simultaneously in ways invisible to any single lender's underwriting model.

11. CONCLUSION

This paper develops a heterogeneous-agent New Keynesian framework with an explicit Buy-Now-Pay-Later contract to assess whether the product is best understood as a consumption-smoothing technology or an extractive debt trap for lower-income households. The answer the framework yields is neither, but rather a redistributive credit instrument whose net welfare effect depends on three parameters with very different empirical footing: a late-fee and credit-spillover channel that the existing literature has measured with reasonable precision and that appears, on recent data, to be moderating; and a merchant cost-shifting channel that the existing literature has not yet measured directly in the BNPL context, but which this paper's general-equilibrium structure shows is capable of determining the sign of the welfare effect for both the most and least liquidity-constrained households in the economy. Resolving the debate in the paper's title, in other words, depends less on further documenting the consumer-facing fee structure of BNPL — already a relatively well-studied object — than on directly estimating how much of the merchant discount fee financing the product is ultimately paid, through higher retail prices, by households who never see a BNPL installment plan at checkout at all.

REFERENCES

1. Aiyagari, S. Rao. 1994. “Uninsured Idiosyncratic Risk and Aggregate Saving.” Quarterly Journal of Economics 109 (3): 659–684.
2. Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. “Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models.” Econometrica 89 (5): 2375–2408.
3. Berg, Tobias, Valentin Burg, Jan Keil, and Manju Puri. 2024. “The Economics of ‘Buy Now, Pay Later’: A Merchant’s Perspective.” NBER Working Paper No. 33152.
4. Consumer Financial Protection Bureau (CFPB). 2023. “Consumer Use of Buy Now, Pay Later: Insights from the CFPB Making Ends Meet Survey.” CFPB Office of Research.
5. Consumer Financial Protection Bureau (CFPB). 2025. “The Buy Now, Pay Later Market.” CFPB Data Spotlight.
6. deHaan, Ed, Jungbae Kim, Ben Lourie, and Chenqi Zhu. 2024. “Buy Now Pay (Pain?) Later.” Management Science 70 (8): 5586–5598.
7. Di Maggio, Marco, Justin Katz, and Emily Williams. 2022. “Buy Now, Pay Later Credit: User Characteristics and Effects on Spending Patterns.” NBER Working Paper No. 30508.
8. Huggett, Mark. 1993. “The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies.” Journal of Economic Dynamics and Control 17 (5–6): 953–969.
9. Kaplan, Greg, and Giovanni L. Violante. 2014. “A Model of the Consumption Response to Fiscal Stimulus Payments.” Econometrica 82 (4): 1199–1239.
10. Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. “Monetary Policy According to HANK.” American Economic Review 108 (3): 697–743.
11. Larrimore, Jeff, Alicia Lloro, Zofsha Merchant, and Anna Tranfaglia. 2024. “‘The Only Way I Could Afford It’: Who Uses BNPL and Why.” Board of Governors of the Federal Reserve System, FEDS Notes, December 20.