Unobservable Heterogeneity in the Marginal Propensity to Consume: Evidence from Korean Household Panel Data

Hyun Hak Kim*

August 25, 2026

Abstract

We estimate the distribution of the marginal propensity to consume (MPC) out of public transfers across Korean households, using fourteen linked waves of the Survey of Household Finances and Living Conditions (2012–2025; 132,965 household-year first differences). A finite mixture model finds that about three-quarters of households have a transfer MPC near zero while roughly a quarter respond with an MPC of 0.16, so the aggregate response is produced by a minority. That minority cannot be identified in advance: income, wealth, liquidity, debt and demographics jointly explain six to ten percent of the variation in responses, indistinguishable from the eight percent reported for the United States. Consequently a doubly robust targeting rule evaluated out-of-sample delivers less stimulus than a universal transfer. Six standard estimators applied to the same data yield average MPCs from 0.032 to 0.107, and unit conversions remove much of the apparent disagreement among published Korean estimates.

Keywords: marginal propensity to consume; latent heterogeneity; generalized random forests; finite mixture models; cash transfers

JEL Classification: E21; D12; H31

*Professor, Department of Economics, Kookmin University, South Korea. E-mail: hyunhak.kim@kookmin.ac.kr. I gratefully acknowledge financial support from the Bank of Korea for an earlier and substantially different report on this topic. The views expressed here are my own and do not represent those of the Bank of Korea, Statistics Korea, or the Financial Supervisory Service. All errors are my own.

1 Introduction

When a government transfers cash to households, how much of it gets spent depends on who receives it. This is now the organising insight of a large literature on fiscal policy with heterogeneous agents (Kaplan and Violante, 2022; Auclert et al., 2024), and it has an obvious corollary for programme design: if some households spend much more of a transfer than others, a fixed budget does more for aggregate demand when it is directed at them. Korea has run the experiment twice in recent memory, paying a universal transfer to every household in May 2020 and a means-tested one to the bottom 88 percent of the distribution in September 2021, and the choice between those two designs remains contested.

The empirical literature that ought to inform this choice reports averages. Estimates of the Korean marginal propensity to consume span roughly 0.07 to 0.61, and where heterogeneity is examined it is usually through comparisons of group means—the response among low-income households against high-income households, or among the liquidity-constrained against the unconstrained. Such comparisons establish that responses differ. They do not establish what a targeting rule requires, which is that the households with high responses can be picked out in advance from information a programme administrator actually has.

This paper separates those two questions. Using every wave of the Survey of Household Finances and Living Conditions from 2012 to 2025—a fourteen-year linked panel of 132,965 household-year first differences on 45,241 households—we estimate the distribution of the marginal propensity to consume out of public transfers and then ask how much of that distribution observable characteristics can account for. We do not claim quasi-experimental identification. We attempted two designs based on the 2020 and 2021 programmes and report in Section 3 why both fail, with the reasons quantified rather than asserted. In place of an identification strategy we offer transparency about what the estimates depend on.

Four findings emerge. First, heterogeneity is real and large. A calibration test rejects a constant response at $p = 3 \times 10^{-10}$, and a finite mixture locates the structure: about three-quarters of households have a transfer MPC close to zero, while roughly a quarter respond with an MPC of 0.16. The aggregate response is produced by a minority.

Second, that minority is not identifiable from observables. Income, wealth, liquidity, debt and demographics jointly account for between six and ten percent of the variation in household-level responses, and most of what they do account for is the familiar gradient in resources rather than any signal of binding constraints. This is statistically indistinguishable from the eight percent that Lewis et al. (2024) report

for the United States.

Third, targeting therefore underperforms. An allocation rule learned by doubly robust policy learning (Athey and Wager, 2021) and evaluated on held-out households delivers less aggregate stimulus than a universal transfer, even though it correctly selects poorer households and even though the low-income response really is higher. The rule raises the response per won by 20 to 60 percent but reduces the total, because the households it excludes are not non-spenders. The relevant trade-off is between efficiency and scale.

Fourth, the average marginal propensity to consume is fragile in a way that helps explain the dispersion in published estimates. Applying six standard estimators to one dataset, one estimand and one covariate set yields values from 0.032 to 0.107, and the coefficient reverses sign when a single control—the change in other income—is omitted. Separately, several Korean studies report elasticities rather than level propensities; converting them removes much of the apparent disagreement, and places Song (2020)'s transitory-income estimate in a range that contains ours.

Our contribution is thus partly substantive and partly methodological. On the substantive side we provide the first estimate of the distribution of the marginal propensity to consume for Korea, built on the longest panel yet assembled from this survey—fourteen linked waves, against the six underlying Song (2020)—and evidence that the limited predictability of that distribution is not specific to the United States. On the methodological side we document how much of the apparent disagreement in this literature is attributable to units, specification and estimator choice, and we report two specification errors—conditioning on post-treatment covariates, and random rather than household-clustered cross-fitting—whose consequences are large; the first came to light only because two estimators disagreed.

These questions sit at the junction of two literatures. Work on MPC heterogeneity offers two accounts of why households differ. One points to circumstances: households with little cash on hand cannot smooth a temporary shock, whether the binding force is a borrowing constraint or a precautionary motive (Deaton, 1991; Carroll, 1997; Zeldes, 1989), and households can be wealthy in net terms yet behave as if hand-to-mouth (Kaplan and Violante, 2014). The evidence here is strong, from Italian survey data (Jappelli and Pistaferri, 2014) to Norwegian lottery windfalls (Fagereng et al., 2021) and experimentally assigned credit lines (Aydin, 2022). The other account points to persistent characteristics such as impatience or rule-of-thumb behaviour (Campbell and Mankiw, 1989; Krusell and Smith, 1998; Gelman, 2021). We do not adjudicate between the two; what our results constrain is how much either account can explain with the variables usually available.

A smaller literature asks what the distribution of the marginal propensity to con-

sume looks like, rather than its conditional means (Carroll et al., 2017; Fishera et al., 2020). The closest paper to ours is Lewis et al. (2024), who estimate the unconditional distribution from the 2008 US rebates and find both that it is wide and that observables explain only about eight percent of it. Whether that finding travels beyond the United States, and beyond a one-time rebate, has not to our knowledge been tested; Section 5 provides such a test. A parallel debate questions micro MPC estimates themselves, whether because their macroeconomic counterfactuals are implausible (Orchard et al., 2025) or because publication selection inflates them (Sokolova, 2023); our estimator-sensitivity results speak to that debate from inside a single dataset, before selection or aggregation can operate.

The paper proceeds as follows. Section 2 describes the panel. Section 3 establishes what these data can say about the average response: two quasi-experimental designs fail for quantifiable reasons, the surviving estimate is fragile in instructive ways, and the published Korean estimates, restated in common units, largely agree with ours. Section 4 estimates the conditional response function, Section 5 turns to the distribution, Section 6 draws out the implications for targeting, and Section 7 concludes.

2 Data

2.1 The Survey

We use the Survey of Household Finances and Living Conditions, compiled jointly by Statistics Korea, the Financial Supervisory Service and the Bank of Korea (Statistics Korea et al., 2025). The survey covers roughly 18,000 to 20,000 households each year and is the country's principal source of linked information on household income, consumption, assets and liabilities. We use every wave available in the public release, 2012 through 2025.

The timing convention matters throughout the paper. Balance-sheet items are measured as of 31 March of the survey year, while income and expenditure refer to the preceding calendar year. The 2021 wave therefore records consumption and income for 2020, which is when the first pandemic transfer was paid; the 2022 wave records 2021, when the second was paid. We index observations by survey year and state the corresponding income year wherever confusion is possible.

2.2 Panel Construction

The survey follows households across waves through an identifier that is stable in the public release, though it has never to our knowledge been used to build a panel of this

length. We link waves on that identifier after verifying that it is unique within each year and that no household appears twice.

Table 1 reports the resulting structure. Match rates with the preceding wave are 92 and 94 percent in the first two links and settle at roughly 76 percent thereafter, consistent with the survey's rotating design. Pooling all waves gives 257,325 household-year observations on 68,595 distinct households. Because our specifications are in first differences, the operative sample is the set of observations with a lagged observation on the same household, which numbers 188,576 before further restrictions.

Table 1. Panel structure by survey wave

Survey year	Households surveyed	With consumption	With a lag	Match with previous year	Receiving transfers
2012	19,744	19,744	0	—	0.373
2013	18,594	9,268	18,076	0.916	0.394
2014	17,863	8,907	17,438	0.938	0.437
2015	18,031	8,974	13,695	0.767	0.447
2016	18,273	9,097	13,850	0.768	0.469
2017	18,497	9,185	14,095	0.771	0.554
2018	18,640	9,261	14,113	0.763	0.572
2019	18,406	18,406	14,292	0.767	0.610
2020	18,064	18,064	13,910	0.756	0.676
2021	18,187	18,187	13,793	0.764	0.999
2022	17,954	17,954	13,760	0.757	0.986
2023	18,094	18,094	13,648	0.760	0.836
2024	18,314	18,314	13,828	0.764	0.714
2025	18,664	18,664	14,078	0.769	0.716

Notes: "With consumption" counts households for which consumption expenditure is recorded; see Section 2.3. "With a lag" counts households observed in the preceding wave. "Match with previous year" is the share of the preceding wave's households that reappear. "Receiving transfers" is the share with positive public transfer income.

The final column of Table 1 documents the two transfer programmes discussed in Section 3. The share of households with positive public transfer income rises steadily from 0.37 in 2012 to 0.68 in 2020, jumps to 0.999 in 2021 and 0.986 in 2022, and falls back to 0.72 by 2025. Figure 1 plots the series. The two spikes correspond exactly to the universal payment of May 2020 and the means-tested payment of September 2021.

2.3 A Break in the Consumption Sample

One feature of the survey design requires attention because it is not visible from the published aggregates. Between 2013 and 2018 the survey was fielded in two sections, a welfare section and a finance section covering disjoint sets of households, and consumption expenditure was collected only in the former. Assets and liabilities were

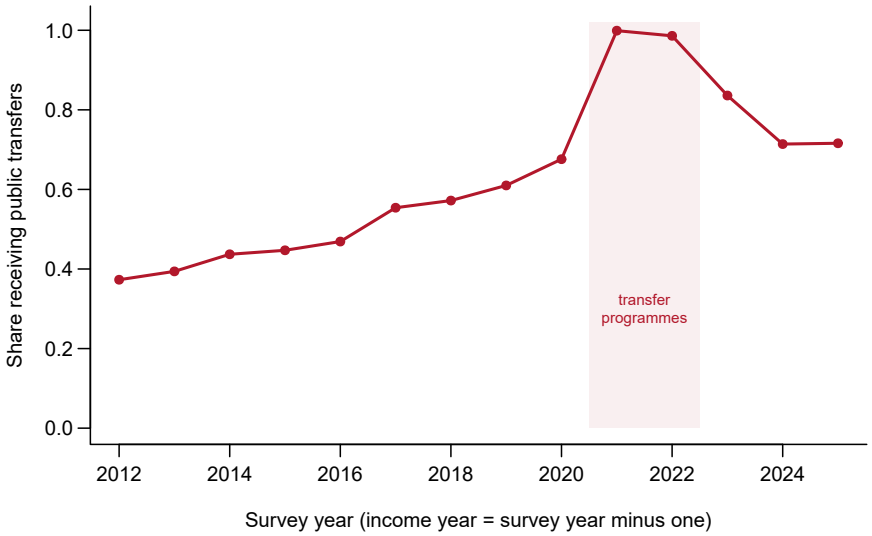


Figure 1. Share of households receiving public transfers

Notes: Share of surveyed households reporting positive public transfer income. The shaded region marks the waves recording the 2020 and 2021 transfer programmes.

collected in both. As a result, roughly half the households surveyed in those years have no consumption data at all.

The consequence is a sample that steps twice: 19,744 households with consumption in 2012, about 9,000 per year from 2013 to 2018, and about 18,000 per year from 2019 onward, when the sections were merged. This is not attrition and it is not selection on consumption; it is a design feature, and the survey documentation confirms it in the variable label for total expenditure. But it has to be handled. Two points limit the damage. First, because consumption is the dependent variable, the estimation sample in 2013–2018 consists of welfare-section households by construction, so there is no selection between sections to correct for; Table A1 in Appendix A2 documents this. Second, all specifications include year effects, which absorb any level shift, and re-estimating on the 2019–25 waves alone—where the sections are merged and the full sample carries consumption data—leaves the conclusions intact. We flag the break here because an analysis that pooled the waves without noticing it would attribute a compositional change to a change in behaviour.

2.4 Variables

The dependent variable is the change in annual consumption expenditure; the survey also records six expenditure categories separately, and Table A2 in Appendix A2 decomposes the transfer response across them. Income is decomposed into labour, business, capital, public transfer and private transfer components; the treatment variable is the change in public transfer income, and the change in the sum of the remaining

components enters as a separate regressor. Section 3 shows that this separation is the single most consequential specification choice in the paper.

Balance-sheet variables comprise liquid assets, which we measure by the stock of savings deposits; net wealth; total debt; and the annual principal and interest payment, from which we construct a debt-service ratio. Demographic variables cover household size, and the age, sex, education, occupation and marital status of the head, together with housing tenure and indicators for elderly, single-parent and metropolitan households.

Every covariate is measured at $t-1$. This is not a stylistic preference. A covariate measured contemporaneously with the transfer is potentially an outcome of it—a household that receives a transfer and uses it to service debt thereby changes its debt-service ratio—and Section 3.7 documents how severely conditioning on such a variable distorts the estimates. We therefore constructed lagged versions of every balance-sheet and classification variable used in the analysis.

We trim the upper and lower half-percent of the distributions of the consumption change and the disposable income change, and winsorise ratio variables at the first and ninety-ninth percentiles, since ratios with small denominators produce extreme values that a linear projection cannot accommodate. Amounts are reported in 10,000 KRW in the estimation and converted to millions in the tables. After these restrictions the estimation sample contains 132,965 household-year observations on 45,241 households.

Two defects in the source materials required correction, and because both bear on reproducibility we document them in Appendix A1. The published variable-name mapping transposes public and private transfer income in every year from 2012 to 2018, an error that is detectable by comparison with the official published aggregates, in which public transfers are several times larger than private. And the 2019 raw file duplicates the label for the principal-repayment variable while omitting the label for interest paid; the second of the duplicated columns is in fact interest, as an accounting identity confirms.

2.5 Descriptive Statistics

Table 2 summarises the estimation sample. The average household has disposable income of 44.2 million KRW and consumption expenditure of 24.9 million, an average propensity to consume of roughly 0.56. Public transfers average 5.6 million KRW a year, with a median of 2.6 million, so the payments examined here—up to one million KRW per household—amount to about a fifth of mean annual transfer receipts, two-fifths of the median, and around four percent of annual consumption.

Panel D reports the first differences that the analysis uses. The standard deviation

Table 2. Descriptive statistics, estimation sample

	Mean	S.D.	P25	Median	P75
Panel A: Income and consumption (million KRW per year)					
Current income	53.7	49.0	20.4	41.5	72.2
Disposable income	44.2	37.7	17.6	35.2	60.3
Public transfers	5.6	8.6	0.0	2.6	7.6
Private transfers	1.2	3.2	0.0	0.0	0.8
Consumption expenditure	24.9	17.0	12.3	21.1	33.4
Panel B: Balance sheet (million KRW)					
Liquid assets (savings)	69.2	157.1	6.3	29.6	77.8
Net wealth	330.3	559.4	54.4	175.2	398.1
Total debt	62.7	160.9	0.0	5.0	61.5
Debt service	8.3	28.2	0.0	0.4	7.2
Panel C: Household characteristics					
Household size	2.5	1.3	1.0	2.0	4.0
Age of head	57.6	15.1	46.0	57.0	69.0
Panel D: First differences (million KRW per year)					
Consumption, Δc	0.7	7.9	-2.7	0.5	4.1
Disposable income, Δy	2.1	15.8	-3.7	1.2	8.0
Public transfers, $\Delta \tau$	0.6	3.8	0.0	0.0	0.9
Other income	1.5	15.9	-4.0	0.7	7.4

Notes: 132,965 household-year first differences on 45,241 households, 2013–2025. Balance-sheet variables are measured at $t-1$. Amounts are in millions of KRW; one million KRW was approximately USD 750 in 2024.

of the annual consumption change is 7.9 million KRW, an order of magnitude larger than the transfers under study. This ratio is what makes the quasi-experimental designs of Section 3 underpowered, and it is worth keeping in view when interpreting the precision of any estimate based on annual survey consumption.

Finally, a trend worth noting because it shapes the balance-sheet variables over our sample. The share of households holding any debt falls steadily from 0.644 in 2012 to 0.536 in 2025, and median debt from 6.0 to 3.0 million KRW. Among indebted households, however, the share making principal or interest payments is stable between 0.81 and 0.87 throughout. A change in how debt service was measured would have moved that ratio; its stability points instead to genuine deleveraging, with fewer households borrowing rather than borrowers reporting differently.

3 The Average MPC: Identification, Fragility, and the Published Record

This section establishes what these data can and cannot say about the average marginal propensity to consume, in three steps. Korea's two pandemic transfers invite a quasi-experimental design, and we begin with two attempts—an instrument built from the household-size schedule of the first programme, and a difference-in-differences design built from the reversal between the two schedules—both of which fail, for reasons we quantify rather than assert. We then turn to what remains: a first-differenced specification whose coefficient proves highly sensitive to a single control and to the choice of estimator. The section closes by using both results to reconcile the published Korean estimates, whose dispersion narrows considerably once units and method are held fixed. The constructive conclusion, developed in Sections 4 and 5, is that the object these data can speak to is the distribution of responses rather than a single number summarising their mean.

3.1 The Two Programmes and a Household-Size Instrument

In May 2020 the government made an unconditional payment to every household, with the amount rising in household size but capped at four members: 400,000 KRW for a single-person household, 600,000 for two, 800,000 for three, and one million for four or more. In September 2021 a second programme paid 250,000 KRW per person to households in the bottom 88 percent of the health-insurance premium distribution, with no cap. Under the survey's timing convention these appear in the 2021 and 2022 waves respectively.

The schedules are visible in the data. The share of households reporting positive

public transfer income rises from 0.68 in the 2020 wave to 0.999 in 2021 and 0.986 in 2022, before falling back to 0.84 and then roughly 0.72. Whatever else is true, the programmes reached almost everyone and the survey recorded them.

The 2020 schedule assigns different amounts to households of different sizes, which suggests using the scheduled amount as an instrument for the observed change in transfer income. On the 13,565 households observed in both the 2020 and 2021 waves, the first stage is strong: regressing the transfer change on the scheduled amount gives a coefficient of 2.67 with an F statistic of 287, and a binary version comparing households of four or more with smaller ones puts the difference in transfer changes at 747,000 KRW, with an F of 79.5. The magnitude of the first coefficient is itself a warning. If the schedule simply metered the relief payment, the coefficient would be close to one; at 2.67, each scheduled won is associated with 2.67 won of additional transfer income, so household size is proxying for a bundle of pandemic-era transfers rather than for the relief payment alone.

The second stage confirms the warning. A simple two-stage least squares estimate returns an MPC of 1.53—consumption rising by more than the entire transfer. An instrumental forest applied to the same design returns a local average treatment effect of -0.72 with a standard error of 3.53, while its own conditional predictions lie between 1.7 and 7.8; that the score-based average and the conditional predictions cannot even agree on sign is a symptom of the breakdown, not a finding about behaviour. The software issues two warnings of its own: the estimated instrument propensities range from 0.003 to 0.856, indicating poor overlap, and some compliance scores are as low as -152.

The diagnosis is straightforward: the exclusion restriction fails. Household size affects consumption growth directly and for reasons unrelated to the transfer—families at different stages of the life cycle have different consumption trajectories—so the reduced form picks up much more than the transfer channel. The 2020 policy environment makes this worse rather than better, because a separate childcare voucher programme operating that year was also keyed to household composition. An instrument correlated with a second, contemporaneous treatment cannot isolate the first.

3.2 A Slope-Reversal Design

A second design is more promising in principle, because it does not require household size to be excluded from the consumption equation. The two programmes assign amounts to household sizes in different ways, and crucially the ordering reverses (Table 3):

Because household-size fixed effects absorb any direct effect of size on consumption

Table 3. Transfer schedules by household size (10,000 KRW)

Household size	1	2	3	4	5	6
2020 relief payment (universal)	40	60	80	100	**100**	**100**
2021 support payment (bottom 88%)	25	50	75	100	**125**	**150**
Difference	−15	−10	−5	0	+25	+50

Notes: The 2020 schedule is concave, capped at four members; the 2021 schedule is linear at 250,000 KRW per person. Households of five or more received relatively less under the first programme and relatively more under the second.

growth, identification comes from the fact that the same household size receives a different *relative* amount under the two programmes. If the direct effect of household size is stable over time, the reversal is attributable to the transfer.

The first stage again works: the scheduled amount predicts the observed transfer change with a coefficient of 1.16 and an F statistic of 146. The second stage again does not. The reduced form is −0.259 with a standard error of 0.200, so the implied Wald estimate of the MPC is −0.223 with a standard error of 0.174. Table 4 shows that the estimate is not merely imprecise but unstable across reasonable variations, running from −0.58 to +2.30.

Table 4. Difference-in-differences estimates of the transfer MPC

Specification	MPC	(s.e.)
Wald estimate, full sample	−0.223	(0.174)
Two-stage least squares, 2013–2022	0.120	(0.221)
Adding household size × programme period	−0.579	(0.427)
Restricting the 2021 schedule to eligible households	0.225	(0.231)
2020 programme only	−0.252	(0.138)
2021 programme only	2.301	(1.937)

Notes: Estimated on the full panel of 132,965 household-year first differences. All specifications include household-size and year fixed effects with standard errors clustered on households. Eligibility for the 2021 programme is proxied by excluding the top income decile, since the health-insurance premium used for means testing is not observed.

3.3 The Statistical Limits of the Design

The decisive problem is statistical power. The variation the design exploits is small relative to the noise it must overcome: the standard deviation of the annual consumption change is 7.88 million KRW, while the standard deviation of the scheduled transfer amount is 282,000 KRW. With a reduced-form standard error of 0.200 and a first-stage coefficient of 1.16, the minimum effect detectable with 80 percent power is 0.48 in MPC units, and every plausible value of the marginal propensity to consume—the

literature's range is roughly 0.05 to 0.30—lies below that threshold. This is arithmetic rather than misfortune. A larger sample would not help, because the constraint is the ratio of the transfer's size to the year-to-year variability of household consumption.

Even with unlimited power, the design would fail its own diagnostic. Figure 2 plots coefficients from interacting the household-size-based schedule with year indicators. Three of the seven pre-programme coefficients are individually significant—−2.13 in 2013, −1.90 in 2014 and −1.39 in 2016—and they are as large as the coefficient in the treatment year itself, −1.40. Consumption growth of different household-size groups was not on parallel paths before the transfers, so the identifying assumption fails on its own terms.

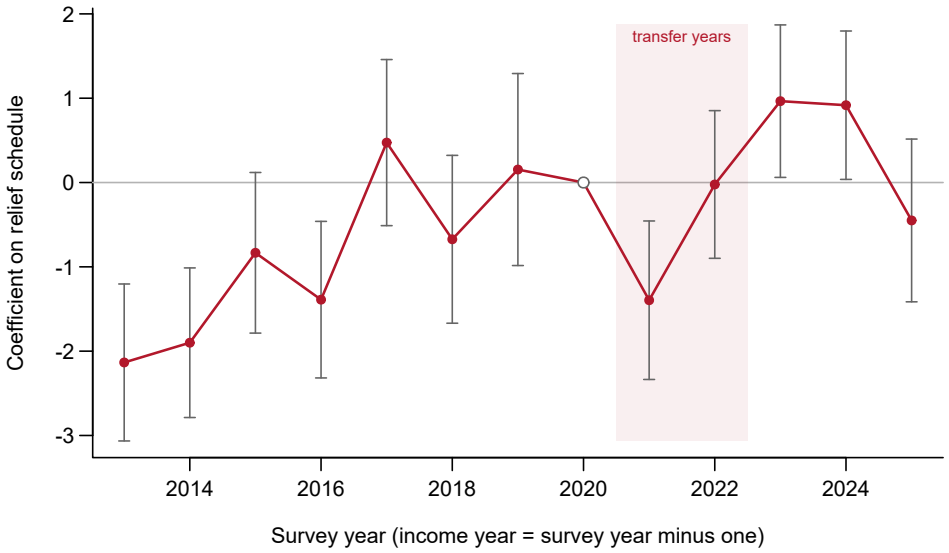


Figure 2. Event study: the schedule interacted with year

Notes: Coefficients from interacting the 2020 relief schedule, treated as a fixed household characteristic, with year indicators, relative to the 2020 wave. Bars are 95 percent confidence intervals with standard errors clustered on households. The shaded region marks the waves in which the transfer programmes appear.

The treatment year compounds the problem. Because the 2021 wave records consumption for calendar year 2020, the year of the transfer is also the year of the largest disruption to consumption patterns in recent memory, and that disruption was not neutral with respect to household size: school closures fell on households with children, and the composition of spending shifted differently for large and small households. The design asks whether households of different sizes changed their consumption differently in 2020, and the pandemic guarantees that they did, for reasons unconnected to the transfer.

There is, finally, no control group to fall back on. The 2020 payment was universal, so no set of untreated households exists against which to compare—only variation in

the amount, and that variation, as the power calculation shows, is too small.

None of this reflects defects of implementation that a better design would repair. It follows from the structure of the data: annual frequency, a transfer small relative to annual consumption, universal coverage, and coincidence with an aggregate shock correlated with the assignment variable. Card-transaction studies avoid all four problems, which is why the credible quasi-experimental evidence on Korean transfers comes from that source (Baek et al., 2023; Kim et al., 2023).

3.4 Specification and the Ladder

We therefore make no causal claim from a natural experiment in what follows. The estimand is the partial association between transfer receipt and consumption growth, holding other income changes and predetermined characteristics fixed, and the transparency device we offer in place of an identification strategy is the specification ladder below, which reports what the estimate depends on rather than asserting that it does not depend on anything. The exercise turns out to have a constructive moral: holding the dataset, the estimand and the covariate set fixed, the estimate still ranges from 0.032 to 0.107 across estimators, and a single omitted control moves it across zero—which is why the later sections shift attention from this single number to the distribution of responses.

Our baseline specification, used throughout the paper, is

$$\Delta c_{it} \;=\; \alpha_t \;+\; \beta\,\Delta\tau_{it} \;+\; \gamma\,\Delta y_{it} \;+\; X'_{i,t-1}\delta \;+\; \varepsilon_{it}, \tag{1}$$

where Δc_{it} is the change in annual consumption expenditure of household i between survey years $t-1$ and t, $\Delta\tau_{it}$ is the change in public transfer income, Δy_{it} is the change in all other disposable income, α_t are year effects, and $X_{i,t-1}$ collects the predetermined characteristics of Section 2. All amounts are in 10,000 KRW, roughly USD 7.5 at 2024 exchange rates. Working in differences removes any time-invariant household component of consumption, and the predetermined covariates $X_{i,t-1}$ absorb variation in the level of resources. Three features of the specification carry most of the weight, and we take them in turn: the treatment of other income, the choice of estimator, and the timing of the covariates.

Table 5 builds the specification up one component at a time. The first three rows are unremarkable and, taken alone, would support the conclusion that transfers are not spent: with no controls the coefficient is -0.013, and adding year and household-size fixed effects leaves it at -0.004. The fourth row changes the picture entirely. Once the change in *other* income is included, the coefficient jumps to $+0.048$ and becomes strongly significant. Adding household fixed effects on top of that leaves it at $+0.036$.

Table 5. Specification ladder for the transfer coefficient

	Specification	Coefficient on $\Delta\tau_{it}$	(s.e.)
(1)	No controls	−0.0126	(0.0069)
(2)	+ year effects	−0.0040	(0.0070)
(3)	+ household-size effects	−0.0037	(0.0070)
(4)	+ change in other income	0.0483	(0.0070)
(5)	(3) and (4) combined	0.0461	(0.0070)
(6)	+ household fixed effects	0.0358	(0.0081)

Notes: Ordinary least squares on 132,965 household-year first differences, with standard errors clustered on households. The dependent variable is the change in annual consumption expenditure. Rows build cumulatively except row (4), which adds the other-income change to row (2); row (5) combines the controls of rows (3) and (4). Row (6) drops 8,854 households observed only once, for which a household effect is not identified.

The reversal in row (4) is not a puzzle once the correlation structure is examined. Public transfers and other income move in opposite directions: $\text{corr}(\Delta\tau_{it}, \Delta y_{it}) = -0.164$. Transfers rise when other income falls, which is what a system of unemployment benefits, means-tested support and pensions is designed to do. Omitting Δy_{it} therefore loads the negative association between the two income sources onto the transfer coefficient and drives it to zero.

This is an omitted-variable problem with a clear diagnosis, not a matter of taste in specification, and it has a direct implication for reading the literature: wherever transfers are countercyclical at the household level—which is what transfer systems are designed to be—a study that regresses consumption on transfer income without controlling for contemporaneous movements in other income will understate the transfer MPC, potentially to zero.

3.5 Which Covariate Does the Work?

It is natural to ask whether the sensitivity in Table 5 is a general feature of the covariate set—many controls each contributing a little—or the work of one variable. Table 6 answers this by removing each block of covariates from the full specification and, separately, by adding each block on its own to a specification with year effects only.

The answer is unambiguous. Removing the change in other income moves the coefficient by −0.056, and adding it alone to a bare specification moves it by +0.053—recovering almost the whole gap between the two extremes of Table 5 by itself. No other block moves the estimate by more than 0.015, and the entire balance sheet—liquidity, net wealth, debt and debt service—moves it by less than 0.002 in total. At the level of individual variables the ranking is the same: other income first at −0.056,

Table 6. Contribution of each covariate block to the transfer coefficient

	Removed from full set		Added on its own	
Block	Coefficient	Change	Coefficient	Change
Change in other income	−0.0173	−0.0555	0.0465	+0.0528
Demographics	0.0528	+0.0146	−0.0051	+0.0012
Baseline income and consumption	0.0454	+0.0072	−0.0139	−0.0076
Net wealth	0.0400	+0.0018	−0.0066	−0.0003
Liquidity	0.0396	+0.0014	−0.0064	−0.0001
Debt and debt service	0.0381	−0.0001	−0.0064	−0.0001

Notes: The full specification includes all blocks and year effects and gives a coefficient of 0.0382; year effects alone give −0.0063. "Change" is measured against the full specification in the left pair of columns and against the year-effects-only specification in the right pair. Because least squares deletes observations with any missing covariate, specifications involving the balance-sheet blocks use the 123,609 complete-case observations.

household size a distant second at +0.022, and every balance-sheet variable below 0.002.

The fragility of the average MPC is therefore not diffuse. It is one omitted variable, with an economically transparent mechanism.

3.6 Estimator Choice

Fixing the specification does not fix the estimate. Table 7 applies six estimators to the same data, the same estimand and the same covariates. They differ only in how the nuisance functions are handled: least squares imposes linearity throughout; the double machine learning estimators of Chernozhukov et al. (2018a) partial out flexible predictions of the outcome and the treatment before estimating the coefficient, with cross-fitting; and the forests of Athey et al. (2019) additionally allow the coefficient itself to vary with covariates.

The estimates range from 0.032 to 0.107, a factor of 3.4. The ordering is systematic rather than random: least squares sits at the bottom and the more flexible estimators above it, which is what one expects if the true conditional expectation is non-linear in the covariates, so that a linear model leaves structure in the residuals that the transfer variable partly absorbs. But the spread among the flexible estimators is itself substantial, from 0.075 to 0.107, and no principle selects among them.

We do not attempt to adjudicate. The point is that a reader encountering any one of these numbers in isolation would form a materially different view of how much of a transfer is spent, and that the dispersion in the published Korean literature is not evidence of disagreement about the world so much as evidence that the estimand is fragile.

For completeness we also applied four meta-learners—the S-, T-, X- and doubly

Table 7. The same estimand under six estimators

Estimator	MPC out of transfers	(s.e.)
Ordinary least squares	0.032	(0.006)
Causal forest DML, gradient boosting	0.075	(0.042)
Linear DML, gradient boosting	0.075	(0.006)
Linear DML, lasso	0.085	(0.007)
Generalized random forest	0.102	(0.009)
Sparse linear DML, gradient boosting	0.107	(0.034)
Range (highest / lowest)	3.40	

Notes: All estimators use the same predetermined covariates and the same treatment definition. The double machine learning variants require complete covariate cases and use 123,477 observations; the generalized random forest handles missing covariates natively and uses all 132,965. Cross-fitting folds are formed by household, following Fuhr and Papies (2024); splitting at random would place different years of the same household in different folds and violate the independence that cross-fitting assumes. Gradient boosting uses GPU-accelerated `xgboost`. Standard errors are clustered on households where the estimator supports it.

robust learners—to a binarised treatment. All four return negative average effects, ranging from -5.8 to -0.02 in units of 10,000 KRW, with the doubly robust learner closest to zero. We regard these as uninformative rather than as evidence of negative effects. Dichotomising the transfer change discards the magnitude information that makes an MPC interpretable, and the resulting treatment definition—receipt above the sample median—is endogenous to the outcome distribution. We report them only to note that the choice between a continuous and a binary treatment is a further degree of freedom.

3.7 Two Traps

Two specification errors that we made and corrected in the course of this work deserve reporting, both because they are easy to make and because their consequences are large.

The first is the inclusion of post-treatment covariates. Our initial specification controlled for the contemporaneous income quintile, the contemporaneous debt-service ratio and contemporaneous interest paid. Each is an outcome of the transfer: a household that receives a transfer and uses it to service debt thereby changes its debt-service ratio. Including them allows a flexible learner to predict the treatment far too well—the out-of-fold R^2 for the treatment rose from 0.12 to 0.64—so that the orthogonalised residual retains little genuine variation in transfer receipt and the estimate is attenuated toward zero. The consequences were severe: the double machine learning estimate of the transfer MPC fell from 0.058 to 0.019, the estimate for disposable income from 0.080 to 0.038, and the 2021 relief-year estimate turned significantly negative at -0.047

before returning to an insignificant -0.004 once the covariates were lagged. All results in this paper use covariates measured at $t-1$.

The second concerns cross-fitting with panel data. Standard implementations assign observations to folds at random, which for a household panel places different years of the same household in different folds and violates the independence that the procedure assumes (Fuhr and Papies, 2024). We form folds by household throughout.

Neither error would have been visible from a single estimator. We found the first only because the double machine learning and forest estimates disagreed by a factor of three, which prompted the diagnostic that located the post-treatment controls. This is a practical argument for the estimator comparison in Table 7: its value is not only in quantifying fragility but in providing a check that a single specification cannot supply.

3.8 The Published Record

Korean estimates of the marginal propensity to consume span roughly 0.07 to 0.61, a factor of nearly nine. Table 8 collects the main studies, and read against the results above, much of that range dissolves into three differences of measurement rather than of behaviour.

The first difference is one of units. The two survey-based studies in Panel A report elasticities, not level marginal propensities: Song (2020) estimates his equations in log residuals, so his coefficients answer how consumption responds in percentage terms to a percentage change in income, and reading his 0.288 as a marginal propensity to consume—as a casual comparison across the table invites—overstates it by roughly a factor of two. Converting at the average propensity to consume brings his transitory-income estimate to between 0.085 and 0.107, a range that contains our own preferred estimate of 0.102. Two studies that appear to disagree by a factor of nearly three turn out to agree once expressed in common units.

The second difference is in what is measured. Card-transaction studies observe spending in the weeks after a payment; they capture the immediate response, but they miss cash and account-transfer spending and cannot net out substitution away from other consumption within the year. Annual survey studies capture the net change over a full year, including any offsetting reduction, and so mechanically produce smaller numbers. The gap between 0.26–0.59 in Panel B and 0.03–0.21 in Panel C is therefore at least in part a difference of horizon and coverage rather than of behaviour, and the two should not be pooled. The third difference is method, and it needs no further demonstration here: Table 7 shows that estimator choice alone spans a factor of 3.4 on fixed data.

Our relationship to Song (2020) deserves a final word, since we use the same survey.

Table 8. Estimates of the marginal propensity to consume for Korea

Study	Data	Method	Reported	Level MPC
Panel A: Survey-based, elasticities				
Song (2020)	SHFLC, restricted release, 2013–16	Variance decomposition; threshold model	0.288 overall 0.180 transitory	0.135–0.171 **0.085–0.107**
Hong (2021)	KLIPS, KLoSA	Variance decomposition	0.3 permanent 0.2 transitory	0.225 0.15
Panel B: Card transactions, level MPCs				
Baek et al. (2023)	Credit bureau card data	Difference-in-differences across provinces	0.36–0.58	0.36–0.58
Kim et al. (2023)	Seoul card data	Difference-in-differences	≥ 0.59	≥ 0.59
Kim and Oh (2020)	Card and sales data	Event study	0.26–0.36	0.26–0.36
Panel C: Survey-based, level MPCs				
Lee et al. (2022)	Household Income and Expenditure Survey, monthly	Triple differences	0.36–0.48	0.36–0.48
Noh (2021)	SHFLC 2019, cross-section	Regression	0.07–0.21	0.07–0.21
This paper	SHFLC, public release, 2013–25	Forests; finite mixture	0.032–0.107	0.032–0.107

Notes: Conversions in Panel A multiply the reported elasticity by an average propensity to consume. For Song (2020) we use 0.592, the aggregate ratio of consumption to disposable income in our data over his sample period, and 0.47 as a lower bound reflecting his use of non-durable consumption; the resulting range is reported. Hong (2021) performs the conversion himself using 0.75.

He works with the restricted release available to Bank of Korea researchers, which records household cash holdings directly and therefore supports a sharper measure of liquidity than the public release we use. His threshold approach asks where in the liquidity distribution the response changes; ours asks which characteristics matter when tested jointly, and what the distribution of responses looks like once one stops conditioning on observables altogether. We regard the two exercises as complementary, and Section 4 is explicit about where his measurement is better than ours.

4 Conditional MPC Heterogeneity

Section 3 showed that a single number summarising the average consumption response is fragile. We now ask a different question: holding the estimator fixed, how much does the response vary across households with different observable characteristics, and which characteristics matter? This section estimates the conditional MPC function non-parametrically and subjects it to two tests—whether the heterogeneity it finds is real, and which covariates account for it.

4.1 Estimating a Conditional MPC Function

Write the object of interest as

$$\beta(x) \;=\; \frac{\partial\, \mathbb{E}[\Delta c_{it} \mid \Delta\tau_{it}, \Delta y_{it}, X_{i,t-1} = x]}{\partial\, \Delta\tau_{it}}, \tag{2}$$

the marginal propensity to consume out of transfer income for a household with characteristics x. The predecessor to this paper, like much of the applied literature that uses machine learning for consumption, estimated a predictive model of the consumption *level* and read a marginal propensity off its partial derivative (Meinshausen, 2006). That approach has two drawbacks: the estimand is a by-product of a prediction problem rather than the target of estimation, and it comes with no distribution theory, so heterogeneity cannot be tested.

We instead use generalized random forests (Athey et al., 2019), which make $\beta(x)$ itself the estimand. The forest solves a local moment condition at each x, weighting observations by an adaptive kernel learned from the data rather than by a fixed distance metric, and the resulting estimates are consistent and asymptotically normal, so confidence intervals are available. Two variants are used. The `lm_forest` estimator fits (1) locally with both income terms as regressors, delivering a pair of conditional coefficients $(\beta(x), \gamma(x))$; the `causal_forest` estimator treats the transfer change as a continuous treatment and orthogonalises the outcome and the treatment against X

before estimating $\beta(x)$, in the manner of Nie and Wager (2021) and Chernozhukov et al. (2018a).

The covariate vector $X_{i,t-1}$ contains twenty predetermined variables: baseline income and consumption, liquid assets and the liquid-asset ratio, net wealth, total debt, the debt-service ratio, interest paid, household size, the age, sex, education, occupation and marital status of the head, housing tenure, indicators for elderly and single-parent households and for metropolitan residence, the lagged income quintile, and year. In the causal forest the change in other income enters as an additional control, so that the estimated effect holds movements in other income fixed as in (1). Every characteristic is measured at $t-1$. Section 3 explains why this matters: a contemporaneous covariate is an outcome of the transfer, and including one attenuates the estimate severely. Standard errors are clustered on households throughout, and the forests use 2,000 trees with a minimum leaf size of 30.

4.2 The Conditional MPC Function

Table 9 summarises the distribution of the fitted coefficients across households. The conditional MPC out of public transfers ranges from -0.030 at the fifth percentile to 0.208 at the ninety-fifth, with a mean of 0.084. The corresponding function for other income is shifted upward and is similarly dispersed, from 0.048 to 0.245 with a mean of 0.117. Households are therefore predicted to differ substantially in how they spend an additional won of transfer income: the central ninety percent of households span a range of 0.24, and in the orthogonalised estimates of Panel B, where all values are positive, the ninety-fifth percentile is seventeen times the fifth.

Table 9. The conditional MPC function

	5%	25%	50%	75%	95%	Mean
Panel A: `lm_forest` *coefficients*						
Public transfers $\beta(x)$	-0.030	0.038	0.080	0.125	0.208	0.084
Other income $\gamma(x)$	0.048	0.073	0.103	0.142	0.245	0.117
Panel B: `causal_forest`*, public transfers*						
$\hat{\tau}(x)$	0.010	0.050	0.079	0.113	0.169	0.102

Notes: Out-of-bag predictions on 132,965 household-year first differences. The average treatment effect from the causal forest is 0.1017 with a standard error of 0.0090. Forests use 2,000 trees, a minimum leaf size of 30, and household clustering.

Panel (a) of Figure 3 plots the two densities. Both are right-skewed and centred well above zero, but the transfer distribution sits to the left of the other-income distribution and has visible mass below zero, which the other-income distribution does not.

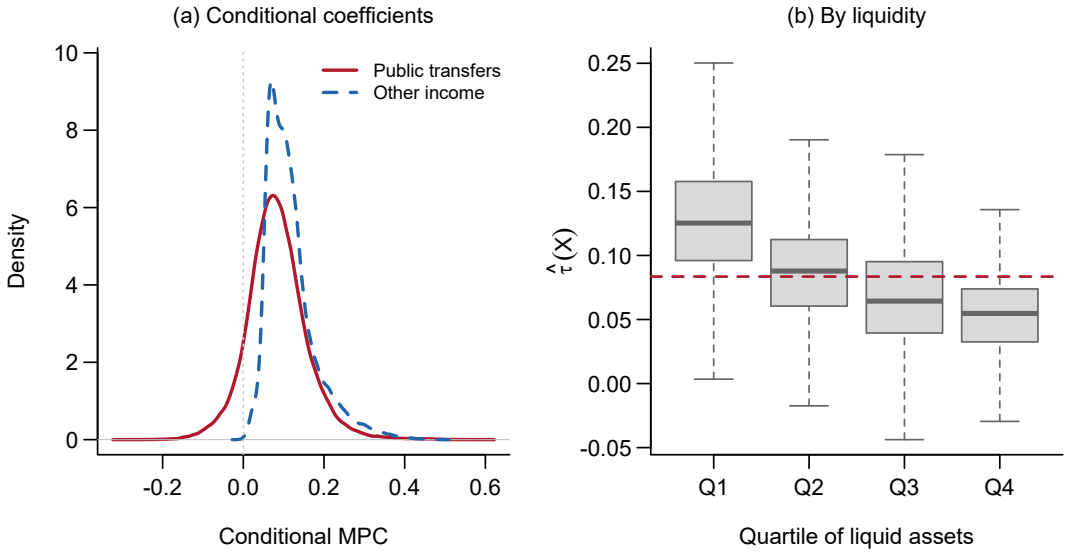


Figure 3. The conditional MPC function

Notes: Panel (a) shows kernel densities of the out-of-bag `lm_forest` coefficients on public transfers and on other income. Panel (b) shows the distribution of the causal-forest conditional effects $\hat{\tau}(X)$ within quartiles of predetermined liquid assets; boxes span the interquartile range, whiskers extend to 1.5 times that range, and the dashed line marks the sample mean.

The causal forest gives an average transfer MPC of 0.102 with a standard error of 0.0090. This sits at the top of the estimator range reported in Table 7, which is expected: the average treatment effect from a forest weights observations by the local variance of the treatment, so it need not coincide with an unweighted regression coefficient. The gap between the least-squares figure of 0.032 and the forest's 0.102 is a further instance of the estimator sensitivity documented in Section 3, and we carry both figures forward rather than choosing between them.

4.3 Is the Heterogeneity Real?

Dispersion in fitted values is not by itself evidence of heterogeneity; a forest will produce a spread of predictions even when the true effect is constant. We therefore apply the omnibus calibration test of Chernozhukov et al. (2018b), which regresses the outcome on two constructed regressors—the mean forest prediction and the demeaned forest prediction—using held-out data. A coefficient of one on the first indicates that the average effect is correctly calibrated; a coefficient of one on the second indicates that the forest's ranking of households captures genuine variation.

Table 10 reports both coefficients at essentially exactly one, each overwhelmingly significant. The heterogeneity the forest detects is real, and its average effect is well calibrated. This is the strongest single piece of evidence in the paper that MPC heterogeneity in Korean data is not a statistical artefact, and it is what licenses the

Table 10. Calibration test for heterogeneity

	Coefficient	(s.e.)	p-value
Mean forest prediction	1.019	(0.073)	$< 2.2 \times 10^{-16}$
Differential forest prediction	1.021	(0.165)	3.15×10^{-10}

Notes: One-sided heteroskedasticity-robust standard errors. Both coefficients are statistically indistinguishable from unity and significantly different from zero.

distributional analysis of Section 5.

We also computed the rank-weighted average treatment effect of Yadlowsky et al. (2025) on a binarised version of the treatment, splitting the sample so that the prioritisation rule and the evaluation come from disjoint households. The estimate is uninformative—−2.18 with a standard error of 9.97—but this reflects a failure of the binarisation rather than of the forest: dichotomising the transfer change at its median produces estimated propensities ranging from 0.10 to 0.995, far too close to the boundary for the required overlap. We report the calibration test as our evidence on heterogeneity and treat the rank-weighted statistic as inapplicable here.

4.4 Which Characteristics Account for the Heterogeneity?

The natural next question is which covariates drive $\hat{\tau}(x)$. A common answer in applied work is to read off the forest's variable importance, which ranks baseline income first (0.36), baseline consumption second (0.26) and net wealth third (0.21), with liquid assets at 0.04 and the debt-service ratio last at 0.01. Variable importance, however, counts splits, and continuous variables with many distinct values attract splits mechanically. It supports no inference.

For that we use the best linear projection of Semenova and Chernozhukov (2021), which regresses the doubly robust scores for the conditional effect on a chosen set of characteristics and delivers valid standard errors. Table 11 reports the result.

Only two characteristics survive a joint test. Baseline income enters negatively and significantly: households with more resources respond less, the standard prediction. The debt-service ratio also enters negatively and significantly. Neither liquid assets, nor net wealth, nor demographics carries independent explanatory power once these are controlled for.

This is worth dwelling on, because it does not match what a subgroup comparison would suggest. Table 12 reports average conditional effects within quartiles of liquid assets and quintiles of income. Both gradients are steep: the least liquid quartile has an estimated MPC of 0.169 against 0.065 for the most liquid, and the bottom income quintile 0.165 against 0.045 for the top. Panel (b) of Figure 3 shows that the

Table 11. Best linear projection of the conditional MPC

	Coefficient	(s.e.)	p-value
Baseline income	-9.24×10^{-6}	(3.31×10^{-6})	0.005
Debt-service ratio	-0.0529	(0.0261)	0.043
Liquid-asset ratio	-0.0047	(0.0034)	0.167
Household size	-0.0092	(0.0091)	0.310
Head's age	-1.45×10^{-4}	(6.00×10^{-4})	0.809
Net wealth	1.74×10^{-8}	(1.95×10^{-7})	0.929
Intercept	0.1781	(0.0465)	< 0.001

Notes: Best linear projection of the causal-forest conditional effects on the listed characteristics, all measured at $t-1$. Confidence intervals are cluster- and heteroskedasticity-robust. Complete cases account for 93.0 percent of the estimation sample.

liquidity gradient is not driven by outliers—the whole distribution of $\hat{\tau}(X)$ shifts down across quartiles. Read on its own, this looks like clear evidence for the buffer-stock mechanism.

Table 12. Conditional MPC by liquidity and income group

Liquid assets			Income		
Quartile	MPC	(s.e.)	Quintile	MPC	(s.e.)
Q1 (lowest)	0.169	(0.014)	Q1 (lowest)	0.165	(0.009)
Q2	0.108	(0.020)	Q2	0.113	(0.016)
Q3	0.064	(0.018)	Q3	0.074	(0.021)
Q4 (highest)	0.065	(0.020)	Q4	0.084	(0.022)
			Q5 (highest)	0.045	(0.036)

Notes: Average conditional effects within groups defined by predetermined liquid assets and by the lagged income quintile, with standard errors from the causal forest.

The two tables are not in conflict; they answer different questions. Liquid assets and income are strongly correlated in these data, so a household in the least liquid quartile is typically also a low-income household. The subgroup comparison attributes the gradient to whichever variable is used to form the groups, while the projection asks what each characteristic contributes holding the others fixed—and by that test the liquidity gradient is largely income operating through a correlated proxy. The distinction matters for the Korean literature. Song (2020), using a threshold model in the liquidity ratio alone, locates a sharp break at roughly two months of after-tax income; a single-variable threshold cannot separate liquidity from the resources with which it co-moves. We regard our finding as complementary rather than contradictory, with one important caveat taken up below.

The negative coefficient on the debt-service ratio admits a straightforward reading.

Because it is predetermined, reverse causation from the transfer to debt service is not at issue. Households already committed to large principal and interest payments convert less of a transfer into consumption, which is what one expects if transfers are partly absorbed by deleveraging. Baker et al. (2023) document exactly this for the 2020 US stimulus payments, a substantial share of which went to rent, mortgage and credit-card balances. The result should not, however, be over-read: Section 5.5 shows that although the debt-service ratio enters this projection significantly, it contributes almost nothing to predicting which households are in fact high responders.

Finally, a limitation specific to the liquidity result. Our measure is the stock of savings deposits in the public release of the survey. Song (2020) works with the restricted version available to Bank of Korea researchers, in which household cash holdings are recorded directly, and cash is precisely the component of liquidity most relevant to the hand-to-mouth classification of Kaplan and Violante (2014). Our variable is a noisier proxy, and classical measurement error attenuates its projection coefficient toward zero. We cannot rule out that liquidity would survive the joint test if measured as well as he measures it, and we return to this in Section 7.

5 The Distribution of the Marginal Propensity to Consume

Sections 3 and 4 characterised the average consumption response and the part of its variation that covariates can track. Both are conditional-mean objects. For the design of transfer programmes, however, what matters is the *distribution* of the marginal propensity to consume across households, and—separately—whether the households at the top of that distribution can be recognised in advance. This section takes up both questions.

Throughout we work with the first-differenced specification (1) of Section 3. One difference in sample deserves note. The forests of Section 4 accommodate missing covariate values natively, but the estimators used here—quantile regression, distribution regression and the finite mixture—require complete cases, which removes the roughly seven percent of observations with a missing balance-sheet ratio. The working sample in this section is therefore the same 123,609 household-year first differences, on 45,029 households, that the least-squares covariate decomposition of Table 6 already uses.

5.1 Quantile Marginal Propensities to Consume

We begin by replacing the conditional mean in (1) with conditional quantiles (Koenker and Bassett, 1978). For $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ we estimate

$$Q_{\Delta c_{it}}(\tau \mid \Delta\tau_{it}, \Delta y_{it}, X_{i,t-1}) \;=\; \alpha_t(\tau) + \beta(\tau)\,\Delta\tau_{it} + \gamma(\tau)\,\Delta y_{it} + X_{i,t-1}'\delta(\tau), \qquad (3)$$

so that $\beta(\tau)$ measures how the τth quantile of the consumption change shifts with transfer income. The contrast with the quantile regression forest used in the predecessor to this paper (Meinshausen, 2006) is the same one drawn in Section 4.1: there the object was a predicted quantile of the consumption level, from which no response coefficient follows directly; here the transfer coefficient itself is the parameter, and it is allowed to differ across the distribution.

Table 13 reports the estimates. The marginal propensity to consume out of public transfers rises monotonically from 0.029 at the tenth percentile to 0.098 at the ninetieth, a factor of 3.44. The gradient for other income is present but much flatter—0.071 to 0.128, a factor of 1.8—so the ratio of the transfer coefficient to the other-income coefficient rises from 0.40 at $\tau = 0.10$ to 0.77 at $\tau = 0.90$. At the bottom of the distribution a won of transfer income moves consumption less than half as much as a won of other income; at the top the two are nearly interchangeable. Transfer income is spent more unevenly across the distribution than ordinary income is.

Table 13. Quantile marginal propensities to consume

	Public transfers		Other income	
Quantile τ	$\beta(\tau)$	(s.e.)	$\gamma(\tau)$	(s.e.)
0.10	0.0285	(0.0067)	0.0713	(0.0019)
0.25	0.0405	(0.0051)	0.0784	(0.0015)
0.50	0.0463	(0.0049)	0.0830	(0.0015)
0.75	0.0646	(0.0070)	0.1037	(0.0022)
0.90	0.0980	(0.0119)	0.1275	(0.0034)
Ratio $\beta(0.90)/\beta(0.10)$	3.44		1.79	

Notes: Quantile regressions of (3) on 123,609 household-year first differences. Standard errors follow the Koenker and Bassett (1978) sandwich form with the `nid` correction for non-i.i.d. errors. All specifications include year effects and the predetermined covariates listed in Section 2. The fit at $\tau = 0.90$ produced a small number of non-positive sparsity estimates; Section 5.2 provides an independent check that does not rely on the extreme-quantile fit.

Two readings of Table 13 should be kept apart. Equation (3) identifies how the transfer shifts the conditional quantiles of the consumption change; it does not identify the response of any particular household, because the household occupying the τth

quantile need not be the same with and without the transfer (Koenker and Hallock, 2001). The statement supported by the table is therefore distributional: the upper tail of the consumption-change distribution moves about three times as much per won of transfer as the lower tail does. Whether that upper tail is populated by an identifiable set of households is the question of Section 5.3.

5.2 Distribution Regression

The quantile estimates at the extremes rest on relatively few observations. As a check that does not depend on the tails, we estimate the effect of transfer income on the entire cumulative distribution function directly. For a grid of thresholds y we run the linear probability model

$$\mathbf{1}\{\Delta c_{it} \le y\} \;=\; \alpha_t(y) + \pi(y)\,\Delta\tau_{it} + \gamma(y)\,\Delta y_{it} + X'_{i,t-1}\delta(y) + u_{it}(y), \tag{4}$$

so that $\pi(y) = \partial\,\mathbb{P}(\Delta c_{it} \le y)/\partial\,\Delta\tau_{it}$. A negative $\pi(y)$ means that transfer income makes a consumption change below y less likely, that is, it pushes the distribution to the right. This is the distribution-regression approach to counterfactual distributions of Chernozhukov et al. (2013).

Table 14 reports $\pi(y)$ at the deciles of Δc_{it}, scaled to a transfer of one million KRW. Every coefficient is negative, and all are individually significant from the third decile upward. The mass of the distribution moves to the right, and it does so most strongly in the middle: a transfer of one million KRW lowers the probability of a consumption change below the median by 0.24 percentage points.

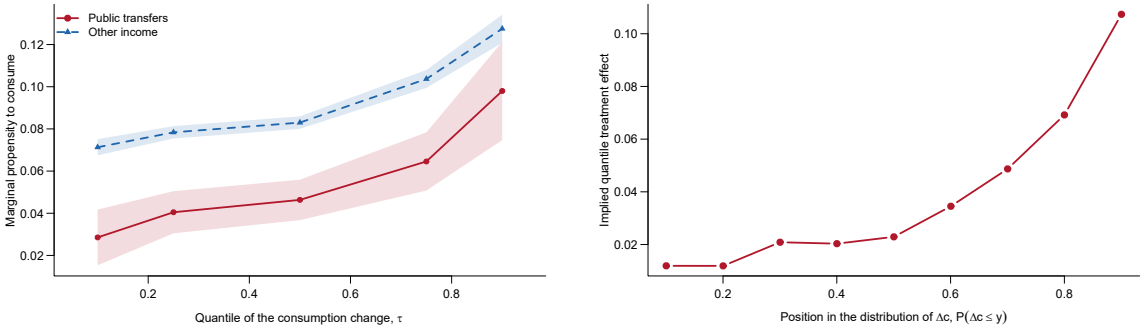
Table 14. Distribution regression and implied quantile treatment effects

Threshold y	$\mathbb{P}(\Delta c_{it} \le y)$	$\pi(y) \times 10^3$	(s.e.)	Implied QTE
−796	0.10	−0.195	(0.249)	0.012
−388	0.20	−0.431	(0.304)	0.012
−175	0.30	−1.296	(0.335)	0.021
−42	0.40	−1.931	(0.351)	0.020
49	0.50	−2.395	(0.355)	0.023
157	0.60	−2.764	(0.354)	0.035
310	0.70	−2.631	(0.337)	0.049
544	0.80	−2.346	(0.307)	0.069
980	0.90	−1.683	(0.240)	0.107

Notes: Estimates of (4) at the deciles of the consumption change. Thresholds are in 10,000 KRW. Coefficients $\pi(y)$ are reported per one million KRW of transfer income and multiplied by 10^3; standard errors are clustered on households. The implied quantile treatment effect is $-\pi(y)/\hat{f}(y)$, where $\hat{f}$ is a kernel estimate of the density of Δc_{it}.

Dividing $-\pi(y)$ by the density of Δc_{it} at y converts the vertical shift of the distri-

bution function into a horizontal one, giving the quantile treatment effect implied by (4). The final column of Table 14 rises from 0.012 at the first decile to 0.107 at the ninth. This independent route reproduces both the level and the slope of Table 13, which is reassuring given the sparsity warnings at $\tau = 0.90$. Figure 4 plots the two sets of estimates.



(a) Quantile regression **(b)** Implied by distribution regression

Figure 4. The consumption response to transfers across the distribution

Notes: Panel (a) plots $\beta(\tau)$ and $\gamma(\tau)$ from Table 13 with 95 percent confidence bands. Panel (b) plots the quantile treatment effects implied by the distribution regression in Table 14.

We also estimated (4) non-parametrically, replacing the linear probability model with a causal forest for the outcome $\mathbf{1}\{\Delta c_{it} \leq y\}$ at the quartiles of Δc_{it}. The average effects are -4.1×10^{-5}, -8.6×10^{-5} and -3.9×10^{-5} per 10,000 KRW at the first, second and third quartiles, each with a standard error an order of magnitude smaller. The forest magnitudes exceed the linear ones, as the forest's average effect exceeds the least-squares coefficient in Table 7 and for the same reason—the variance weighting of forest averages—so the comparison of levels across the two methods is not informative. What is informative is that the same inverted-U pattern in the size of the shift appears and the household-level effects are dispersed around it: neither the sign nor the shape of the distributional response depends on the linear functional form.

5.3 The Unconditional Distribution of the MPC

The estimates so far describe how transfers move the distribution of consumption changes. They do not tell us how the marginal propensity to consume is distributed *across households*, because a given quantile of Δc_{it} mixes households with different responses. To recover that distribution we follow Lewis et al. (2024) and treat the coefficient on transfer income as heterogeneous with a discrete support that is not observed. Households belong to one of K latent types, and within type k,

$$\Delta c_{it} \;=\; \alpha_k + \beta_k \, \Delta \tau_{it} + \gamma_k \, \Delta y_{it} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_k^2), \tag{5}$$

with type probabilities ω_k. We estimate (5) by maximum likelihood using the EM algorithm (Grün and Leisch, 2008). The object of interest is the implied distribution of β, that is, the set $\{(\beta_k, \omega_k)\}_{k=1}^{K}$.

Two practical matters deserve comment. First, the likelihood is multimodal, and a single random start does not reliably find the global optimum: two single-start runs on the same data returned different solutions. We therefore take the best of five random starts at every estimation, including inside the bootstrap. Second, mixture components are identified only up to a permutation, so before aggregating across bootstrap replications we order components by β_k. Both choices matter mainly for the reported precision; the point estimates differ across starting schemes by well under one bootstrap standard error, and Table A1 in Appendix A2 compares single-start and multi-start results.

Table 15 reports the three-component model, our preferred specification, with bootstrap standard errors and percentile confidence intervals from 300 replications resampled at the household level. The picture is stark. About 44 percent of households have a transfer MPC that is statistically indistinguishable from zero, and a further 34 percent have an MPC of 0.018. The remaining 23 percent respond with an MPC of 0.163, and both this coefficient and the share are precisely estimated. The probability-weighted mean, 0.042, sits inside the range of the linear estimates in Section 3, so the mixture is not manufacturing a different average—it is decomposing the same average into very unequal parts.

The between-type standard deviation of the MPC is 0.066 with a confidence interval that excludes zero, so the dispersion is not an artefact of sampling noise. Figure 5 displays the components and their precision.

Interpreting the three types as behavioural categories requires care. The Bayesian information criterion continues to fall as components are added, out to $K = 7$, the largest number we estimate, which is a familiar sign that the mixture is absorbing non-normality of the residuals as well as genuine heterogeneity in β. At $K = 7$ the component MPCs range from -0.048 to 0.260, with a high-response type of MPC 0.26 holding a 15 percent share, so the broad conclusion—wide dispersion, with a minority of high responders—survives. We report $K = 3$ in the text because it is the smallest specification that separates a high-response group, and treat the larger models as robustness in Appendix A2. What the mixture identifies under any K is the distribution of the slope coefficient in (5); reading it as a distribution of structural marginal propensities also requires the exogeneity discussed in Section 3.

Table 15. Finite mixture estimates of the MPC distribution

	Estimate	Bootstrap s.e.	95% CI	
			Lower	Upper
Component MPCs				
Type 1	−0.0031	0.0197	−0.0514	0.0247
Type 2	0.0183	0.0113	0.0010	0.0433
Type 3	**0.1632**	0.0420	0.0695	0.2425
Population shares				
Type 1	0.4382	0.0282	0.3344	0.4475
Type 2	0.3357	0.0285	0.3184	0.4407
Type 3	**0.2261**	0.0185	0.2073	0.2746
Implied moments				
Weighted mean MPC	0.0417	0.0066	0.0274	0.0542
Between-type s.d.	0.0664	0.0230	0.0201	0.1180

Notes: Three-component mixture (5) estimated by EM with five random starts, on 123,609 household-year first differences. Standard errors and percentile intervals come from 300 bootstrap replications resampling households, with components ordered by β_k within each replication to resolve label switching; all 300 replications converged to three components. Point estimates are those of the bootstrap's original sample. Re-estimating the same model from an independent set of starts gives $(-0.007, 0.018, 0.160)$ for the component MPCs and $(0.424, 0.335, 0.240)$ for the shares, differences well inside one bootstrap standard error.

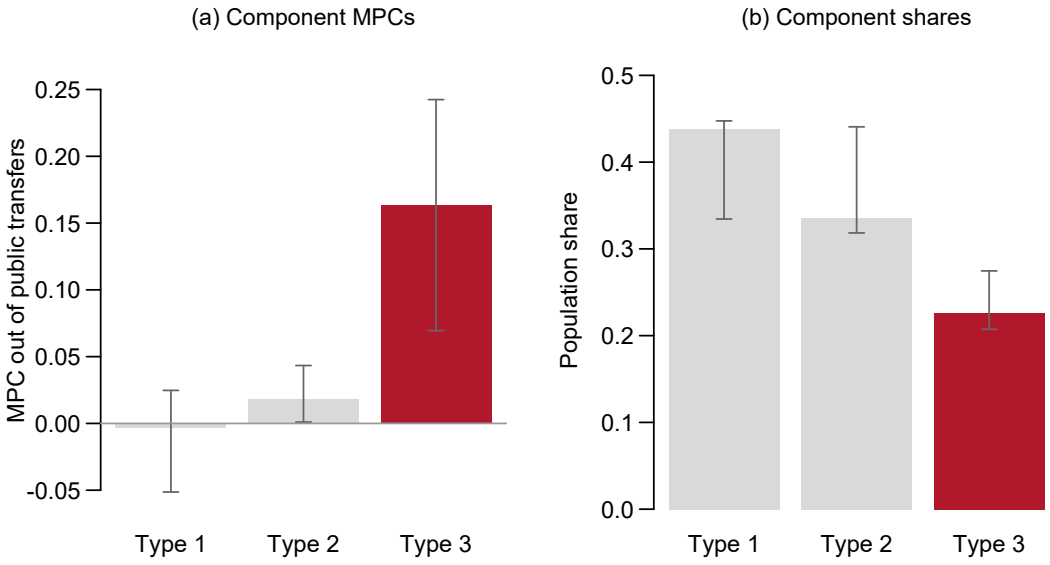


Figure 5. Latent types in the marginal propensity to consume

Notes: Component MPCs and population shares from Table 15, with 95 percent bootstrap percentile intervals. Type 3, in dark red, is the high-response group.

5.4 How Much Do Observables Explain?

The mixture gives each household a posterior distribution over types and hence a posterior expected MPC, $\widehat{\text{MPC}}_i = \sum_k \mathbb{P}(k \mid i)\,\hat{\beta}_k$. The policy question is whether this quantity can be predicted from what a programme administrator can observe. We answer it by projecting $\widehat{\text{MPC}}_i$ on the covariate set: a linear regression on income, consumption, liquid assets, the liquid-asset ratio, net wealth, debt, the debt-service ratio, household size, head's age and year effects, and a regression forest on the same variables that allows arbitrary non-linearity and interaction.

Table 16 reports the results. Under the three-component specification the linear projection recovers 5.6 percent of the variance of $\widehat{\text{MPC}}_i$, with a bootstrap confidence interval running from 1.2 to 12.0 percent. The seven-component model gives 7.0 percent, and allowing the forest to fit arbitrary non-linearities raises this to 9.7 percent out-of-bag—the most generous figure we can construct. On any of these measures, roughly ninety percent of the variation in how households spend transfer income is invisible to the variables in a rich household balance-sheet survey.

Table 16. Share of MPC heterogeneity explained by observables

	Explained share	Bootstrap s.e.	95% CI
Linear projection, $K = 3$ (preferred)	0.056	0.028	[0.012, 0.120]
Linear projection, $K = 7$	0.070	—	—
Regression forest, out-of-bag, $K = 7$	0.097	—	—
R^2 of $\widehat{\text{MPC}}_i$ on $\hat{\tau}(X)$, $K = 3$	0.044	—	—
R^2 of $\widehat{\text{MPC}}_i$ on $\hat{\tau}(X)$, $K = 7$	0.037	—	—
Memo: Lewis et al. (2024), US	≈ 0.08	—	—

Notes: The dependent variable is the posterior expected MPC from the mixture, estimated with five random starts. Covariates are the predetermined characteristics listed in Section 2 plus year effects. $\hat{\tau}(X)$ is the conditional average effect from the causal forest of Section 4. Bootstrap statistics are from the 300 household-level replications underlying Table 15, whose point estimate for the $K = 3$ linear share is 0.055.

The lower rows of Table 16 make the point in a different way. The causal forest of Section 4 estimates $\hat{\tau}(X)$, the best single predictor of the transfer response that the covariates can construct. Its correlation with the mixture-based household MPC is 0.21, so the two objects share about four percent of their variation. They are measuring largely different things.

Figure 6 shows how. Panel (a) contrasts the two distributions. The mixture-based series inherits the discreteness of the underlying types: mass piles up near the component values, and most households sit at an intermediate posterior expectation, because the data seldom assign a household to one type with confidence. Its standard deviation is 0.027. The forest's conditional means are smoothly distributed and in fact more

dispersed, with a standard deviation of 0.045—covariates do spread households out. Panel (b) shows that the two spreads have little to do with one another. The households the forest ranks as high responders are largely not the households the mixture identifies as high responders, and the fitted line through the scatter is correspondingly shallow. The covariates generate variation in predicted responses; that variation simply does not line up with the latent heterogeneity that drives aggregate spending.

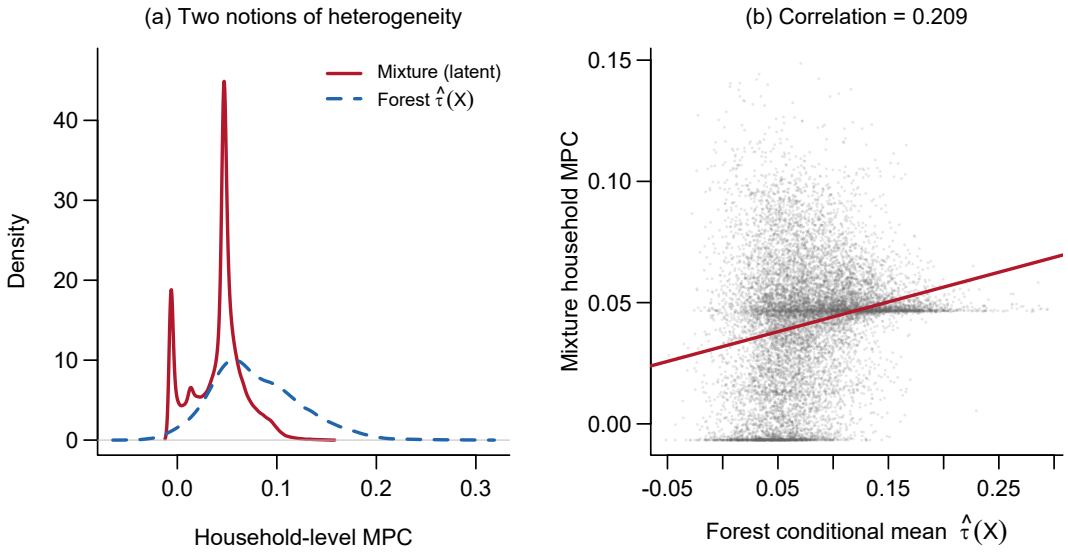


Figure 6. Latent and observable heterogeneity are nearly orthogonal

Notes: Panel (a) compares the kernel density of the household-level MPC implied by the three-component mixture (standard deviation 0.027) with that of the causal-forest conditional mean $\hat{\tau}(X)$ from Section 4 (standard deviation 0.045). Panel (b) plots the two series against each other for a random subsample of 12,000 household-year observations, with the ordinary least squares fit superimposed.

The forest is recovering the component of the response that covariates can track; the mixture is recovering a distribution that is mostly made of something else. Section 4 found that only lagged income and the debt-service ratio survive a joint test of the covariates, and the present result explains why that finding is not merely a matter of statistical power: there is little systematic variation for the covariates to capture in the first place.

Our estimates straddle the roughly 8 percent that Lewis et al. (2024) report for US households responding to the 2008 rebates, and the US figure lies comfortably inside our confidence interval. We therefore do not read the comparison as showing that Korean heterogeneity is more or less predictable than American. Figure 7 shows the bootstrap distribution together with the US benchmark. The agreement is striking given how much differs between the two exercises: a different country, a different survey, a different transfer programme, and a different estimator. That two such different settings deliver the same answer suggests the finding is a general feature of

household consumption behaviour rather than a feature of any one dataset.

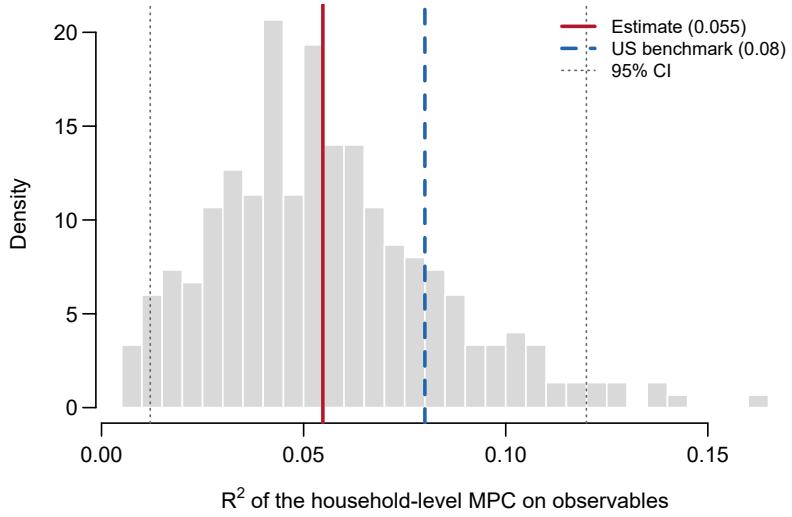


Figure 7. Bootstrap distribution of the explained share

Notes: Distribution of the R^2 from projecting the household-level MPC on observables across 300 household-level bootstrap replications. The solid line is the point estimate, the dashed line the US benchmark of Lewis et al. (2024), and the dotted lines the 95 percent percentile interval.

5.5 Which Observables Do the Work?

If observables explain a little, it is worth knowing which ones, not least because economic theory is specific about what should matter. The buffer-stock tradition points to liquid assets and borrowing capacity; the permanent-income tradition points to the household's resource level. Table 17 decomposes the explained share by variable block, reporting both the contribution of each block on its own and the loss from removing it from the full set.

The answer is unambiguous. Baseline consumption and income account for essentially the whole of the explained variation: on their own they deliver 98 percent of it, and they are the only block whose removal matters, costing 41 percent of the explained increment. Everything the buffer-stock literature directs attention to—liquid assets, net wealth, debt, the debt-service ratio—adds almost nothing once resources are controlled for. Removing the entire liquidity block costs 1.4 percent of the explained variation; removing debt and debt service costs 0.3 percent. A regression forest that allows arbitrary non-linearity reaches the same verdict, assigning 75 percent of its variable importance to baseline consumption and 20 percent to baseline income, with every balance-sheet variable below 2 percent.

The direction of the resource effect is the conventional one. Sorting households

Table 17. Which observables explain the latent MPC?

	Block on its own		Removed from full set	
Block	ΔR^2	Share	ΔR^2	Share
Baseline consumption and income	0.0534	98.1%	−0.0222	40.8%
Demographics	0.0223	40.9%	−0.0000	0.0%
Net wealth	0.0127	23.3%	−0.0000	0.1%
Liquidity	0.0088	16.1%	−0.0007	1.4%
Debt and debt service	0.0084	15.5%	−0.0002	0.3%
Other income change	0.0002	0.3%	−0.0000	0.0%
Individual variables, own contribution				
Baseline consumption	0.0521	95.8%		
Baseline income	0.0346	63.5%		
Household size	0.0210	38.7%		
Net wealth	0.0127	23.3%		
Liquid assets	0.0087	16.0%		
Debt	0.0079	14.5%		
Debt-service ratio	0.0013	2.4%		

Notes: The dependent variable is the household-level MPC from the three-component mixture. All specifications include year effects, which alone give $R^2 = 0.0019$; the full covariate set gives $R^2 = 0.0563$, so the explained increment is 0.0544. "Share" expresses each entry as a percentage of that increment. Shares in the first pair of columns exceed 100 percent in total because the blocks are correlated.

into quartiles of baseline consumption, the average latent MPC falls monotonically from 0.048 to 0.033. This is not an artefact of the size of the income shock: the same sort on the absolute size of the transfer change leaves the latent MPC essentially flat, between 0.040 and 0.044, and controlling directly for the magnitudes of both income changes still leaves baseline consumption and income contributing 0.041 of the 0.057 total R^2.

Two implications follow. First, the modest predictability documented in Section 5.4 is not evidence that we can identify liquidity-constrained households; it is close to a restatement of the fact that better-off households spend less at the margin. Second, and more sharply, the economically interesting part of MPC heterogeneity—the part that distinguishes a household with an MPC of 0.16 from an otherwise similar household with an MPC of zero—is essentially invisible in these data. This is what makes the targeting exercise of Section 6 difficult, and it also tempers the debt-service result of Section 4: the debt-service ratio enters the best linear projection of the conditional mean significantly, but it contributes almost nothing to predicting which households are actually high responders.

The measurement caveat of Section 4.4 applies here with particular force. Because our liquidity variable is a noisier proxy than the cash holdings available in the restricted release, some part of the low explained share may reflect measurement error rather

than a genuine absence of structure.

Taken together, the exercises in this section point the same way. The consumption response to transfer income is concentrated in the upper tail of the distribution of consumption changes; roughly a quarter of households account for essentially all of the aggregate response; and the identity of those households is largely beyond the reach of income, wealth, liquidity, debt and demographics, which together account for somewhere between six and ten percent of the variation in how transfers are spent—and what little they do account for is mostly the familiar fact that better-off households spend less at the margin, not any ability to detect liquidity constraints. Section 6 draws out what this implies for the design of transfer programmes.

6 Policy Implications: The Limits of Targeting

The case for targeting cash transfers rests on a simple observation: if some households spend more of a transfer than others, then concentrating a fixed budget on the high spenders raises aggregate consumption. Korea's two pandemic transfer programmes were designed around exactly this debate, the first universal and the second means-tested. The evidence in Sections 4 and 5 speaks to both halves of the argument, and it separates them: heterogeneity in the marginal propensity to consume is real and large, but the households at the top of that distribution are largely unidentifiable from the information a programme administrator would have. This section asks what that combination implies for allocation.

6.1 Learning an Allocation Rule

We use the policy learning framework of Athey and Wager (2021). A policy is a map $\pi : \mathcal{X} \to \{0, 1\}$ assigning each household either a transfer or nothing, and its value is the aggregate consumption response it generates. Given doubly robust scores $\hat{\Gamma}_i$ for the effect of receiving a transfer, the empirical value of a policy is $\hat{V}(\pi) = n^{-1} \sum_i \pi(X_i)\, \hat{\Gamma}_i$, and the learned policy maximises this over a class of decision trees.

Two design choices matter for what follows. First, tree-based policy learning requires a binary treatment, so we define receipt as a transfer increase at or above the sample median; the continuous causal forest of Section 4 remains our estimate of the response magnitude. This binarisation inherits the overlap problem that made the rank-weighted statistic of Section 4.3 uninformative, and we therefore do not rest the section's conclusion on the doubly robust scores alone: the decisive result below is corroborated by the continuous-treatment effects, which are positive essentially everywhere. Second, and more importantly, the rule must be learned and evaluated on

different households. Ranking households by an effect estimated on the same data and then reading off the average effect in the top group is subject to a winner's curse: the households that happen to have large estimated effects are disproportionately those with large positive estimation error. We therefore split the sample by household, learn the prioritisation on one half, and evaluate it on the other using doubly robust scores computed there.

This precaution is not cosmetic. Applied naively to the same sample, our estimates imply that targeting the top fifth of households would raise the effect per won by a factor of 1.84 relative to uniform allocation. The honest split-sample evaluation, reported next, gives a much weaker and less orderly picture. We report both because the gap is itself a methodological finding: the targeting gains that appear in a single-sample calculation are substantially an artefact of the ranking.

6.2 The Targeting Curve

Table 18 traces the value of allocating a fixed budget to the top-ranked k percent of households, evaluated out of sample. The estimates are positive throughout and larger than uniform allocation, but the ratios are modest—between 1.2 and 1.6—and they do not increase monotonically as targeting tightens. Narrowing from the top half to the top tenth raises the point estimate only from 1.45 to 1.62, while the standard error more than doubles, from 13.6 to 29.4. Tighter targeting concentrates the budget on households the ranking is progressively less confident about.

Table 18. The value of targeting, honestly evaluated

Share of households receiving	Effect	(s.e.)	Relative to uniform
All (uniform allocation)	43.98	—	1.00
Top 50%	63.84	(13.55)	1.45
Top 40%	65.06	(15.43)	1.48
Top 30%	71.75	(17.72)	1.63
Top 20%	53.79	(20.89)	1.22
Top 10%	71.24	(29.39)	1.62
Memo: naive same-sample calculation			
Top 50%	—	—	1.44
Top 20%	—	—	1.84

Notes: Effects are average doubly robust scores in 10,000 KRW, evaluated on households held out from the estimation of the prioritisation rule. The sample of 132,965 household-year observations is split by household into 22,620 training and 22,621 evaluation households. The memo lines rank households by their in-sample conditional effect and report the ratio of the mean effect in the top group to the overall mean.

6.3 The Learned Rule Does Not Beat Universal Transfers

The targeting curve holds the budget fixed and asks how far it should be concentrated. A policy tree answers a different and more practical question: given the covariates, which households should be excluded? Figure 8 displays the curve, and the learned depth-two tree splits first on the liquid-asset ratio at 1.33 years of disposable income, then on net wealth at 267 million KRW in the low-ratio branch and on liquid assets at 296 million KRW in the high-ratio branch. Both thresholds fall well into the upper tail—the sample medians are 173 million KRW for net wealth and 30 million for liquid assets—so the rule is excluding the visibly affluent rather than drawing a line near the middle of the distribution. It assigns transfers to 80 percent of households, and those it selects are markedly poorer than those it excludes: median baseline income of 29.0 against 64.6 million KRW, median liquid assets of 25.0 against 50.4 million, and median net wealth of 122 against 537 million.

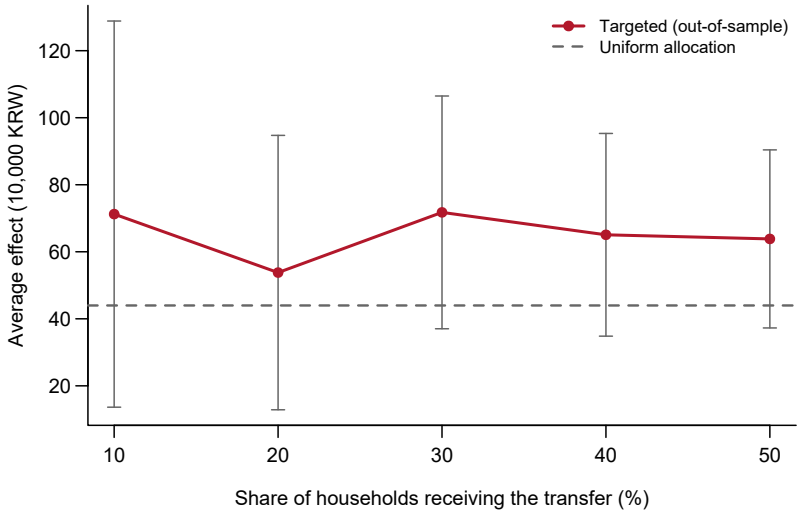


Figure 8. The targeting curve

Notes: Average doubly robust effect among households ranked in the top k percent by a conditional effect estimated on a disjoint set of households. The dashed line marks uniform allocation.

The rule therefore does what one would expect: it targets the asset-poor. The question is whether it is worth applying. Evaluated on the held-out households, the policy value of the learned rule is 409,000 KRW per household, against 482,000 for giving to everyone. (The universal benchmark differs from the 440,000 of Table 18 because applying the tree requires complete covariate cases; both figures in the comparison are computed on the same subsample, which is what matters.) *The learned exclusion rule delivers less aggregate consumption stimulus than a universal transfer.*

The reason is visible in the estimated conditional effects. For this exercise we re-

estimate the causal forest on the reduced covariate set an allocation rule could plausibly use—balance-sheet variables, household size, age, education and tenure, but not, for instance, lagged consumption. On that set the average transfer MPC is 0.088 with a standard error of 0.011, slightly below the 0.102 of Section 4, and the conditional effects are positive essentially everywhere: the fifth percentile of $\hat{\tau}(X)$ is 0.007. Among the 20 percent of households the rule excludes, the average conditional effect is still 0.036. Excluding them removes a fifth of the budget from households who would have spent a non-trivial share of it, and the concentration gain on the remaining 80 percent does not compensate.

6.4 Efficiency versus Scale

These two results are consistent, and together they define the policy problem more sharply than either does alone. Targeting does raise the consumption response per won: Table 18 shows gains of roughly 20 to 60 percent in effect per unit of budget. But a rule that excludes households in order to achieve that concentration reduces the total response, because the excluded households are not in fact non-spenders. The choice is between efficiency and scale, not between a good policy and a bad one, and which side to prefer depends on whether the binding constraint is the budget or the target level of aggregate stimulus.

Section 5 explains why the efficiency gain is so limited. A targeting rule can only exploit heterogeneity that is visible in the covariates, and the covariates capture at best a tenth of the variation in how households spend transfers—most of that being the unremarkable fact that better-off households spend less at the margin. The households with a marginal propensity to consume of 0.16 are not systematically the low-liquidity or low-wealth households; they are distributed through the population in a way that income, wealth, liquidity and debt do not track. A rule built on those variables selects households who are poorer, not households who spend more.

This qualifies a common argument in the Korean policy debate. That the consumption response is larger among low-income households is well documented, here and elsewhere, and it is often taken to establish that means-tested transfers dominate universal ones. The subgroup gradient is real—the bottom income quintile has an estimated MPC of 0.165 against 0.045 for the top—but it does not survive translation into an allocation rule. What our results suggest is not that targeting is pointless, but that its value should be assessed with an honest out-of-sample evaluation rather than inferred from a subgroup comparison, and that on that test the case for excluding households from a transfer programme is weak in these data.

Two caveats bound the conclusion. First, our evaluation concerns aggregate con-

sumption only. Transfers also serve distributional and insurance objectives, for which targeting the asset-poor may be desirable irrespective of the consumption response; a household that repays debt rather than consuming is not thereby made no better off. Second, the argument is conditional on the information set. Administrators with access to richer data—in particular to cash holdings, which Song (2020) shows are recorded in the restricted version of this survey but not the public one—might construct rules that perform better than ours. What we can say is that the variables available in one of the most detailed household balance-sheet surveys in the world do not support effective targeting.

7 Conclusion

Using fourteen waves of a Korean household survey, we have asked not how much of a transfer the average household spends but how that response is distributed and whether the households at the top of the distribution can be recognised in advance. The answers point in a consistent direction, and they matter for how cash transfer programmes are designed.

Heterogeneity in the marginal propensity to consume is real. The calibration test in Section 4 rejects a constant response decisively, with a differential prediction coefficient of 1.02 that is indistinguishable from unity and significant at $p = 3 \times 10^{-10}$. A finite mixture puts structure on it: roughly three-quarters of households have a transfer MPC close to zero, while about a quarter respond with an MPC of 0.16. The probability-weighted mean is 0.042, so the aggregate response that a representative-agent calculation would attribute to all households is in fact produced by a minority of them.

That minority cannot be identified from what surveys observe. Income, wealth, liquid assets, debt, debt service and demographics together account for somewhere between six and ten percent of the variation in household-level responses, and the part they do account for is largely the unremarkable fact that better-off households spend less at the margin rather than any ability to detect binding liquidity constraints. Our estimate is statistically indistinguishable from the roughly eight percent that Lewis et al. (2024) report for the United States. The agreement across two countries, two surveys, two transfer programmes and two estimators suggests that this is a general feature of consumption behaviour rather than an artefact of any one setting.

The policy implication is not the one the subgroup literature suggests. Because poorer households respond more, means testing is routinely presented as the more efficient design. But an allocation rule learned from the covariates and evaluated on held-out households delivers less aggregate stimulus than giving to everyone—409,000

against 482,000 KRW per household—because the households it excludes are not non-spenders. Targeting does raise the response per won of budget, so the real choice is between efficiency and scale rather than between a better and a worse policy. And the efficiency gain available is small, for the reason the preceding paragraph gives: the heterogeneity that observable characteristics track is not the heterogeneity that determines who spends.

A methodological finding runs alongside these. Estimates of the average marginal propensity to consume from a single dataset, a single estimand and a single covariate set range from 0.032 to 0.107 across standard estimators, and the coefficient reverses sign when the change in other income is omitted—a single omitted variable, with a transparent mechanism, since transfers rise when other income falls. Separately, we show that several Korean studies report elasticities rather than level propensities, and that converting them removes much of the apparent disagreement in the literature: Song (2020)'s transitory-income estimate, expressed in level terms, contains ours. Researchers comparing estimates across studies should check units before concluding that findings conflict.

These conclusions carry limits that we have tried to keep visible throughout. Identification is selection on observables—the two quasi-experimental designs we attempted fail for the reasons quantified in Section 3—so our estimates are partial associations conditional on predetermined characteristics. Survey-reported consumption carries measurement error, which attenuates estimated propensities and makes our levels best read as lower bounds. The mixture identifies the distribution of a slope coefficient under within-component normality, and the information criterion favours more components than we report, so the three types are a parsimonious summary rather than behavioural categories. And because our liquidity measure is coarser than the cash holdings available in the survey's restricted release, part of the small explained share may reflect measurement error rather than a true absence of structure.

The same limits point to the natural next steps. Linking survey households to card or administrative transaction records would combine the survey's balance-sheet detail with the timing precision that identification requires; the panel is now long enough to support estimation of the intertemporal responses that heterogeneous-agent models take as inputs (Auclert et al., 2024); and the finding that observables explain so little of who spends makes survey instruments that elicit intended responses to hypothetical windfalls (Colarieti et al., 2024) the most promising route to the preferences and expectations that balance-sheet data cannot reach.

Data Availability Statement

The data that support the findings of this study are the public-use microdata of the Survey of Household Finances and Living Conditions (2012–2025), produced jointly by Statistics Korea, the Bank of Korea and the Financial Supervisory Service and distributed through Statistics Korea's Microdata Integrated Service (MDIS) at `https://mdis.kostat.go.kr` (Statistics Korea et al., 2025). The microdata are available free of charge to registered users; under the MDIS terms of use they may not be redistributed by third parties, so the raw files cannot be deposited with the journal. Complete replication materials—the R and Python code, the corrected variable mapping documented in Appendix A1, the MDIS extraction settings for each survey wave, and the aggregate estimation outputs from which every table and figure is produced—will be deposited in the *Review of Income and Wealth* repository maintained by the IARIW upon acceptance, so that any researcher who obtains the microdata from MDIS can reproduce the results in full.

References

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2):1148–1178.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

Auclert, A., Rognlie, M., and Straub, L. (2024). The intertemporal Keynesian cross. *Journal of Political Economy*, 132(12):4068–4121.

Aydin, D. (2022). Consumption response to credit expansions: Evidence from experimental assignment of 45,307 credit lines. *American Economic Review*, 112(1):1–40.

Baek, S., Kim, S., Rhee, T.-h., and Shin, W. (2023). How effective are universal payments for raising consumption? evidence from a natural experiment. *Empirical Economics*, 65:1–30.

Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., and Yannelis, C. (2023). Income, liquidity, and the consumption response to the 2020 economic stimulus payments. *Review of Finance*, 27(6):2271–2304.

Campbell, J. and Mankiw, N. (1989). Consumption, income and interest rates: Reinterpreting the time series evidence. NBER Chapters, National Bureau of Economic Research.

Carroll, C. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics*, 112(1):1–55.

Carroll, C., Slacalek, J., Tokuoka, K., and White, M. (2017). The distribution of wealth and the marginal propensity to consume. *Quantitative Economics*, 8(3):977–1020.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018a). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2018b). Generic machine learning inference on heterogeneous treatment effects in randomized experiments. NBER Working Paper 24678, National Bureau of Economic Research.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Colarieti, R., Mei, P., and Stantcheva, S. (2024). The how and why of household reactions to income shocks. NBER Working Paper 32191, National Bureau of Economic Research.

Deaton, A. (1991). Saving and liquidity constraints. *Econometrica*, 59(5):1221–1248.

Fagereng, A., Holm, M. B., and Natvik, G. J. (2021). MPC heterogeneity and household balance sheets. *American Economic Journal: Macroeconomics*, 13(4):1–54.

Fishera, J. D., Johnsonb, D. S., Smeeding, T. M., and Thompsond, J. P. (2020). Estimating the marginal propensity to consume using thedistributions of income, consumption, and wealth. *Journal of Macroeconomics*, 65:103–118.

Fuhr, J. and Papies, D. (2024). Double machine learning meets panel data: Promises, pitfalls, and potential solutions. *arXiv preprint arXiv:2409.01266*.

Gelman, M. (2021). What drives heterogeneity in the marginal propensity to consume? temporary shocks vs persistent characteristics. *Journal of Monetary Economics*,

117:521–542.

Grün, B. and Leisch, F. (2008). FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*, 28(4):1–35.

Hong, K. (2021). Analysis of the marginal propensity to consume considering household characteristics. Technical report, National Assembly Budget Office. In Korean.

Jappelli, T. and Pistaferri, L. (2014). Fiscal policy and mpc heterogeneity. *American Economic Journal*, 6:107–136.

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

Kaplan, G. and Violante, G. L. (2022). The marginal propensity to consume in heterogeneous agent models. *Annual Review of Economics*, 14:747–775.

Kim, M. and Oh, Y. (2020). Analysis of the effects of the emergency disaster relief payments. Technical report, Korea Development Institute. In Korean.

Kim, S., Koh, K., and Lyou, W. (2023). Means-tested COVID-19 stimulus payment and consumer spending: Evidence from card transaction data in South Korea. *Economic Analysis and Policy*, 78:1359–1371.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156.

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

Lee, W., Kang, C., and Woo, S. (2022). The effects of the 2020 COVID-19 emergency income support on household consumption in Korea. *Korean Journal of Economic Studies*, 70(1). In Korean.

Lewis, D. J., Melcangi, D., and Pilossoph, L. (2024). Latent heterogeneity in the marginal propensity to consume. *Review of Economic Studies*. Also NBER Working Paper 32523 and FRB New York Staff Report 902.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7:983–999.

Nie, X. and Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319.

Noh, Y.-H. (2021). Consumption effects of transfer payments: Estimating the marginal propensity to consume. *Health and Social Welfare Review*, 41(2). In Korean.

Orchard, J., Ramey, V. A., and Wieland, J. F. (2025). Micro MPCs and macro counterfactuals: The case of the 2008 rebates. *Quarterly Journal of Economics*, 140(3):2001–2052.

Semenova, V. and Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *Econometrics Journal*, 24(2):264–289.

Sokolova, A. (2023). Marginal propensity to consume in recessions: A meta-analysis. *Review of Economic Dynamics*.

Song, S.-y. (2020). Leverage, hand-to-mouth households, and heterogeneity of the marginal propensity to consume: Evidence from South Korea. *Review of Economics*

of the Household, 18(4):1213–1244.

Statistics Korea, Bank of Korea, and Financial Supervisory Service (2012–2025). [Dataset] Survey of Household Finances and Living Conditions: Public-Use Microdata, 2012–2025 Waves. Statistics Korea Microdata Integrated Service (MDIS), `https://mdis.kostat.go.kr`. Accessed August 2026.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., and Wager, S. (2025). Evaluating treatment prioritization rules via rank-weighted average treatment effects. *Journal of the American Statistical Association*, 120(549):38–51.

Zeldes, S. (1989). Optimal consumption with stochastic income: deviations from certainty equivalence. *Quartlerly Journal of Economics*, 104(2):275–298.

A1 Data Appendix

This appendix documents two defects in the source materials, both corrected before any of the analysis reported in the paper, and the construction of the panel.

Transposed transfer income. The variable-name mapping distributed with the microdata transposes public and private transfer income in every wave from 2012 through 2018. The error is detectable by comparison with the official published aggregates: for Korean households public transfers are several times larger than private transfers, whereas the mapping as distributed implies the reverse. Under the corrected mapping, mean public transfer income rises smoothly from 1.92 million KRW in 2012 to 5.00 million in 2020, and private transfers are stable between 0.73 and 1.16 million; under the mapping as distributed the two series are exchanged. Because the predecessor analysis on which this project builds used disposable income and net wealth rather than the transfer components, its results are unaffected; all results in the present paper use the corrected mapping.

Duplicated label in the 2019 file. The 2019 raw file contains the label for the principal-repayment variable twice and omits the label for interest paid. The second of the duplicated columns is in fact interest, which the accounting identity confirms: with that assignment, principal plus interest equals total debt service exactly in every observation and in every year, with mean absolute error zero. Our construction routine applies this recovery automatically and verifies the identity.

Panel construction. Waves are linked on the household identifier after verifying uniqueness within each wave. Every characteristic used as a covariate is stored at $t-1$ as well as t, so that the analysis can be restricted to predetermined variables. The consumption sample break described in Section 2.3 is recorded in a section indicator available for 2012–2018.

A2 Robustness

This appendix reports five sample-based robustness exercises, each re-estimating two quantities—the transfer coefficient from the full specification of Section 3, and the three-component mixture of Section 5, from which we take the high-response component, its population share, and the share of household-level variation explained by observables—followed by two checks specific to the mixture: the number of components and the starting scheme. Table A1 collects the results.

Table A1. Robustness of the main estimates

	Transfer MPC		Mixture, $K = 3$		
Sample	Estimate	(s.e.)	High type	Share	R^2
Baseline	0.038	(0.006)	0.163	0.226	0.055
Pandemic waves					
Excluding 2021–22 waves	0.044	(0.008)	0.152	0.222	0.040
Excluding 2021–23 waves	0.045	(0.009)	0.195	0.240	0.042
Trimming					
Trimming at 0.1%	0.037	(0.007)	0.071	0.485	0.060
Trimming at 1.0%	0.039	(0.006)	0.189	0.225	0.050
Trimming at 2.5%	0.043	(0.006)	0.266	0.200	0.033
Sample composition					
2019–25 waves only	0.032	(0.007)	0.158	0.263	0.048
Attrition					
Long-spell households (5+ waves)	0.042	(0.008)	0.140	0.249	0.064
Non-leavers only	0.038	(0.007)	0.204	0.209	0.061

Notes: Each row re-estimates the specification of Table 5, column (5), and the mixture of Table 15 on the indicated sample. Standard errors are clustered on households; the mixture is estimated from five random starts.

Consumption categories. Table A2 decomposes the total response into the six expenditure categories the survey records separately. Food alone accounts for 0.018 of the total 0.038, close to half, while housing, health, transport and communication each contribute between 0.003 and 0.005 and education is slightly negative. The six categories sum to 0.031 against a total of 0.038, the residual falling in the "other consumption" item that the survey does not break out; that the components nearly reconcile is a useful check, since each is collected as a separate question. Transfers are converted into necessities rather than spread evenly across the consumption basket—a pattern consistent with the finding in Section 5 that the response is concentrated among lower-resource households.

Pandemic waves. The 2021 and 2022 waves record the two transfer programmes and also the years of greatest disruption to consumption. Dropping them raises the estimated transfer MPC from 0.038 to 0.044, and dropping 2023 as well raises it to 0.045; the high-response component and its share are essentially unchanged. Our results are not produced by the pandemic episode. If anything the pandemic waves attenuate the estimate, which is consistent with consumption in 2020 having been constrained by factors other than resources.

Table A2. Transfer MPC by consumption category

Consumption category	MPC	(s.e.)
Food	0.018	(0.003)
Housing	0.004	(0.001)
Education	-0.002	(0.002)
Health	0.005	(0.002)
Transport	0.004	(0.001)
Communication	0.003	(0.001)
Total consumption	0.038	(0.006)

Notes: Each row estimates the specification of Table 5, column (5), with the change in the indicated expenditure category as the dependent variable. Categories are collected as separate survey items and need not sum exactly to total consumption.

Trimming. We trim the top and bottom half-percent of the consumption and income change distributions. The transfer coefficient is insensitive to this choice, moving between 0.037 and 0.043 as the threshold varies from 0.1 to 2.5 percent. The mixture is not equally insensitive. At the mildest trimming the algorithm finds a quite different configuration, with a high-response component of 0.071 holding 48.5 percent of the population; at the most aggressive it finds a component of 0.266 holding 20.0 percent. Between 0.5 and 2.5 percent the share is stable between 0.20 and 0.23 while the component MPC rises with the threshold.

We read this as a genuine limitation. The qualitative conclusion—that a minority of households accounts for the aggregate response—holds throughout, and the explained-share result is stable at 0.033 to 0.060. But the magnitude of the high-response component and, at the extreme, its share depend on how outliers are handled, and readers should treat the point estimates in Table 15 as conditional on our trimming rule rather than as sharply identified.

Sample composition. Section 2.3 describes the welfare and finance sections into which the survey was divided between 2013 and 2018. Because consumption was collected only in the welfare section, the estimation sample in those waves consists entirely of welfare-section households by construction: of the 109,898 household-year observations in 2013–2018, all 54,692 with consumption data are in the welfare section and all 55,206 without are in the finance section. There is therefore no selection between sections to correct for, and restricting to the welfare section throughout leaves the estimates numerically identical. The remaining question is whether the smaller mid-sample waves distort the pooled estimate. Restricting to the 2019–25 waves, in which the sections were merged and the full sample carries consumption data, gives a

slightly lower transfer coefficient of 0.032 with an unchanged mixture structure.

Attrition. The survey's rotating design means households leave the panel; 23.3 percent of household-year observations are not followed into the next wave. Attrition does not appear to be selective on the dimensions that matter here. Median baseline income among households that leave is 35.2 million KRW against 33.8 million among those that stay, liquid assets 32.9 against 29.1 million, net wealth 184 against 177 million, and the age and size of the household are the same. Restricting to households observed in at least five waves gives a transfer MPC of 0.042, and excluding observations that are not followed gives 0.038, both within a standard error of the baseline.

Number of mixture components. The Bayesian information criterion falls monotonically from 859,151 at $K = 1$ to 825,280 at $K = 7$, the largest number of components we estimate. We report $K = 3$ in the text as the smallest specification that separates a high-response group, and note that at $K = 7$ the component MPCs range from -0.048 to 0.260, so the qualitative conclusion is unchanged.

Multi-start estimation. The EM algorithm converges to different local optima from different random starts: two single-start runs on the same data returned component MPCs of $(0.002, 0.019, 0.164)$ and $(-0.025, 0.017, 0.175)$ respectively. All reported estimates take the best of five random starts. The bootstrap in Table 15 was run under both schemes; point estimates are materially identical, while the multi-start scheme narrows the confidence interval for the high-response share from $[0.187, 0.403]$ to $[0.207, 0.275]$.

A3 Methodological Details

Label switching. Mixture components are identified only up to a permutation. Before aggregating across bootstrap replications we order components by the transfer coefficient β_k within each replication. Without this correction the bootstrap distribution of any component-specific quantity is a mixture over labellings and its dispersion is meaningless.

Quantile treatment effects from distribution regression. Let $F(y) = \mathbb{P}(\Delta c_{it} \leq y)$ and let $\pi(y) = \partial F(y)/\partial \Delta\tau_{it}$ be the distribution-regression coefficient at threshold y. A shift of the distribution function at y translates into a horizontal shift of the quantile function through $\partial Q(\tau)/\partial \Delta\tau_{it} = -\pi(y_\tau)/f(y_\tau)$, where $y_\tau = Q(\tau)$ and f is

the density. We estimate f by a Gaussian kernel on the pooled distribution of Δc_{it} and evaluate it at the decile thresholds.

Cross-fitting with panel data. Double machine learning requires that observations used to fit the nuisance functions be independent of those used to estimate the parameter of interest. Assigning household-year observations to folds at random violates this, since different years of the same household are then split across folds (Fuhr and Papies, 2024). We form folds by household throughout, and likewise split by household in every sample-splitting exercise in the paper.