Latent Heterogeneity in the Marginal Propensity to Consume*

Daniel Lewis
FRBNY

Davide Melcangi
FRBNY

Laura Pilossoph†
FRBNY

November 13, 2020

Abstract

We estimate the distribution of marginal propensities to consume (MPCs) using a novel clustering approach, generalizing the fuzzy C-means algorithm to regression settings. We find that households spent at least one fourth of their 2008 stimulus payments, with considerable MPC heterogeneity which varies by consumption good. We document observable determinants of this heterogeneity, without imposing *ex ante* assumptions on such relationships, and evaluate potential determinants jointly. MPCs correlate positively with income and the average propensity to consume, but much heterogeneity remains unexplained. The partial equilibrium response to the stimulus is twice the homogeneous estimate, highlighting the importance of heterogeneity.

Keywords: Marginal Propensity to Consume, Consumption, Tax Rebate, Heterogeneous Treatment Effects, Clustering, C-means
JEL Codes: D12, D91, E21, E32, E62

---

*This paper was previously circulated as "A New Approach to Estimating Heterogeneous Effects and the Distribution of the Marginal Propensity to Consume". The views expressed herein are solely those of the authors and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System. We thank René Chalom and Meghana Gaur for excellent research assistance. For helpful comments, we also thank Michael Boutros, Sara Casella (discussant), Richard Crump, Marco Del Negro, Keshav Dogra, Domenico Giannone, Simon Gilchrist, Greg Kaplan, Ulrich Müller, Eric Nielsen (discussant), Mikkel Plagborg-Møller, Isaac Sorkin, Mary Wootters, and various seminar and conference participants.

†daniel.lewis@ny.frb.org, davide.melcangi@ny.frb.org, laura.pilossoph@ny.frb.org.

1 Introduction

Recent work highlights the importance of heterogeneity in the marginal propensity to consume (MPC) out of transitory income shocks for fiscal policy, the transmission of monetary policy, and welfare.[1] Despite their importance, estimates of the *distribution* of MPCs are largely elusive. Even with plausibly identified transitory income shocks, estimating individual-level MPCs requires panel data with long horizons, which are typically not available; it also usually requires the unappealing assumption that an individual's marginal propensity to consume (MPC) is time invariant.[2] The existing literature, therefore, has followed one of two avenues: estimating a fully structural model and simulating a distribution of MPCs, or grouping observations by some presupposed observable characteristics and estimating group-specific MPCs out of transitory income shocks.[3] However, because both of these approaches require taking a stance on the source of MPC heterogeneity, they may fail to uncover the true degree of heterogeneity, miss other relevant dimensions of heterogeneity that predict an individual's MPC, or both.

In this paper, we propose a novel methodology to estimate the distribution of MPCs directly. We develop an estimator based on the fuzzy C-means (FCM) clustering algorithm (Dunn (1973), Bezdek (1973)), which jointly (i) groups households together that have similar latent consumption responses to the 2008 tax rebate and (ii) provides estimates of the MPCs within these groups. Specifically, the algorithm takes a standard regression of consumption changes on the tax rebate receipt and basic controls (Johnson, Parker, and Souleles (2006), Parker, Souleles, Johnson, and McClelland (2013)), but allows the coefficient on the rebate to be heterogeneous across unknown groups; the groups as well as their rebate coefficients are jointly estimated.

This approach offers several advantages over existing efforts to recover the distribution of MPCs. First, it allows us to estimate the full unconditional distribution of MPCs, which can be driven both by latent factors and observable characteristics, broadly defined. Standard methods that rely on sample splitting by the latter cannot recover heterogeneity

[1] The MPC distribution is a crucial object in Heterogeneous Agent New Keynesian (HANK) models of monetary policy (see Kaplan, Moll, and Violante (2018)). For example, Auclert (2019) shows that the response of aggregate consumption to monetary policy shocks depends on the covariance of the distribution of MPCs with the cyclicality of income, net nominal position, and unhedged interest rate exposure.

[2] Nearly all theories of MPC heterogeneity have some form of state dependence. For example, in Carroll (1992) the MPC is a declining function of gross household wealth.

[3] For the former, see for instance Kaplan and Violante (2014) and Carroll, Slacalek, Tokuoka, and White (2017). For the latter, Fagereng, Holm, and Natvik (2016) exploit randomized lottery winnings to identify transitory income shocks, and subsequently group observations on observables to estimate group-level MPCs. See also Johnson et al. (2006), Blundell, Pistaferri, and Preston (2008), Parker et al. (2013), Kaplan, Violante, and Weidner (2014), and Crawley and Kuchler (2018).

in MPCs associated with the former (by definition). Yet, we find that the majority of MPC heterogeneity can be attributed to such latent characteristics. Second, because our approach does not require taking an ex-ante stand on what observables correlate with MPC heterogeneity, we can "let the data speak" by investigating *ex post* which observables predict the recovered individual MPCs. We find that a household's MPC is individually correlated with various observable characteristics, and that these relationships are statistically significant. Hence, our approach overcomes the loss of statistical power which may have plagued the sample-splitting approach in existing studies. Importantly, however, we further show that the significance of most of these univariate relationships with the MPC disappears when we instead examine predictors *jointly*, with the exception of non-salary income and the average propensity to consume (APC). Finally, we formally show that household observable characteristics explain only a small portion of the variance in MPCs.

Our contribution hinges on the fact that clustering algorithms like the one we adopt assign individuals to groups not based on observable characteristics, but based on how well each set of estimated group-specific parameters describe the observations within the group. This feature allows us to bypass the decision on which observables matter for MPC heterogeneity, and instead estimate the heterogeneity directly first. FCM is a "fuzzy" clustering approach, in which individuals are not assigned to groups in a binary fashion, but instead have continuous weights. When there is no panel dimension to the data (as is the case in our empirical setting) it is unrealistic to think that group assignment can be determined binarily in the presence of noise, so continuous weights are desirable to represent the level of uncertainty that exists in the assignment. If the econometrician knows the true distribution of the data, a likelihood-based approach like a Gaussian mixture model can be used, but, absent any information on the distribution, the form of weights used by FCM has optimality properties (Bezdek (1973)).

We therefore generalize FCM from the cluster means case to the regression setting, where it constitutes a weighted least squares (WLS) problem, and call the resulting generalized estimators fuzzy C-parameters (FCP).[4] We establish the consistency and asymptotic normality of a GMM formulation for the FCP problem. In a simulation study calibrated to our empirical setting, we find that FCP is able to accurately recover the distribution of MPCs, more so than quantile regressions and binary classification or likelihood-based competitors. We further show how FCP can be used in the second stage of two-stage least squares (TSLS) to accommodate the use of instrumental variables, which we

[4]Further generalizations to nonlinear models, while outside the scope of this paper, appear trivial under suitable assumptions.

exploit when we instrument the rebate value with its receipt. These characteristics mean that FCP is well-suited to studying heterogeneity in a wide variety of economic settings with cross-sectional or short-panel data, and can be used in contexts well beyond our current focus.

Applying our estimator to study the MPC distribution using the 2008 Economic Stimulus Act, we uncover a substantial degree of heterogeneity. In particular, households spend at least one fourth of the rebate within one quarter, with some households displaying an MPC above one. While the share of households with different MPCs generally decreases with the magnitude of the MPC, we estimate sizable MPCs even at the bottom of the distribution. Fagereng et al. (2016) and Olafsson and Pagel (2018) find evidence of similar behavior in Norway and Iceland respectively, and recent papers have proposed models of limited cognitive perception to rationalize such behavior (e.g., Ilut and Valchev (2020)). However, we also find that, while all households spend a non-trivial portion of the rebate overall, they often smooth nondurable consumption. Specifically, a third of the households do not use the rebate to buy additional nondurable goods. Our results are consistent across different specifications and sample restrictions. For instance, instrumenting the rebate with an indicator for its receipt, as in Parker et al. (2013), leaves the results qualitatively unchanged and in fact increases the estimated heterogeneity in MPCs.[5]

Having characterized the distribution of marginal propensities to consume, we describe its main drivers. Historically, the literature has found mixed empirical evidence and generally weak relationships between MPC and observable household characteristics, with the possible exception of liquid wealth.[6] This is likely due to a significant loss of statistical power when re-estimating the MPC with interactions or sample splits. Our approach allows us to sidestep these issues. Indeed, we find that our estimated MPCs are significantly correlated, individually, with many observable drivers, despite the fact that we use the same identification strategy and dataset that previously delivered insignificant relationships. For example, we find that homeowners have significantly higher MPCs than renters, and households with a mortgage display even greater marginal propensities to consume than outright homeowners.

Our estimates for household-level MPCs also allow us to study multivariate relationships without further losses of power. We find that only two observables are robust to the inclusion of additional controls and positively correlate with MPCs: households' non-

[5]The same is true when we exclude from the sample households that never received a rebate, or when we include lagged values of the rebate to control for persistent effects of receiving the rebate.

[6]Parker et al. (2013) find statistically insignificant differences by age, income, and liquid wealth. Broda and Parker (2014) and Fagereng et al. (2016) find significant relationships for the latter.

salary income and their APC. Kueng (2018) also finds that high-income households have higher MPCs in Alaskan data. We highlight how our result crucially hinges on the non-salary component of income, such as business and financial income. Examining how MPCs vary jointly with income and the APC, we uncover three groups of households. "Poor-savers", with low total income and a low APC, have the lowest MPCs. Households with high total income and a low APC, or vice versa, display intermediate marginal propensities to consume. The greatest MPCs are found among "rich-spenders", who not only have high total income, but also typically spend a large portion of it. This group of households has not received much attention in models of consumption and savings.

Importantly, our best array of observable predictors is able to explain only about 11% of the variation in estimated MPCs. With the vast majority of heterogeneity unexplained by standard observables, our results suggest that a relevant portion of MPC heterogeneity is driven by latent household traits. For example, heterogeneity in discount rates and/or intertemporal elasticities of substitution (Aguiar, Boar, and Bils (2019)) would deliver heterogeneity in MPCs, and is further supported by the aforementioned significance of APCs in predicting MPCs, as APCs can also be a function of the same unobserved traits. This type of unobserved heterogeneity could never be recovered by simply splitting the sample on observable characteristics and estimating within-subsample homogeneous MPCs, as is typically done in the literature.[7]

Finally, correctly accounting for MPC heterogeneity also matters for the aggregated consumption effects of the fiscal stimulus. We show that the sample average of our estimated heterogeneous responses is larger than the homogeneous marginal propensity to consume. When considering the cumulated heterogeneous responses over two quarters, the aggregated response from our distribution is almost twice as large as its homogeneous counterpart. While still a partial-equilibrium object in nature, this result suggests that correctly accounting for heterogeneity is important in order to correctly evaluate the impact of the 2008 fiscal stimulus.

Our paper is related to an extensive literature estimating the marginal propensity to consume out of transitory income shocks, and a smaller, complementary literature examining how it varies across households. As previously mentioned, the vast majority of existing papers study *observable* drivers of the MPC; we relate our findings to this literature. A burgeoning literature has turned its attention to unobserved household traits and preference heterogeneity. Our findings corroborate the importance of this dimension,

[7] This is true unless preference heterogeneity is explicitly elicited in survey questions so that it can be used as an observable control. Using Nielsen panel data, Parker (2017) finds that the MPC out of the tax rebate is indeed strongly correlated with a self-reported measure of impatience.

recently highlighted by Alan, Browning, and Ejrnaes (2018), Parker (2017), Aguiar et al. (2019), and Gelman (2019).

Our approach allows us to flexibly and non-parametrically combine observed and unobserved MPC heterogeneity. In this respect, Misra and Surico (2014) is closest in spirit to our work.[8] They estimate a quantile regression of consumption responses to the 2008 tax rebate using the same data, and find substantial heterogeneity. However, quantile regression estimates the role of regressors at specific points in the overall conditional distribution of the dependent variable. We discuss how this approach is sensitive to the correlation of MPC heterogeneity with other forms of heterogeneity, since other factors may be quantitatively larger drivers of the conditional distribution of consumption changes than the tax rebate. Indeed, we find in simulations that quantile regression overstates the dispersion of MPCs, particularly in the left tail. We not only estimate a different MPC distribution relative to Misra and Surico (2014) (in particular, one featuring a lower bound on MPCs that is considerably higher), but we also find new relationships with observables.

Finally, we believe our findings can be used to discipline heterogeneous agent models, for three main reasons. First, our estimated full distribution of MPCs is an agnostic target, regardless of a model's characteristics. Second, individual correlations between MPCs and observable characteristics are crucial objects for many heterogeneous agent models, and we are able to estimate them with statistical precision. Third, we provide researchers with an explicit number for the joint importance of observable and unobservable drivers for the distribution of MPCs.

The paper proceeds as follows. In Section 2 we describe our empirical strategy based on the 2008 tax rebate. In Section 3, we formulate the problem at hand and present the FCP estimator. We extend FCM from the cluster means case to a fully-general regression setting, as well as instrumental variables regression, and derive asymptotic properties of the corresponding estimators. Our results are outlined in Section 4, where we provide estimates of the distribution of MPCs for various consumption categories. Section 4.3 discusses observable characteristics that are correlated with the estimated MPCs. Section 4.5 aggregates the estimated household MPCs to arrive at a partial equilibrium effect on aggregate consumption. Section 5 concludes.

[8] Other papers have used the "reported preference" approach, eliciting MPC heterogeneity directly from responses to survey questions. Recent examples include Sahm, Shapiro, and Slemrod (2010), Jappelli and Pistaferri (2014) and Fuster, Kaplan, and Zafar (2018).

2 Empirical methodology

In order to estimate the marginal propensity to consume, and how it varies across households, we look at an off-the-shelf well-identified quasi-natural experiment: the 2008 Economic Stimulus Act (ESA), as studied by Parker et al. (2013), among others. Between April and July of 2008, $100 billion in tax rebates was sent to approximately 130 million US tax filers.[9] The timing of rebate receipt was determined by the last two digits of the recipient's Social Security Number (SSN), making the timing of receipt random. As in Parker et al. (2013), we also exploit the randomized timing of the rebate receipt, but instead estimate heterogeneous (and unobserved) marginal propensities to consume rather than a homogeneous marginal propensity to consume. Our data come from the Consumer Expenditure Survey (CEX), which contains comprehensive and detailed measures of household-level consumption expenditures. The 2008 CEX wave also includes supplemental questions on the ESA, including the amount of each stimulus payment received. While CEX expenditures are reported at the quarterly frequency, new households enter the survey at each month, making the frequency of our data monthly. Since we depart from Parker et al. (2013) by allowing for treatment heterogeneity, we present their homogeneous specification first as a useful benchmark, introducing our generalizations thereafter.

2.1 Homogeneous MPC

Parker et al. (2013) consider the following specification:

$$\Delta C_j = \beta' W_j + \phi R_j + \alpha + \epsilon_j, j = 1, \ldots, N, \tag{1}$$

where ΔC_j is the first difference of consumption expenditure of household i in quarter t.[10] W_j is a set of controls including month dummies aimed at absorbing common time effects such as aggregate shocks, as well as seasonal factors.[11] The independent variable of interest is R_j, which denotes the amount of the tax rebate received by each household. ϕ

[9] We defer to Parker et al. (2013) and Sahm et al. (2010) for an exhaustive discussion of the Economic Stimulus Act.

[10] To maintain consistent notation throughout the paper, we refer to j as the (i, t) combination of household i in quarter t. We wish to emphasize that while we have information on the same households i in different periods t, identification is not obtained by comparing individual responses over time. We do not exploit any limited panel structure, except to construct consumption changes for the left-hand-side variable. We return to this point below.

[11] In Parker et al. (2013), the other controls are age, change in number of adults in the household, and change in the number of children in the household. The controls we will use are the same, but additionally include age squared.

is then interpreted as the causal effect of the rebate on expenditures, where identification is achieved by comparing expenditure changes of households that received the rebate in a certain period to expenditure changes of households that did not receive the rebate in the same period.[12]

2.2 Heterogeneous MPCs

We depart from the homogeneous specification in Equation (1) and allow for heterogeneity in the expenditure responses to the tax rebate across households. In particular, we augment Parker et al. (2013)'s specification as follows:

$$\Delta C_j = \beta' W_j + \sum_{g \in G} \left(\phi_g \mathbf{1}\left[j \in g\right] R_j + \alpha_g \mathbf{1}\left[j \in g\right] \right) + \epsilon_j, j = 1, \ldots, N, \tag{2}$$

where $\mathbf{1}\left[j \in g\right]$ is an indicator that takes a value of 1 if household i in period t belongs to a certain group $g = 1, \ldots, G$. That is, we assume that heterogeneity in responses to the rebate can be summarized with G groups, characterized by the vector of coefficients $\{\alpha_g, \phi_g\}$. We include group-specific intercepts α_g to correctly interpret ϕ_g as a marginal propensity to consume. For example, since we do not observe quarterly changes in income, failing to include group-specific level effects may bias MPC estimates due to heterogeneity in income changes unrelated to the tax rebate. Our object of interest is $\phi = \begin{pmatrix} \phi_1 & \ldots & \phi_G \end{pmatrix}'$, which describes MPC heterogeneity, while $\mathbf{1}\left[j \in g\right]$ tells us the group membership of each household. The vector of coefficients, combined with $1\left[j \in g\right]$, gives an approximation of the MPC distribution. In the next section we introduce a new methodology to jointly estimate ϕ and $\mathbf{1}\left[j \in g\right]$. We also discuss how we choose the total number of groups, G.

3 A fuzzy clustering approach to MPC estimation

In this section, we describe our approach to recovering the distribution of MPCs. We first outline the FCM algorithm for the cluster means case, before generalizing it to our FCP methodology, and outlining simulation results. Finally, we compare the FCP approach to alternative methodologies.

[12] Kaplan and Violante (2014) discuss why ϕ may not correctly measure the marginal propensity to consume out of a transitory income shock, but is instead better thought of as a "rebate coefficient". We address these issues in Supplement D.5.

3.1 Fuzzy clustering

To estimate group-specific MPCs we must first assign individuals to groups. Previous papers have grouped individuals based on observable characteristics, but doing so presupposes the determinants of the MPCs *a priori*. However, we do not suppose to know these determinants in advance (there is indeed considerable empirical and theoretical disagreement on this point).[13] Moreover, we aim to investigate correlates of the MPC *ex post*, requiring us to remain agnostic while recovering the MPCs. For these reasons, we group individuals on the basis of their heterogeneous MPCs themselves (and potentially other group-specific parameters). Clustering methods are tailored to this goal.

To build intuition for our approach, we start with the cluster means case, where the econometrician simply wishes to characterize the distribution of means across groups. This problem consists of minimizing

$$L_1(P, \psi) = \int \sum_{g=1}^{G} w_g(y; \psi) \|y - \psi_g\|^2 P(dy), \tag{3}$$

where $y \in \mathbb{R}$ is the observed variable with probability measure P on $\mathbb{R}$, $g = 1, \ldots, G$ indexes groups, and $\psi \in \mathbb{R}^G$ is the vector of cluster means.[14] The weights, w_g, satisfy $\sum_{g=1}^{G} w_g(y; \psi) = 1 \forall y$. Equation (3) constitutes a weighted least-squares (WLS) problem, where the weights are unknown. Dunn (1973) shows that if the econometrician does not have any additional information about the distribution of y, the optimal solution to Equation (3) consists of binary weights $w_g^* = \mathbf{1}\left[\|y - \psi_g\|^2 \leq \|y - \psi_h\|^2 \ \forall\, h \neq g\right]$. This "hard" classification corresponds to the so-called hard K-means (HKM) approach. If the econometrician has information to provide a likelihood for y, then w_g can be derived from that likelihood, as in a Gaussian mixture model, for instance. However, performance can deteriorate (e.g., Winkler et al. (2011)) if the likelihood is misspecified or there are outliers, which motivates the use of the (optimal) non-parametric weights, w_g^*.

Moreover, the binary nature of w_g^* is not well-suited to empirical settings like ours. Some clustering applications focus on cases where y is a $T \times 1$ vector of outcomes (a panel structure), sometimes even assuming $T \to \infty$. In contrast, in our data, $T = 1$. Given a single observation for each individual, in the presence of noise – including measurement error, which is common in survey data like the CEX – definitive group assignment is an unrealistic representation of the econometrician's information. Assigning individuals for

[13] We cover the extensive literature on MPC heterogeneity in Sections 1 and 4.3.

[14] For simplicity, we present our theory in the text in terms of scalar-valued y, but our theoretical results in Supplement A are presented in full generality for vector-valued $y \in \mathbb{R}^T$ to accommodate a panel structure, for instance.

whom $\|y - \psi_g\|^2 \approx \|y - \psi_h\|^2, g \neq h$ discretely to a single group has the potential to mischaracterize group membership and to distort WLS estimates. Indeed, in a simple example with two Gaussian clusters, the cluster means are shifted outwards, as we show analytically in Proposition 1 in Supplement A.1. The "fuzzy clustering" literature recognizes that group assignments may be better represented in practice as uncertain, with observations belonging to multiple groups with some weight in $(0, 1)$. To obtain a set of weights that is both optimal and non-binary, we must modify Equation (3) so that the derivative of the objective function with respect to w_g is positive at $w_g = 1$ for all g (see Rousseeuw et al. (1995)). Dunn (1973) and Bezdek (1973) propose a simple solution:

$$L_m(P, \rho) = \int \sum_{g=1}^{G} w_g^m(y; \rho) \|y - \rho_g\|^2 P(dy), \tag{4}$$

where m is a "fuzziness parameter" that determines the deviation from binary assignment. If $m = 1$, then $L_m = L_1$, with binary weights, but if $m > 1$, the optimal weights are "fuzzy", with $w^*_{g,m} \to 1/G$ as $m \to \infty$. In particular, Bezdek (1973) shows that the minimization of Equation (4) results in the optimal weights:

$$\mu_g(y; \rho) \equiv w^*_{g,m}(y; \rho) = \left(\sum_{h=1}^{G} \frac{\|y - \rho_g\|^{2/(m-1)}}{\|y - \rho_h\|^{2/(m-1)}} \right)^{-1}, g = 1, \ldots, G.$$

The resulting objective function,

$$J_m(P, \rho) = \int \sum_{g=1}^{G} \mu_g^m(y; \rho) \|y - \rho_g\|^2 P(dy) \tag{5}$$

is known as "fuzzy C-means" (FCM). Without specifying the distribution of y, and given the formulation in Equation (4), FCM provides the optimal non-parametric solution to clustering data in cross-sectional or small-T settings like the one we face empirically (Bezdek (1973)).

Asymptotic properties for FCM in this simple cluster means case are derived by Yang and Yu (1992) and Yang (1994), who show that sample estimates of ρ_g are consistent for their population counterparts with normal limiting distributions. FCM has frequently been found to perform very well in practice, sometimes even outperforming well-specified mixture models (e.g., Chapter 6 of Bezdek (1981)).

3.2 Fuzzy clustering regression

While FCM has been widely adopted in practice as a tool to recover cluster means, the MPC, which we wish to uncover, is a regression coefficient on the tax rebate as outlined in Section 2. We therefore extend the FCM approach to a regression setting by considering the model

$$y_i = \sum_{g=1}^{G} \mathbf{1}\left[i \in g\right] \theta_g' x_i + \varepsilon_i, \quad i = 1, \ldots, N, \tag{6}$$

where group membership indicators $1\left[i \in g\right]$ are unobserved and x is a $K \times 1$ vector. We propose fuzzy C-parameters (FCP) as a novel tool to recover the distribution of heterogeneous treatment effects θ_g in Equation (6). FCP replaces cluster means ρ_g in Equation (5) with the conditional mean, $\theta_g' x$:

$$J_m^{reg}\left(\Pi, \theta\right) = \int \int \sum_{g=1}^{G} \mu_g^{reg}\left(y, x; \theta\right)^m \left\| y - \theta_g' x \right\|^2 \Pi_{y|x}\left(dy \mid x\right) \Pi_x\left(dx\right). \tag{7}$$

where $\theta \in \Theta \subset \mathbb{R}^{K \times G}$, Π denotes the distribution of (y, x) and

$$\mu_g^{reg}\left(y \mid x; \theta\right) \equiv \left(\sum_{h=1}^{G} \frac{\left\| y - \theta_g' x \right\|^{2/(m-1)}}{\left\| y - \theta_h' x \right\|^{2/(m-1)}} \right)^{-1}, g = 1, \ldots, G. \tag{8}$$

As before, Equation (7) has the interpretation of a WLS regression, where the weights are unknown. We define θ^* as the population minimizer of the objective in Equation (7). Our approach determines group weights based on the data alone, without requiring a classification *a priori*.

Asymptotic properties of fuzzy C-parameters

We now characterize the asymptotic properties of an estimator, $\hat{\theta}$, based on the sample counterpart of Equation (7). In Supplement A.2, we provide detailed proofs of these properties, generalized to allow y to be vector valued (as in panel data).

Substituting μ_g^{reg} in Equation (8) into Equation (7) and taking first-order conditions

with respect to θ yields a set of moment conditions

$$E\left[\frac{\partial J_m^{reg}}{\partial vec\left(\theta\right)}\right] \equiv E\left[\eta\left(\theta, y_i, x_i\right)\right] = E\left[\left(\sum_{h=1}^{G} \frac{\left\|y_i - \theta_g' x_i\right\|^{2/(m-1)}}{\left\|y_i - \theta_h' x_i\right\|^{2/(m-1)}}\right)^{-m} \left(y_{it} - \theta_g' x_i\right) x_i\right] = 0$$
$$\text{for } g = 1, \ldots, G,$$

which are satisfied by θ^*. These sample counterparts of these moment conditions can be used to estimate θ^* by (just-identified) GMM using the objective function

$$S_N\left(\theta\right) = \frac{1}{N}\sum_{i=1}^{N} \eta\left(\theta, y_i, x_i\right)' \sum_{i=1}^{N} \eta\left(\theta, y_i, x_i\right). \tag{9}$$

Define $\hat{\theta}$ as the solution to (9). To conserve space, we detail all assumptions in Supplement (A.2), but sketch them here. Assumption 1 presents standard OLS assumptions, extended to require exogeneity of x_i for each group, as well as finite G. Assumption 2 adds standard additional assumptions for consistency of GMM, in particular the uniqueness of θ^* (up to group ordering) and compactness of Θ. Theorem 1 shows that $\hat{\theta}$ is consistent for θ^*, the population solution to Equation 7:

Theorem 1. *(Consistency) Under Assumptions 1-2, $\hat{\theta} \xrightarrow{p} \theta^*$ as $N \to \infty$.*

To characterize the asymptotic distribution, we require further standard technical conditions for the asymptotic normality of GMM estimators, outlined in Assumption 3. The limiting distribution of $\hat{\theta}$ is given by Theorem (2):

Theorem 2. *(Asymptotic Normality) Under Assumptions 1 - 3,*

$$\sqrt{N}\left(vec\left(\hat{\theta}\right) - vec\left(\theta^*\right)\right) \xrightarrow{d} \mathcal{N}\left(0, H^{-1}VH^{-1}\right),$$

where

$$V = E\left[\eta\left(\theta^*, y_i, x_i\right)\eta\left(\theta^*, y_i, x_i\right)'\right],$$

and H is the Hessian of Equation (7).

We provide expressions for H (including several extensions) in Supplement A.2. With these results, we are able to conduct inference for the population parameter θ^*.

Practical considerations

In practice, the econometrician must choose both the number of groups, G, and the smoothing parameter, m; we adopt data-dependent rules to select them. For G, we adapt the "gap statistic" of Tibshirani et al. (2001) to the regression setting. This approach runs the clustering algorithm on data with no cluster structure (homogeneous coefficients) and compares the FCP objective function in this reference distribution to the value obtained for the empirical data. The gap statistic normalizes declines in (7) as G increases by expected declines in (7) that would arise mechanically even without a group structure; the optimal G maximizes the gap between the reference and empirical FCP objective function.

We choose m to minimize the loss function

$$Q(m) = \left(\bar{J}_m^{reg}\left(\hat{\theta}\right)\right)^2 + \left(\bar{E}_m\left(\hat{\theta}\right)\right)^2,$$

where $\bar{J}_m^{reg}$ is the sample analog of Equation (7), normalized to the range $[0,1]$, and $\bar{E}_m$ is the sample cluster entropy (see e.g., Bezdek et al. (1984)), measuring dispersion with clusters, normalized to the range $[0,1]$. We trade off the objective function against the similarity of observations within each group. This approach is related to the method of Cui et al. (2010) for cluster means. In our data, we select $m = 2.35$, in line with the range of values found in the FCM literature.[15] Since the optimal value for G is invariant to m in our data, we choose G and m sequentially.

We implement FCP by numerically minimizing the GMM objective function, as detailed in Section (3.2). The FCM literature has solved the minimization problem iteratively, but we find GMM to be both considerably faster and more accurate in our setting than the iterative approach. As such, we view casting the problem in a GMM formulation as another contribution of the paper. Supplement C.1 contains details.

FCP instrumental variables regression

In our empirical setting, the value of the rebate an individual receives is potentially endogenous, so we propose an instrumental variables extension of FCP (FCP-IV). In particular, we consider a two-stage least squares (TSLS) estimator, where we denote the rebate as x^e, additional controls as ω, and the instrument as z (in our case, an indicator for the

[15] The literature generally suggests $1.5 \leq m \leq 2.5$ (e.g., Bezdek et al. (1984); Pal and Bezdek (1995); Yu et al. (2004); Jing et al. (2014); Wu (2012)). Optimality results are not generally available (theoretically or numerically).

rebate receipt). We estimate the first stage

$$x_i^e = \gamma z_i + \tau' \omega_i + u_i,$$

via OLS, and then generate $\tilde{x}_i^e = \gamma z_i + \tau' \omega_i$. We define $\tilde{x} = \begin{pmatrix} \tilde{x}^e & \omega' \end{pmatrix}'$, and estimate Equation (7), replacing x with $\tilde{x}$ as the second stage. In Supplement A.3, we establish the consistency and asymptotic normality of this estimator.

3.3 Simulation results

To evaluate the performance of FCP in our setting, we conduct a simulation study calibrated to our empirical data. We summarize the main results below, with full details reported in Supplement C.2. We simulate data according to Equation (6) using the empirical parameter estimates, taking an individual's true group membership as the group with maximal weight. We consider three specifications: $G = 5$ (our empirical baseline), with Gaussian errors ε_i (with group-specific variance); $G = 5$, with ε_i resampled from each group's empirical errors; and $G = 10$, with Gaussian errors (with group-specific variance). For each specification, we estimate the model using FCP ($m = 2.35$, our empirical choice), HKM extended to a regression framework (as in Bonhomme and Manresa (2015)), and a Gaussian mixture model (using the EM algorithm). For each estimator, we report the mean and RMSE for each MPC, the share of observations misclassified (by maximal weight), and the median CDF of MPCs (by maximal weight), across 500 simulations.

For $G = 5$ and Gaussian errors, both FCP and HKM perform quite well by all metrics: the average point estimates are accurate, the RMSEs are fairly low, only 9% of individuals are misclassified, and the median CDF recovers the true distribution of MPCs very closely. The Gaussian mixture model performs poorly, recovering a spuriously wide distribution of MPCs. For $G = 5$ and empirical errors, FCP continues to perform well, demonstrating a clear advantage over HKM in terms of average point estimates and continuing to match the true CDF of MPCs. The Gaussian mixture model deteriorates further, recovering a large mass of negative MPCs. Finally, for $G = 10$, a much more challenging clustering problem, FCP performs remarkably well. Average point estimates remain reasonably close to the truth, the RMSEs generally fall, the share misclassified rises only slightly, and the median CDF matches the true distribution of MPCs very closely. HKM recovers a large mass of negative MPCs, while the Gaussian mixture model recovers many MPCs greater than one. Additionally, for FCP, tests of the null of the true MPCs are generally well-sized. Together, these results support the ability of FCP to recover the distribution of MPCs in our empirical setting.

Apart from these clustering approaches, quantile regression is used by Misra and Surico (2014) to characterize heterogeneous responses to the 2008 tax rebate. Quantile regression differs from clustering; rather than recovering groupings of the data that share similar parameters and estimating those parameters, quantile regressions investigate the relationship between the dependent variable and the regressors at different points of the conditional distribution of the dependent variable. Because quantile regression computes relationships at percentiles of the overall conditional distribution, the estimated MPC distribution depends on the correlation of MPCs with other forms of heterogeneity. If the "ranking" of the conditional distribution is mostly driven by factors other than the responsiveness to the rebate (like fixed effects or other covariates), and these factors are uncorrelated with the rebate, heterogeneity of the MPC distribution will be underestimated in the presence of noise. We provide a simple example in Supplement C.3.

We confirm quantile regression may provide a poor approximation of the MPC distribution using the empirically-motivated simulation setup described above. We find that, in this setting, quantile regression estimates spurious dispersion in the MPC, particularly in the left tail, computing a sizable mass of responses below the lowest MPC and even below zero. As we argue below, the lower tail of the MPC distribution is indeed a crucial distinction between our estimates and those of Misra and Surico (2014). Quantile regression also less accurately assigns individuals to MPCs in our simulations. This latter point is important since we examine determinants of MPC heterogeneity *ex post*; if individual MPCs are not accurate, results of these regressions will be spurious. Such exercises also pose a conceptual problem for quantile regression. Following quantile regression, it is possible to regress *percentiles* on the characteristics of their members, similar to how Misra and Surico (2014) plot observables across percentiles. However, it is not possible to examine the relationship between MPCs and covariates directly. Thus, this exercise relies on percentiles and MPCs being well-correlated for validity.

4 Results

We apply our FCP approach to the 2008 tax rebate. In particular, our heterogeneous formulation in Equation (2) is a special case of Equation (6), where the coefficients on W_j are assumed to be common across groups; we then estimate (6) using FCP with $m = 2.35$, in particular minimizing

$$\frac{1}{N}\sum_{j=1}^{N}\left(\sum_{g=1}^{G}\left\|\Delta C_j - \alpha_g - \phi_g R_j - \beta' W_j\right\|^{-2/(m-1)}\right)^{1-m}$$

with respect to $\{\alpha_g, \phi_g\}_{g=1}^{G}$.[16,17] Our findings highlight a considerable degree of MPC heterogeneity whose extent varies depending on the consumption category considered. We first show the distribution of marginal propensities to consume for total expenditures and illustrate how our results are robust to different specifications and sample selection procedures. We then investigate how the MPC distribution changes as we consider non-durable and durable goods as the dependent variables. Importantly, our approach also allows us to directly test whether households display similar propensities for different consumption goods, or instead substitute across expenditure types when they receive a transitory income shock such as a tax rebate. Finally, we explore which observable household characteristics are correlated with the estimated marginal propensities to consume, both individually and jointly, and discuss the partial equilibrium response of expenditures to the 2008 tax rebates in our heterogeneous setting relative to the homogeneous one.

4.1 The distribution of marginal propensities to consume

We start by considering total expenditures, defined as in Parker et al. (2013). Following Kaplan and Violante (2014), who show that properly accounting for outliers reduces the homogeneous rebate coefficient, while increasing precision, we drop the top and bottom 1.5% of consumption changes.[18] The gap statistic indicates that 5 is the optimal number of groups. For each household that receives the rebate, we compute the weighted average MPC, using the household-specific weights $\mu_{i,g}$ and the group-specific MPCs ϕ_g estimated by the algorithm.[19] Figure 1 shows the distribution of this object for the (i,t) pairs receiving the rebate.

The vast majority of households display a relatively low (but certainly non-negligible) MPC between 0.25 and 0.4, and the share of households with a given MPC slowly decays as the MPC increases. While under this specification no household can be strictly defined as having an MPC of one, the majority of the sample exhibits a sizable propensity to consume; our findings suggest that most households consume at least part of the rebate.[20]

[16] The weights are given by Equation 8, which are a function of the estimates themselves.

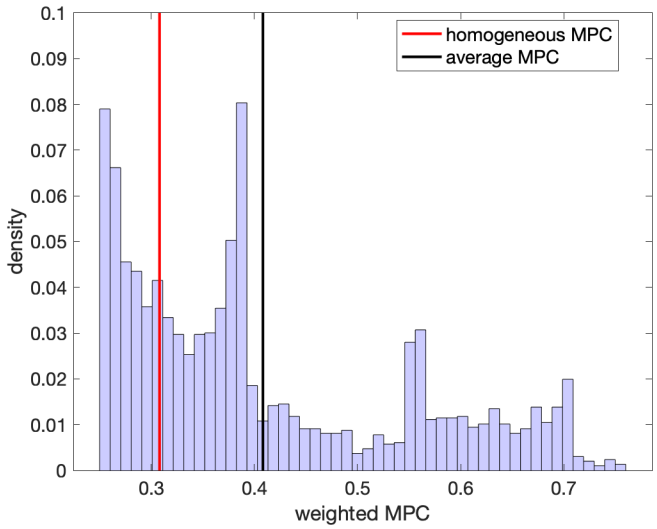
[17] We show in Appendix D.1 that the estimated MPC distribution is robust to also allowing for heterogeneity in the household level controls.

[18] This is the only way in which our sample departs from Parker et al. (2013), and explains why the homogeneous MPC we estimate for total consumption differs from theirs.

[19] Supplement D.3 shows the distribution of the MPC associated with the maximal weight at the individual level. Because we find that many such weights are close to one, the average of MPCs associated with maximal weights is very close to the average of weighted MPCs.

[20] In Section D.2, we report the MPC distribution estimated by HKM and a Gaussian mixture model. The distribution is qualitatively robust to the estimator used, although there is variation in the degree of dis-

Figure 1: Estimated distribution of MPCs out of the tax rebate



Notes: Figure 1 plots a histogram (light blue bars) of the estimated distribution of MPCs for total expenditures among households that received the rebate, defined as in Parker et al. (2013). The homogeneous MPC (red vertical line) is estimated assuming a homogeneous response to the tax rebate as in Parker et al. (2013), following Equation (1). For each household we compute the weighted MPC across groups. The black vertical line shows the average weighted MPC in our sample.

Fagereng et al. (2016) and Olafsson and Pagel (2018) find sizable spending responses even for households with high liquid wealth, in Norwegian administrative data and Icelandic app user data, respectively. We confirm that even the smallest MPCs are substantially larger than zero, even when estimating the full unconditional MPC distribution, in standard U.S. survey data. In contrast, in this same data, Misra and Surico (2014) use quantile regression to estimate a substantial share of MPCs at or *below* zero. As we discuss in Section 3.3 and Appendix C.3, our simulations suggest that an extended left tail could be a spurious feature of the quantile regression approach in this setting, which would explain the difference in our recovered MPC distributions. A lower bound of the MPC distribution above zero can be explained by bounded rationality. Ilut and Valchev (2020), for instance, develop a model in which MPCs can be high for all households, even those with slack liquidity constraints. Due to limited cognitive perception, households can find themselves in the midst of a "learning trap", "which makes the high MPC behavior the norm, rather than exception."

Aggregating the individual-level responses, we also find a larger average marginal propensity to consume than from the homogeneous regression, as shown by the black

persion in the tails. HKM does not estimate any MPC higher than 0.45. The Gaussian mixture model, in contrast, predicts higher dispersion on both ends of the distribution, also estimating some negative MPCs; we find exaggerated dispersion to be a property of the Gaussian mixture model in simulations, see Supplement C.2.

Table 1: Test for MPC equality

(a) Analytical standard errors

	MPC				
	0.25	0.38	0.57	0.71	0.76
0.25	6.02				
	(0.01)				
0.38	0.37	1.87			
	(0.54)	(0.17)			
0.57	6.02	0.37	0.17		
	(0.68)	(0.82)	(0.48)		
0.71	8.05	1.63	0.04	10.51	
	(0.00)	(0.20)	(0.84)	(0.00)	
0.76	0.21	0.14	0.02	0.00	0.43
	(0.65)	(0.71)	(0.89)	(0.97)	(0.51)

(b) Conditional on FCP weights

	MPC				
	0.25	0.38	0.57	0.71	0.76
0.25	154.55				
	(0.00)				
0.38	8.28	78.70			
	(0.00)	(0.00)			
0.57	2.23	0.75	7.33		
	(0.14)	(0.39)	(0.01)		
0.71	115.12	32.48	0.45	304.07	
	(0.00)	(0.00)	(0.50)	(0.00)	
0.76	4.37	2.35	0.35	0.04	9.89
	(0.04)	(0.13)	(0.55)	(0.85)	(0.00)

Notes: MPCs for total expenditures. The two tables show $F-$statistics from pairwise two-sided Wald tests of equality across MPCs (the diagonals show tests of equality with zero). The left panel uses the standard errors outlined in Theorem 2. The right panel repeats the exercise, taking the weights as given. These are equivalent to weighted least squares estimates where the weights are those in Equation (8), raised to the power m, as in (12). Therefore, to run the tests in Table 1(b), we replicate the sample by the number of groups and estimate $\Delta C_j = \beta' W_j + \sum_{g \in G} \left(\phi_g \mathbf{1}\left[j \in g\right] R_j + \alpha_g \mathbf{1}\left[j \in g\right] \right) + \epsilon_j$ via weighted least squares, with standard errors corrected for heteroskedasticity, and compute the Wald tests. $p-$values are reported in parentheses.

and red vertical lines, respectively. This is not surprising; in general, estimates from a homogeneous specification like Equation (1) will not equal the average of estimates from our heterogeneous specification, unless the distribution of right-hand-side variables is invariant across groups. We discuss this point further in Supplement E, and provide a statistical decomposition to determine which variables drive this discrepancy.[21] We find that variation in the mean of regressors across groups (particularly for certain time dummies and the age variable) is principally responsible.

In Table 1 we examine whether the estimated MPCs differ from zero and are statistically different from one another. In the left panel, we make use of the analytical formulas outlined in Theorem 2 to compute Wald tests of pairwise equality across MPCs. Groups 3 and 5 (ordered from lowest to highest MPC), whose MPCs are not statistically significantly different from zero, are also those with the smallest share of households.[22] Relative to the equivalent weighted least squares regression in which the weights are treated as given, the FCP standard errors are larger because they take estimation error of the weights into account. Table 1(b) shows that – when group assignment is taken as given – all MPCs are statistically different from zero. Moreover, most MPC groups are statis-

[21] See Słoczyński (2020) for a recent discussion of this issue in a model with binary treatment.

[22] 4% and 3% of rebate recipients have maximum weight on groups 3 and 5, respectively.

tically different from each other, at least at the 68% confidence level. Supplement D.3 further shows that the full *distribution* of MPCs is largely invariant when re-estimated on bootstrap samples drawn from the data.[23]

The flexibility of the FCP methodology allows us to nest instrumental variable estimation. This is particularly relevant in our framework, since the exogenous source of the transitory income shock is driven by the random timing of rebate receipt, but the value of the rebate itself may be endogenous. We therefore follow the literature and instrument the tax rebate with an indicator variable for its receipt. In this TSLS specification, we first regress the rebate value on the rebate indicator and the same controls as in Equation (2), and then use the predicted values in the second stage. Moreover, we also address potential bias arising from the inclusion of households that never get the rebate, excluding them from the analysis.[24]

Figure 2 plots the resulting distribution of weighted MPCs, and shows how it remains qualitatively unchanged relative to the OLS specification. If anything, instrumentation uncovers a non-negligible portion of households that consume the rebate in its entirety and that even display an MPC slightly larger than 1. Moreover, the gap between aggregated and homogeneous response is even larger than in OLS.

4.2 The MPC distribution for different consumption goods

We have shown how households differ with respect to their propensity to consume the rebate. How does the distribution of these propensities change across consumption goods? The granularity of the CEX data allows us to tackle this question, while our approach allows us to explore how good-specific MPCs vary at the household level.

First, in the left panel of Figure 3, we report the weighted MPC distribution for nondurable goods.[25] As expected, the distribution is shifted to the left with respect to the distribution corresponding to total expenditures in Figure 1, as nondurable goods account for, on average, only 57% of household total expenditures.

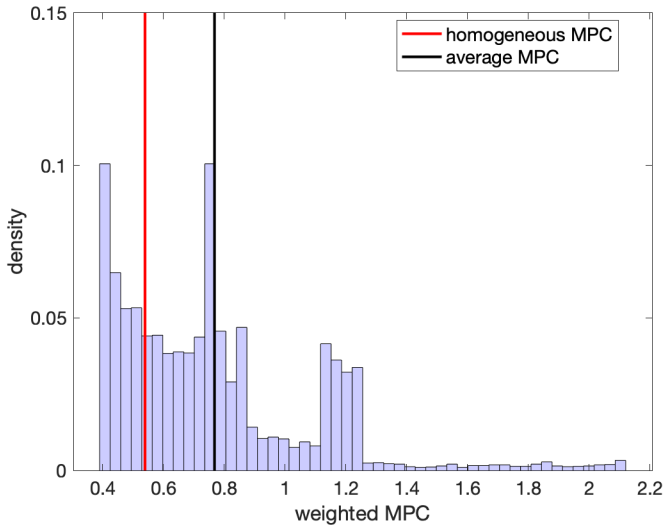
An important share of households consumes a value of nondurables consistent with the annuity value of the rebate, as suggested by the Permanent Income Hypothesis (Fried-

[23] In particular, we repeat the estimation of the distribution of MPCs for total expenditures, with 5 groups, over 250 samples obtained with bootstrap with replacement. We find that the average quantiles across bootstraps are very close to those estimated in the baseline sample, and fairly stable across bootstraps.

[24] Households may never receive the rebate because they have different characteristics, such as higher income. In Supplement D.5 we show robustness to the exclusion of this group for OLS too.

[25] Nondurable goods are defined, following Parker et al. (2013), as strictly nondurables (Lusardi (1996)) plus apparel goods and services, health care expenditures (excluding payments by employers or insurers), and reading material (excluding education).

Figure 2: Estimated distribution of MPCs out of the tax rebate: two-stage least squares



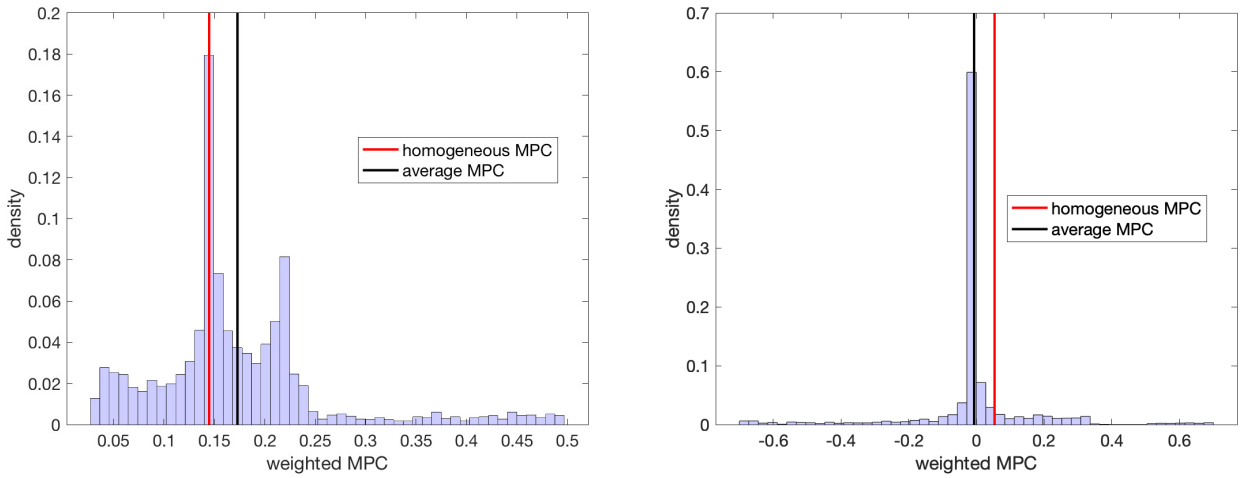
Notes: Figure 2 plots a histogram (light blue bars) of the estimated distribution of MPCs for total expenditures, defined as in Parker et al. (2013), using the two-stage least squares specification. We also drop from the sample the households that never receive a rebate. The homogeneous MPC (red vertical line) is estimated assuming a homogeneous response to the tax rebate, as in Parker et al. (2013), following the two-stage least squares equivalent of Equation (1). We estimate the model with $G = 5$. For each household we compute the weighted MPC across groups. The black vertical line shows the average weighted MPC in our sample.

man (1957)): roughly one third of households have an MPC that is not statistically distinguishable from zero. This suggests that the large lower bound on MPCs, recently found in the literature and discussed in the previous section, could depend on the type of good purchased. The majority of households have a small, but non-zero, propensity to consume nondurable goods, and around 8% of households consume nearly half of their rebate in nondurables. The heterogeneity in nondurable MPCs is not only economically meaningful, but also statistically significant. In Supplement D.3 we show that nearly all the estimated MPCs are statistically different from each other. Instrumenting the rebate with the rebate receipt indicator slightly increases both the mass and the values at the right tail, similar to the results for total expenditures.

As shown in the right panel of Figure 3, we estimate that 73% of households do not change their durable expenditures in response to the rebate; their weighted MPC is in a neighborhood of zero and the associated group-specific MPC is not statistically different from zero. A small fraction of households, however, has a durable MPC larger than one half.[26] The dichotomy of this MPC distribution is in line with the discrete nature of durable goods purchases. This discreteness implies lumpy adjustment and is consistent with the fact that most households either use most of the rebate to purchase durables, or

[26] In line with the tendency shown for other consumption categories, estimating durable MPCs with TSLS uncovers a group with larger propensity, up to 1.49.

Figure 3: MPCs out of the tax rebate: nondurables and durables



(a) Nondurables (b) Durables

Notes: Nondurable goods are defined, following Parker et al. (2013), as strictly nondurables (Lusardi (1996)) plus apparel goods and services, health care expenditures (excluding payments by employers or insurers), and reading material (excluding education). The homogeneous MPC (red line) is estimated assuming homogeneous response to the tax rebate. For each household we compute the weighted MPC across groups. The black line shows the average weighted MPC in our sample. We estimate the models with $G = 5$ to allow direct comparability with the distribution of MPCs for total expenditures. We follow Coibion et al. (2017) and define durables as durable health expenditures, entertainment durables, furniture, jewelry, durable personal care, vehicle purchases, durable vehicle expenditures, housing durable expenditures (e.g., maintenance and repair commodities such as paint, materials.).

do not adjust their durable goods consumption at all.[27]

Finally, we assess whether households with high propensities to consume nondurable goods are also more likely to consume durable goods after receiving the rebate. While we can rule out substitution between goods, the estimated complementarity – at the margin – is, however, quantitatively small. The correlation between household-level weighted MPCs for nondurable goods with those for durables is 0.04, significant at the 5% level, while the rank correlation is also 0.04. Albeit small, the complementarity might signal the presence of heterogeneous preferences or a small share of "spender" types, who are more prone to adjust any type of consumption in response to transitory income shocks. While the structure of our data does not allow us to draw conclusions regarding permanent unobserved heterogeneity in MPCs, we can investigate what observable characteristics explain the estimated MPC distributions that we recover. We tackle this issue in the next section.

[27]Our definition of durables follows Coibion et al. (2017). Parker et al. (2013), instead, define durable expenditures as the difference between total and nondurable expenditures. Using this categorization delivers very similar results.

4.3 What drives MPC heterogeneity?

Our approach uncovers the distribution of marginal propensities to consume without taking a stance, *ex ante*, on its observable determinants. Consequently, we can use the estimated distribution to understand how MPCs correlate, *ex post*, with observable characteristics. We start by examining how observables are individually correlated with MPCs. We then turn to investigate the joint relationship between the estimated MPCs and various household characteristics. As such, we contribute to the literature in two ways. First, we show that, with our approach, a large number of statistically significant individual correlations between MPCs and observable drivers emerge. This is true despite the fact that we use a dataset and an identification strategy which previously failed to find statistically significant relationships (see Parker et al. (2013)). We believe that this contrast underscores the power of our approach. Second, we show how the distribution of MPCs is jointly correlated with observable characteristics, and can be confident that any lack of significant correlations is not due to loss of statistical power introduced by progressive interactions.

Table 2 reports individual correlations. Our estimated weighted MPCs (column (1)) are positively correlated with salary and non-salary income, the mortgage interest-to-income ratio, the average propensity to consume (APC), and log liquid wealth; however, they are negatively correlated with age.[28] As shown in columns 2-4, these findings barely change when considering the MPC associated with an individual's maximal weight, or the MPC distribution estimated via TSLS.[29] We also find that homeowners have larger MPCs, a result that echoes findings in Parker et al. (2013). Moreover, having a mortgage is associated with an even higher propensity to consume, as shown in Figure 4.

Table 3 reports multivariate regressions of the estimated MPCs on the same observables. Since the inclusion of liquid assets reduces the number of observations, we also estimate the MPC drivers without taking liquidity into account. The absence of a correlation between the MPC and liquid assets is in line with findings by Parker et al. (2013), who suggest it might be driven by non-response bias.[30] Our results are robust across specifications and to using alternative MPC definitions.[31]

[28] Additional relationships hold unconditionally. For instance, we find that households that put money into a tax-deferred or tax-free educational savings plan have a significantly higher MPC. Moreover, MPCs increase with education. All these relationships, however, are insignificant when tested jointly with other observables as in Table 3.

[29] In unreported results we also find that correlations with observables are robust to the exclusion of observations associated with statistically insignificant MPCs for total expenditures.

[30] Considering the ratio of liquid assets to total income leaves the coefficients unaffected. Moreover, this variable is unconditionally uncorrelated with the marginal propensities to consume.

[31] All of the regression results are broadly unaffected if using the MPC associated with an individual's

Table 2: Individual correlations with the MPC for total expenditures

	(1)	(2)	(3)	(4)
log salary income	0.11***	0.09***	0.10***	0.09***
log non-salary income	0.19***	0.17***	0.18***	0.18***
mortgage interest/income	0.06***	0.07***	0.05**	0.04*
APC	0.05*	0.05**	0.02	0.00
age	-0.04**	-0.04*	-0.04**	-0.03*
log liquid assets	0.12***	0.11***	0.14***	0.13***

Notes: Table 2 shows the correlations between MPC estimates listed in columns and observables listed in rows. Column (1) uses the weighted MPC from the WLS specification, while column (2) uses the MPC associated with the maximal weight from the WLS specification. Columns (3) and (4) use the weighted and maximal weight MPCs, respectively, based on the TSLS specification, respectively. All logged variables take a value of log(0.001) when the raw value is 0 or negative. *, ** and *** denote significance of the correlation at 10, 5 and 1% levels, respectively.

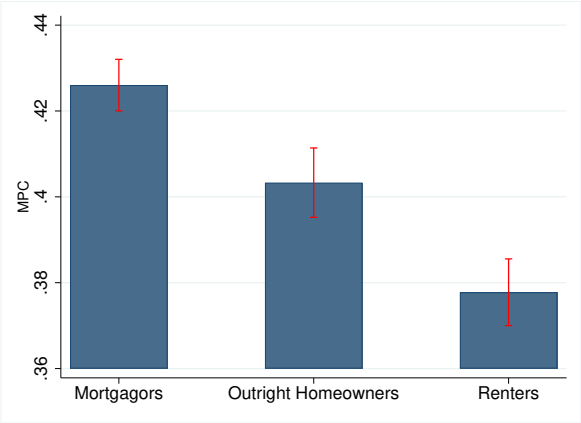
Importantly, only two explanatory variables remain statistically significant after the inclusion of additional covariates: non-salary income and the average propensity to consume, both of which are positively correlated with the marginal propensity to consume. We expand on these two drivers in the remainder of the section.

While higher income households have higher MPCs, it is mainly the non-salary component of income that drives this relationship.[32] This effect is partly the result of a particular category of households, such as entrepreneurs or investors (for example, those with a positive business or financial income), who have a significantly higher MPC. The intensive margin, however, seems to play the most prominent role. Putting the estimates together, we find that a 1% increase in non-salary income is associated with an increase in the MPC between 5 and 13 cents for each dollar of the rebate, depending on the specification. Put differently, a 2% increase in non-salary income predicts a 1 standard deviation increase in the MPC, in our preferred specification in column (2). The importance of business and financial income for the MPC might suggest the presence of wealthy hand-to-mouth households, as first posited by Kaplan and Violante (2014). Finally, the positive correlation between income and MPCs does not only hold for total expenditures, but also for nondurable expenditures, even when controlling for additional covariates.

maximal weight as the dependent variable.

[32]Income in the CEX is measured in the first interview and relates to income over the prior 12 months. Non-salary income consists of farm and business income, financial income (e.g., income from interest, dividends, pensions and annuities) and all other income except foodstamps (e.g., retirement, supplemental security, unemployment compensation), following the categorization in Coibion et al. (2017). Business and financial income are positively correlated with the MPC.

Figure 4: MPCs by housing status



Notes: Figure 4 shows the average weighted MPC for total expenditures by housing status of households receiving a rebate. Error bars show 90% confidence intervals, computed with the standard deviation of weighted MPC within each group.

Some studies find that low-income households have a higher marginal propensity to spend: see, for instance, Johnson et al. (2006) for the 2001 tax rebate and Jappelli and Pistaferri (2014), with respect to cash on hand, for Italian data on self-reported MPCs. Other studies, however, find mixed results or even the opposite relationship, as we do. While Broda and Parker (2014) find that low-income households had larger propensities to spend in the month of the 2008 rebate receipt than households in the top income tercile, this difference "becomes indistinguishable by the end of the quarter". Misra and Surico (2014) also find that median income is higher at the top of the conditional distribution of consumption changes, which they find to be associated with higher propensities to consume, although the overall relationship is U-shaped. We instead find a monotonic relationship with income, as in Kueng (2018), who studies consumption responses to regular and predetermined payments from the Alaska Permanent Fund. Boutros (2020) finds that households whose 2008 rebate was a smaller fraction of their income – typically higher-income households – had a higher MPC. He explains this finding with a model of bounded intertemporal rationality, in which the smaller the relative size of the payment, the more planning costs dominate the benefits of consumption smoothing. The theory of limited cognitive perception developed by Ilut and Valchev (2020) also delivers rich agents with high MPCs. Shapiro and Slemrod (2009) use data on self-reported propensities to spend the 2008 rebate and show that low-income individuals were more likely to pay off debt. They also find that 21% of households making more than $75,000 of total annual income reported to spend most of the rebate, compared to 18% for households with total income below $20,000. Miranda-Pinto, Murphy, Walsh, and Young (2020) develop a model that can rationalize these findings via time-varying consumption thresholds.

Table 3: Explanatory variables: weighted MPC for total expenditures

	(1)	(2)	(3)	(4)
dummy for no salary	-0.032	-0.125	-0.361	-0.379
	(0.108)	(0.097)	(0.282)	(0.235)
log salary income	0.001	-0.006	-0.016	-0.021
	(0.007)	(0.006)	(0.018)	(0.015)
log non-salary income	0.046***	0.066***	0.097***	0.125***
	(0.012)	(0.009)	(0.029)	(0.022)
mortgage interest/income	0.054	0.010	0.146	0.073
	(0.041)	(0.034)	(0.092)	(0.075)
APC	0.063***	0.092***	0.116***	0.146***
	(0.016)	(0.013)	(0.031)	(0.025)
homeowner dummy	0.032*	0.024	0.026	0.022
	(0.015)	(0.012)	(0.039)	(0.031)
dummy for mortgage	-0.025	-0.013	-0.042	-0.026
	(0.015)	(0.012)	(0.036)	(0.028)
log liquid assets	-0.001		0.004	
	(0.003)		(0.006)	
dummy for liquid assets <= 0	-0.004		0.054	
	(0.040)		(0.089)	
N	723	1079	723	1079
R^2	0.101	0.112	0.101	0.092

Notes: Columns (1) and (2) use the weighted MPCs from our WLS specification. Columns (3) and (4) use the weighted MPCs from our TSLS specification. All logged variables take a value of log(0.001) when the raw value is 0 or negative. Non-salary income is positive for all observations. Standard errors are robust to heteroskedasticity and reported in parentheses. We control for marriage dummies, education dummies, number of children, age and age squared; those coefficients are not reported. Age and its square are controls in our FCP estimation. While this does not pose an issue for the point estimates shown in this table, it could potentially affect inference. However, we repeat the same regressions, excluding age and and age squared, and find that the remaining coefficients are unaffected. , *, and *** denote significance at the 10, 5, and 1% levels respectively.

Marginal propensities to consume also increase with the average propensity to consume (APC). Empirically, we define the APC as the ratio between lagged consumption and lagged total income. We consider income as measured in the first interview for each household, which refers to the previous 12 months. We lag expenditures to avoid the possibility of a mechanical positive correlation with the MPC. To ensure stability of APCs, we average expenditures over all the available lagged quarters at the household-level, but the results are virtually unchanged if we only consider the first lag. Households that spent 1 percentage point more of their income before receiving the rebate spent 9 additional cents out of each rebate dollar. This effect is significant also for nondurable MPCs and conditional on a wide array of controls.[33]

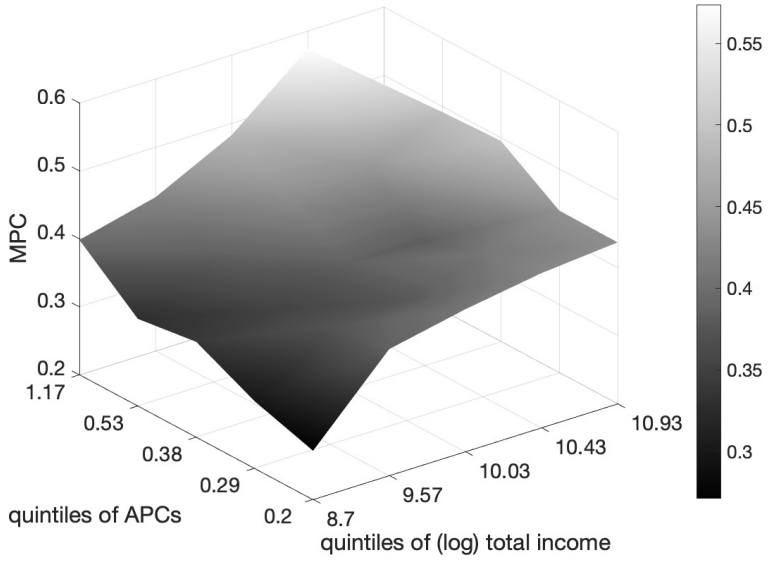
In Figure 5 we show how the MPC varies jointly with the APC and total income. We separately compute quintiles of the APC and total income, and calculate the average weighted MPC for each quintile pair. The MPC increases with income, conditional on the APC, and vice versa. As the figure shows, our analysis uncovers three main groups. First, households with low total income and a low APC display the lowest marginal propensity to consume. We label these households "poor savers". Second, households with a high APC and low total income, and vice versa, display intermediate MPCs. Third, the greatest marginal propensity to consume is found among households with a high APC and high total income. We label this group "rich spenders".[34]

We regard the results presented in this section as particularly relevant for disciplining macro models of household consumption. First, the average propensity to consume can easily be computed in a large number of micro datasets with minimal information. Second, these correlations can be directly tested in even the simplest of consumption/savings models. Existing models, however, predict very different relationships between MPC and APC. Hand-to-mouth, constrained agents, will typically have large MPC and APC. As they save towards their target level of wealth, both propensities fall. If agents are infinitely-lived, they eventually reach the target level of wealth, at which they stop saving (i.e.: APC = 1) and have an MPC equal to the annuity value of the transitory income shock. A life-cycle model, in contrast, can generate the empirically observed positive relationship between MPC and APC, as old households dissave, but also have a high MPC due to a low effective discount factor. This standard model, however, generates a clear

[33] A 1 percentage point increase in the APC for total expenditures predicts 2 additional cents per rebate dollar spent on nondurables. This effect goes up to 6 when considering the APC for nondurable expenditures only.

[34] We find similar relationships for MPC for nondurable goods, especially the presence of "rich spenders", as we show in Supplement (D.4).

Figure 5: The relationship between MPCs, APCs, and income



Notes: Figure 4 shows the average weighted MPC for total expenditures for pairs of quintiles of APC and log total income. The colorbar on the right represents the MPC.

relationship between the MPC and age, which our results do not bear out.[35]

All these characterizations are conditional on homogeneous preferences. Preference heterogeneity, in contrast, can break these relationships and rationalize some of our findings. Aguiar et al. (2019), for instance, highlight the importance of heterogeneity in the intertemporal elasticity of substitution in order to generate heterogeneous target levels of wealth; high-IES households have high MPCs and high APCs. Consistent with this, Parker (2017) finds that the majority of consumption responsiveness to the tax rebate in the Nielsen data is driven by a measure of impatience, defined as households reporting to be "the sort of people who would rather spend their money and enjoy today rather than save more for the future."

An additional finding underscores the importance of unobserved heterogeneity. All the observable drivers mentioned in this section — as well as other household characteristics that do not strongly correlate with the MPC — explain a relatively small portion of the variance of the weighted MPC distribution. Indeed, our linear regression framework of weighted MPCs on observable characteristics delivers an R^2 of 11%, at best. While other recent studies also find that most observable households characteristics are jointly

[35]Moreover, most incomplete markets models typically fail to generate savings rates (APCs) that increase (decrease) with wealth and permanent income, at odds with what is observed in the data and documented by Dynan, Skinner, and Zeldes (2004) and Straub (2017).

uncorrelated with MPC heterogeneity, we can exclude that this finding is solely due to loss of statistical power.[36] Moreover, through the R^2 we can formally obtain a statistical measure of the portion of the variance in the MPC distribution explained by observable characteristics.

A low R^2 could be partly explained by non-linear relationships that are either difficult to parametrize, or are not captured by variables in our dataset.[37] For example, the CEX contains only sparsely populated information on wealth. In Supplement D.4, we show the relationship between the MPC and liquid wealth, aware of the potential nonresponse bias highlighted by Parker et al. (2013), but we refrain from showing any relationship with total wealth, given the lack of reliable data. While such unmeasured characteristics could potentially explain some of the variation in MPCs, our results strongly suggest the presence of latent drivers, since some of those unobserved characteristics may give rise to the observables we analyze in the first place, such as the APC.

4.4 Robustness to spurious heterogeneity

In this section we show that neither the MPC heterogeneity we uncover, nor its correlation with observables, is a spurious product of our estimation approach. For this exercise, we generate data using estimates from the homogeneous regression, with errors drawn from a Gaussian distribution with the empirical variance. We then obtain FCP estimates under the faulty assumption that five groups are present, and repeat the same analysis for 250 Monte Carlo samples. Table 4 shows that imposing spurious heterogeneity on a homogeneous distribution does not significantly bias moments of the distribution.

Limited spurious heterogeneity, however, arises towards the tails, driven by particularly noisy draws in the simulation. To show that these spuriously estimated MPCs do not invalidate our headline results, we regress the estimated weighted MPCs for each sample on the array of observable predictors used in specification (2) of Table 3. On average across samples, all the estimated correlations are almost exactly 0. For illustrative purposes, Figure 6 displays the distribution of the $t-$statistic for the coefficient on the APC, across samples. In only 6.8% of the samples is there significant evidence of a relationship between MPC heterogeneity and APC at the 5% level, a size distortion within the scope of Monte Carlo error. The same is also true for coefficients on all other observables, with even lower shares of significant coefficients. Further, a common concern when

[36] See, for instance, Table 6 in Fagereng et al. (2016).

[37] As discussed, our results are robust to different sets of explanatory variables. We also run a linear Lasso for the selection of the array of predictors. Regressing the MPC on the selected right hand side delivers the same R^2.

Table 4: Over-fitting G: average quantiles of the MPC distribution across simulated samples

	Average	p25	p50	p75
truth:	0.31	0.31	0.31	0.31
$G = 1$	0.31 (0.11)	–	–	–
$G = 5$	0.30 (0.27)	0.25 (0.27)	0.30 (0.27)	0.35 (0.28)

Notes: Table 4 reports the average of various summary statistics of the distribution of (weighted) MPCs across Monte Carlo samples. The first row reports the truth, which is 0.31 for all statistics, since the distribution is homogeneous. The second row corresponds to a correctly-specified homogeneous regression in repeated samples (with the standard deviation across samples below in parentheses) and the fourth to FCP incorrectly assuming the presence of five distinct groups. pxx denotes the xx^{th} percentile. The final row reports the standard deviation of each moment for the $G = 5$ specification across simulated samples. Results are roughly unchanged computing medians across simulated samples, as well as distributions of the MPC associated with an individual's maximal weight.

using the CEX data is the role of measurement error. These exercises also serve to show that grouping on noise alone – like measurement error – does not lead to either the type of distribution that we recover or the correlations with observable characteristics that we estimate.

4.5 Aggregate partial equilibrium effects of the 2008 ESA

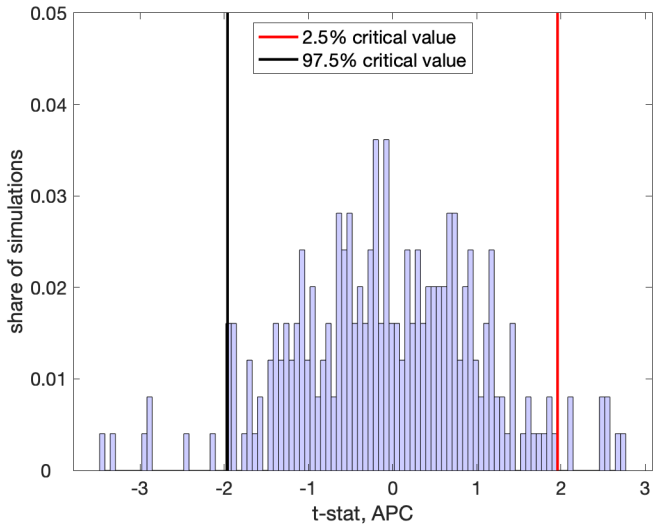
In this section, we estimate the partial equilibrium (PE) aggregate response to the 2008 tax rebate based on our estimated heterogeneous coefficients. For this exercise, we use a lagged specification that takes into account the possible persistent effects of rebate receipt, as in Parker et al. (2013). In particular, we estimate the following model:

$$\Delta C_j = \beta' W_j + \sum_{g \in G} \left(\phi_g \mathbf{1}\left[j \in g\right] R_j + \phi_g^{\text{lag}} \mathbf{1}\left[j \in g\right] R_j^{\text{lag}} + \alpha_g \mathbf{1}\left[j \in g\right] \right) + \epsilon_j, j = 1, \ldots, N, \tag{10}$$

where the coefficient ϕ_g^{lag} represents the lagged effect of the rebate for group g.[38] We do not force a household group membership for household i to be fixed across t, since we want to preserve flexibility. Even if individuals' preferences may be constant, the MPC may be time-varying, due, for instance, to changes in state variables such as income and wealth. To correctly estimate the cumulative response to the rebate, we therefore

[38] Kaplan and Violante (2014) suggest that the rebate coefficient might differ from the marginal propensity to consume because some households in the control group have already received the rebate, and some households might anticipate receiving the rebate in the future. Adding lagged rebate partially address this concern. See Supplement D.5 for further discussion of this specification.

Figure 6: (Lack of) spurious correlations with observables



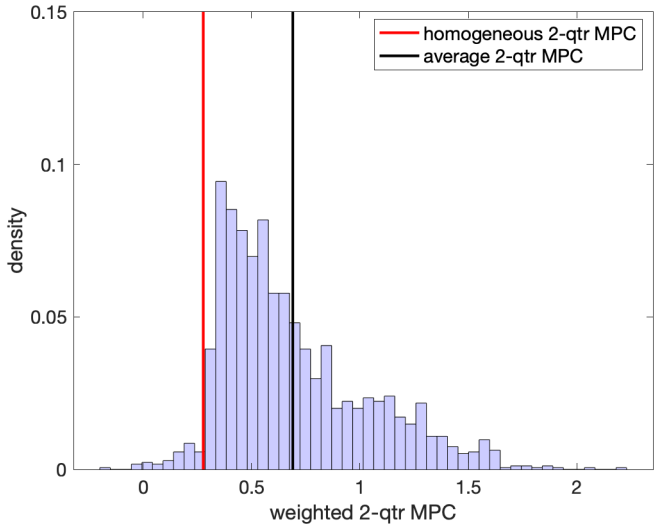
Notes: For each of 250 simulated samples, we regress the weighted MPCs for total expenditure estimated imposing spurious heterogeneity on the set of observables used in specification (2) of Table 3. Figure 6 plots the histogram (light blue bars) of the$t-$statistics for the coefficient on the APC. Total expenditures and baseline OLS specification. The red and black lines represent the critical values for a 5% test of equality with zero.

track individual weights over the two quarters following the rebate. We use these to construct the individual 2-quarter total effect of the rebate, by adding twice the weighted contemporaneous rebate coefficient to the weighted lagged coefficient.[39]

Figure 7 plots a histogram of this object among those who received the rebate. Relative to the baseline results depicted in Figure 1, the distribution spreads out, with some households having a total effect near zero. Moreover, as depicted in Figure 7, the estimated partial equilibrium effect of the tax rebate more than doubles relative to its homogeneous counterpart, from 0.28 to 0.69. Indeed, accounting for individual heterogeneity of treatment effects becomes even more important, even if the object of interest is simply the average response; as a share of 2008-Q1 aggregate consumption in our sample, the total effect is equivalent to 3.6%. Broda and Parker (2014) find that the stimulus increased consumption by 1.3% and 0.6% in Q2 and Q3, respectively.

[39]For example, a household may be categorized to be in some group a in the period in which they receive the rebate, and then in some group b the period after they receive the rebate. For such an individual, we construct the individual 2-quarter total effect of the rebate by adding twice the contemporaneous rebate coefficient for group a to the lagged rebate coefficient of group b, since ϕ_g^{lag} captures the change in consumption *relative* to consumption in the period of rebate receipt.

Figure 7: Estimated distribution of total 2-quarter effect of the tax rebate



Notes: Figure 7 plots the histogram (light blue bars) of the estimated distribution of the total effect of the 2008 ESA for total expenditures, defined as in Parker et al. (2013), using the lagged specification in Equation (10). The homogeneous MPC (red vertical line) is estimated assuming a homogeneous contemporaneous and homogeneous lagged response to the tax rebate, as in Parker et al. (2013). For each household we compute the weighted MPC, weighted across groups $g \in G$. The black vertical line shows the average weighted MPC in our sample.

5 Conclusion

We develop a flexible clustering approach to uncover heterogeneity in the marginal propensity to consume. In particular, our fuzzy C-parameters methodology extends the fuzzy C-means algorithm to regression problems, is well-suited to cross-sectional settings and has a better simulation performance than competing clustering approaches and quantile regressions. Our approach has three main advantages, which allow us to obtain as many empirical contributions. First, we can estimate the full unconditional distribution of MPCs, which could be driven both by latent factors and observable characteristics, without having to take a stand on the latter. Second, we can investigate, *ex post*, how MPCs individually correlate with observables, without being subject to the loss of statistical power that is likely to affect existing methodologies. Third, we explore how MPCs are jointly correlated with observables, and obtain a formal estimate of the portion of the MPC variance they explain.

Taking FCP to the data on the 2008 Economic Stimulus Payments, we find that households display a considerable degree of heterogeneity in their marginal propensities to consume. Even after estimating the full unconditional distribution, households spent at least one fourth of the rebate within one quarter. Nondurable consumption, however, is characterized by smoothing for most households. We do not find evidence of individual-

level substitution across consumption goods in response to transitory income shocks, but rather a very mild positive correlation. Finally, we explore what observables – individually and jointly – best predict different portions of the MPC distribution. We are able to retrieve various statistically significant relationships with the MPC, and show that only those with non-salary income and the APC survive the inclusion of additional drivers. Our findings suggest that there exists a tight relationship between marginal and average propensities to consume, which is easy to derive in many models of consumption behavior and yet has received relatively little attention. Moreover, since observable characteristics explain a minor portion of the estimated MPC heterogeneity, we posit that other latent factors, such as preference heterogeneity, might be important in determining marginal propensities to consume.

Finally, two caveats help to highlight possible avenues for future work. Importantly, we measure the distribution of MPCs to the 2008 tax rebate. This means our estimated distribution uses a single cross-section of data during a recession; if an individual's MPC is a function of the aggregate state, extrapolating our estimates requires caution. Second, because our empirical setting is one in which individuals only experience positive transitory shocks, we cannot speak to income windfalls, to which households may respond differently (Fuster et al. (2018)). However, the fuzzy C-parameters approach we develop can easily be applied to other datasets with suitably identified transitory income shocks, making comparisons straightforward. We leave such exercises for future work.

References

AGUIAR, M., C. BOAR, AND M. BILS (2019): "Who Are the Hand-to-Mouth?" in *2019 Meeting Papers*, Society for Economic Dynamics, 525.

ALAN, S., M. BROWNING, AND M. EJRNAES (2018): "Income and Consumption: A Micro Semistructural Analysis with Pervasive Heterogeneity," *Journal of Political Economy*, 126, 1827–1864.

AUCLERT, A. (2019): "Monetary Policy and the Redistribution Channel," *American Economic Review*, 109, 2333–67.

BEZDEK, J. (1981): *Pattern Recognition With Fuzzy Objective Function Algorithms*, New York: Plenum Press.

BEZDEK, J. C. (1973): "Fuzzy Mathematics in Pattern Classification," *Ph.D. dissertation, Cornell University*.

BEZDEK, J. C., R. EHRLICH, AND W. FULL (1984): "FCM: The Fuzzy C-Means Clustering Algorithm," *Computers & Geosciences*, 10, 191 – 203.

BLUNDELL, R., L. PISTAFERRI, AND I. PRESTON (2008): "Consumption inequality and partial insurance," *American Economic Review*, 98, 1887–1921.

BONHOMME, S. AND E. MANRESA (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147–1184.

BOUTROS, M. (2020): "bounded Intertemporal rationality and the Marginal Propensity to Consume," Tech. rep.

BRODA, C. AND J. A. PARKER (2014): "The economic stimulus payments of 2008 and the aggregate demand for consumption," *Journal of Monetary Economics*, 68, S20–S36.

CARROLL, C., J. SLACALEK, K. TOKUOKA, AND M. N. WHITE (2017): "The Distribution of Wealth and the Marginal Propensity to Consume," *Quantitative Economics*, 8, 977–1020.

CARROLL, C. D. (1992): "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity*, 23, 61–156.

COIBION, O., Y. GORODNICHENKO, L. KUENG, AND J. SILVIA (2017): "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 88, 70–89.

CRAWLEY, E. AND A. KUCHLER (2018): "Consumption Heterogeneity: Micro Drivers and Macro Implications," *Danish National Bank Working Paper 129*.

CUI, J., Q. LI, J. WANG, AND D. ZONG (2010): "Research on Selection Method of the Optimal Weighting Exponent and Clustering Number in Fuzzy C-means Algorithm," in *2010 International Conference on Intelligent Computation Technology and Automation*, vol. 3, 104–107.

DUNN, J. C. (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *Journal of Cybernetics*, 3, 32–57.

DYNAN, K. E., J. SKINNER, AND S. P. ZELDES (2004): "Do the Rich Save More?" *Journal of Political Economy*, 112, 397–444.

FAGERENG, A., M. B. HOLM, AND G. J. NATVIK (2016): "MPC Heterogeneity and Household Balance Sheets," Discussion Papers 852, Statistics Norway, Research Department.

FRIEDMAN, M. (1957): *A Theory of the Consumption Function*, Princeton University Press.

FUSTER, A., G. KAPLAN, AND B. ZAFAR (2018): "What Would You Do With $500? Spending Responses to Gains, Losses, News, and Loans," Staff Reports 843, Federal Reserve

Bank of New York.
GELMAN, M. (2019): "The Self-Constrained Hand to Mouth," .
ILUT, C. L. AND R. VALCHEV (2020): "Economic agents as imperfect problem solvers," Tech. rep., National Bureau of Economic Research.
JAPPELLI, T. AND L. PISTAFERRI (2014): "Fiscal Policy and MPC Heterogeneity," *American Economic Journal: Macroeconomics*, 6, 107–36.
JING, L., D. DENG, AND J. YU (2014): "Weighting Exponent Selection of Fuzzy C-Means via Jacobian Matrix," in *Knowledge Science, Engineering and Management*, ed. by R. Buchmann, C. V. Kifor, and J. Yu, Cham: Springer International Publishing, 115–126.
JOHNSON, D. S., J. A. PARKER, AND N. S. SOULELES (2006): "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review*, 96, 1589–1610.
KAPLAN, G., B. MOLL, AND G. L. VIOLANTE (2018): "Monetary Policy According to HANK," *American Economic Review*, 108, 697–743.
KAPLAN, G. AND G. L. VIOLANTE (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82, 1199–1239.
KAPLAN, G., G. L. VIOLANTE, AND J. WEIDNER (2014): "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, 45, 77–153.
KUENG, L. (2018): "Excess sensitivity of high-income consumers," *The Quarterly Journal of Economics*, 133, 1693–1751.
LUSARDI, A. (1996): "Permanent Income, Current Income, and Consumption: Evidence from Two Panel Data Sets," *Journal of Business & Economic Statistics*, 14, 81–90.
MIRANDA-PINTO, J., D. MURPHY, K. J. WALSH, AND E. R. YOUNG (2020): "A model of expenditure shocks," .
MISRA, K. AND P. SURICO (2014): "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs," *American Economic Journal: Macroeconomics*, 6, 84–106.
OLAFSSON, A. AND M. PAGEL (2018): "The liquid hand-to-mouth: Evidence from personal finance management software," *The Review of Financial Studies*, 31, 4398–4446.
PAL, N. R. AND J. C. BEZDEK (1995): "On Cluster Validity for the Fuzzy C-Means Model," *IEEE Trans. Fuzzy Systems*, 3, 370–379.
PARKER, J. A. (2017): "Why Don't Households Smooth Consumption? Evidence from a $25 Million Experiment," *American Economic Journal: Macroeconomics*, 9, 153–83.
PARKER, J. A., N. S. SOULELES, D. S. JOHNSON, AND R. MCCLELLAND (2013): "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, 103, 2530–53.
ROUSSEEUW, P., E. TRAUWAERT, AND L. KAUFMAN (1995): "Fuzzy clustering with high contrast," *Journal of Computational and Applied Mathematics*, 64, 81 – 90.
SAHM, C. R., M. D. SHAPIRO, AND J. SLEMROD (2010): "Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications," in *Tax Policy and the Economy, Volume 24*, National Bureau of Economic Research, Inc, NBER Chapters, 69–110.
SHAPIRO, M. D. AND J. SLEMROD (2009): "Did the 2008 Tax Rebates Stimulate Spending?" *American Economic Review*, 99, 374–79.
SŁOCZYŃSKI, T. (2020): "Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights," *The Review of Economics and Statistics*,

forthcoming, 1–27.

STRAUB, L. (2017): "Consumption, Savings, and the Distribution of Permanent Income," Working Paper.

TIBSHIRANI, R., G. WALTHER, AND T. HASTIE (2001): "Estimating the Number of Clusters in a Data Set via the Gap Statistic," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63, 411–423.

WINKLER, R., F. KLAWONN, AND R. KRUSE (2011): "Fuzzy clustering with polynomial fuzzifier function in connection with m-estimators," *Special Issue on Fuzzy Set Theory and its Applications*, 10, 146–163.

WU, K.-L. (2012): "Analysis of Parameter Selections for Fuzzy C-Means," *Pattern Recognition*, 45, 407–415.

YANG, M.-S. (1994): "On Asymptotic Normality of a Class of Fuzzy C-Means Clustering Procedures," *International Journal of General Systems*, 22, 391–403.

YANG, M.-S. AND K. F. YU (1992): "On Existence and Strong Consistency of a Class of Fuzzy C-Means Clustering Procedures," *Cybernetics and Systems*, 23, 583–602.

YU, J., Q. CHENG, AND H. HUANG (2004): "Analysis of the Weighting Exponent in the FCM," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34, 634–639.

Supplemental Appendix to "Latent Heterogeneity in the Marginal Propensity to Consume"

Daniel Lewis Davide Melcangi Laura Pilossoph

A Theoretical results

In this section, we provide additional theoretical results and proofs for the theorems included in the text. In particular, we characterize the behavior of HKM and FCM, relative to true cluster means, in a simple two Gaussian cluster example. Second, we provide detailed proofs establishing FCP regression as a GMM problem and the consistency and asymptotic normality of the resulting estimator. We include analytical expressions for the Hessian matrix required for the asymptotic variance. We also establish asymptotic properties of the FCP-IV estimator discussed in the text. Finally, we include a brief discussion of the relationship between FCP and finite mixture models.

A.1 Characterizing HKM and FCM in a simple case

The population optima of both HKM in fixed-T data (ψ^*) and FCM in general (ρ^*) represent pseudo-true parameters. For HKM, this is because the group assignment cannot be consistently estimated. For FCM, this is because the introduction of non-binary weights does not match the true model. However, we are not aware of any existing analytical results characterizing the properties of these pseudo-true parameters. In this section, we consider the problem of estimating the cluster means of two groups, where the data are generated from two Gaussian distributions with distinct means, the same finite variance, and equal mass, where the econometrician observes a single outcome for each observation. We show that there exists a smoothing parameter m for which ρ^* will recover the true parameter. While this is a highly simplified setting and m is unknown in practice, our simulation results show that for our regression problem and data-driven choice of m, the estimates are indeed quite close to the true parameter values.

First, we characterize the behavior of HKM in this homoskedastic two-Gaussian setting.

Proposition 1. *In the homoskedastic two Gaussian cluster case, ψ_1^*, the HKM global optimum for the lower mean, ζ_1, is given by*

$$\psi_1^* = \Phi\left(-\zeta_1/\sigma\right)\left(\zeta_1 + \frac{-\Phi'\left(-\zeta_1/\sigma\right)}{\Phi\left(-\zeta_1/\sigma\right)}\right) + \Phi\left(-\zeta_2/\sigma\right)\left(\zeta_2 + \frac{-\Phi'\left(-\zeta_2/\sigma\right)}{\Phi\left(-\zeta_2/\sigma\right)}\right),$$

and similarly for ζ_2, the higher mean, where σ is the standard deviation of both Gaussian clusters. ζ_1 is negatively biased and ζ_2 is positively biased unless $\sigma \to 0$.

Proof. Without loss of generality, we assume that $\zeta_1 = -\zeta_2$, so $(\zeta_1 + \zeta_2)/2 = 0$ (we can always demean all data before clustering). Denote the Gaussian with mean ζ_1 as G_1 and similarly G_2 for ζ_2, with $\zeta_1 < \zeta_2$, and denote their variance by σ^2. By symmetry, the groups are separated at 0, so values $y < 0$ are assigned to cluster 1 and $y > 0$ are assigned to cluster 2; $y = 0$ is a measure-zero event. Thus, the observations assigned to cluster 1 correspond to the portion of G_1 left of zero and the left tail of G_2. To compute ψ_1^*, it suffices to compute the mean over these two truncated normal distributions, weighted by their relative contribution to the cluster's mass:

$$\begin{aligned} E\left[y \mid g\left(y\right) = 1\right] = &\frac{\Pr_{G_1}\left(g\left(y\right) = 1\right)}{\Pr_{G_1}\left(g\left(y\right) = 1\right) + \Pr_{G_2}\left(g\left(y\right) = 1\right)} E_{G_1}\left[y \mid g\left(y\right) = 1\right] \\ &+ \frac{\Pr_{G_2}\left(g\left(y\right) = 1\right)}{\Pr_{G_1}\left(g\left(y\right) = 1\right) + \Pr_{G_2}\left(g\left(y\right) = 1\right)} E_{G_2}\left[y \mid g\left(y\right) = 1\right], \end{aligned}$$

where $g\left(y\right)$ denotes the group to which a value y is assigned. By symmetry, the total mass of the cluster is unity, so the relative contributions are simply $\Phi\left(-\zeta_1/\sigma\right)$ and $\Phi\left(-\zeta_2/\sigma\right)$ respectively, where Φ is the standard normal CDF. Finally, it remains to compute the means for each of the two truncated normals. Using standard results for the mean of the truncated normal distribution, with lower bound equal to $-\infty$ and upper bound equal to 0, we obtain the result,

$$\psi_1^* = \Phi\left(-\zeta_1/\sigma\right)\left(\zeta_1 + \frac{-\Phi'\left(-\zeta_1/\sigma\right)}{\Phi\left(-\zeta_1/\sigma\right)}\right) + \Phi\left(-\zeta_2/\sigma\right)\left(\zeta_2 + \frac{-\Phi'\left(-\zeta_2/\sigma\right)}{\Phi\left(-\zeta_2/\sigma\right)}\right),$$

with a symmetric argument giving a similar expression for ψ_2^*.

To conclude that ψ_1^* is negatively biased, note that in computing the mean over cluster 1, the right tail of G_1 (right of zero) has been replaced by an equal mass to the left of zero, shifting the overall mean to the left. ψ_1^* in general only recovers ζ_1^* by taking the limit as $\sigma \to 0$. □

Proposition 1 shows that each of the global minimum of the HKM objective function for each of the cluster means is biased *outwards* (away from zero), with the lower mean being further reduced, and the upper mean being increased. The intuition is that ψ_1^* represents the mean over two truncated normals (portions of the mean ζ_1 and ζ_2 distributions), with the right tail of the distribution with mean ζ_1 truncated and replaced with an equal-mass portion of the ζ_2 distribution that is situated to the left of that tail. In practical terms, this means that HKM overstates cluster heterogeneity in this simple case.

While we cannot similarly characterize the FCM optima analytically, Proposition 2 provides an existence result for a value $\tilde{m}$ such that FCM recovers the truth.

Proposition 2. *In the homoskedastic two Gaussian case, assuming ρ_1^* and ρ_2^* are unique, there exists some $\tilde{m} \in (1, \infty)$ such that $\rho_1^*(\tilde{m}) = \zeta_1, \rho_2^*(\tilde{m}) = \zeta_2$.*

Proof. We begin by showing that $\rho_i^*(m)$ is everywhere differentiable in m. ρ_i^* is implicitly defined by the moment equation

$$f_i(m, \rho) = \int_{-\infty}^{\infty} \left(1 + \frac{\left\| y_i - \rho_i^* \right\|^{\frac{2}{m-1}}}{\left\| y_i - \rho_j^* \right\|^{\frac{2}{m-1}}} \right)^{-m} (y - \rho_i^*) \, dP(y) = 0, i \neq j$$

(for a formal argument that the FCM clustering problem can be represented as a method of moments problem see Proposition 5). By the implicit function theorem, since ρ_1^* and ρ_2^* are unique,

$$\frac{d\rho_i^*(m)}{dm} = \left[-\frac{\partial f}{\partial \rho'}(m, \rho^*(m)) \right]^{-1} \left[-\frac{df}{dm}(m, \rho^*(m)) \right],$$

where ρ stacks ρ_1, ρ_2 in a vector (and similarly f stacks f_1, f_2). Since ρ^* is assumed to be the unique solution to $f(m, \rho)$, the first term (the inverse of the Jacobian of the moments) exists. The second term can be simplified to

$$\int_{-\infty}^{\infty} 2(m-1)^{-2} \left[1 + \frac{\left\| y - \rho_i^* - \zeta_i \right\|^{\frac{2}{m-1}}}{\left\| y - \rho_j^* - \zeta_j \right\|^{\frac{2}{m-1}}} \right]^{-m} \ln \left[1 + \frac{\left\| y - \rho_i^* - \zeta_i \right\|^{\frac{2}{m-1}}}{\left\| y - \rho_j^* - \zeta_j \right\|^{\frac{2}{m-1}}} \right]$$
$$\times \frac{\left\| y - \rho_i^* - \zeta_i \right\|^{\frac{2}{m-1}}}{\left\| y - \rho_j^* - \zeta_j \right\|^{\frac{2}{m-1}}} \ln \left(\frac{\left\| y - \rho_i^* - \zeta_i \right\|}{\left\| y - \rho_j^* - \zeta_j \right\|^{\frac{2}{m-1}}} \right) (y - \rho_i^* - \zeta_i) \, dP(y).$$

For a given m (and thus ρ^*), the integrand is clearly finite for finite y except for at the point where $y = \rho_j^* + \zeta_j$ (and infinite y are probability zero since the variance of each

Gaussian component is assumed to be finite). Further, denoting $b_j = \left\| y - \rho_j^* - \zeta_j \right\|^{\frac{2}{m-1}}$, $a_i = \left\| y - \rho_i^* - \zeta_i \right\|^{\frac{2}{m-1}}, \tilde{a}_i = \left\| y - \rho_i^* - \zeta_i \right\|$

$$
\begin{aligned}
&\lim_{b_j \to 0} \left[1 + \frac{a_i}{b_j}\right]^{-m} \ln\left[1 + \frac{a_i}{b_j}\right] \frac{a_i}{b_j} \ln\left(\frac{\tilde{a}_i}{b_j}\right) \\
&= \lim_{b_j \to 0} \frac{\left[\frac{b_j + a_i}{b_j}\right]^{-m} a_i \ln\left[\frac{a_i + b_j}{b_j}\right] \ln\left(\frac{\tilde{a}_i}{b_j}\right)}{b_j} \\
&= \lim_{b_j \to 0} \frac{b_j^{m-1} a_i \left(\ln\left(a_i + b_j\right) - \ln\left(b_j\right)\right) \left(\ln\left(\tilde{a}_i\right) - \ln\left(b_j\right)\right)}{\left(b_j + a_i\right)^m} \\
&= \lim_{b_j \to 0} \frac{a_i}{\left(b_j + a_i\right)^m} \left[b_j^{m-1} \ln\left(a_i + b_j\right) \ln\left(\tilde{a}_i\right) - b_j^{m-1} \ln\left(a_i + b_j\right) \ln\left(b_j\right) \right. \\
&\qquad \left. - b_j^{m-1} \ln\left(b_j\right) \ln\left(\tilde{a}_i\right) + b_j^{m-1} \ln\left(b_j\right) \ln\left(b_j\right)\right]
\end{aligned}
$$

The term outside the brackets is finite in the limit. We take the second part term-by-term. The first is zero in the limit. The remaining terms are also each zero in limit by application of l'Hôpital's rule. Having argued that the integrand is finite for all points with positive probability under P, the integral exists. Thus $\frac{\partial f}{\partial m}\left(m, \rho^*\left(m\right)\right)$ exists everywhere, so $\frac{\partial \rho_i^*(m)}{\partial m}$ exists everywhere. Since $\rho_i^*\left(m\right)$ is a univariate function, existence of the derivative is sufficient for differentiability to hold. Since $\rho_i^*\left(m\right)$ is thus everywhere differentiable for $m \in (1, \infty)$, $\rho_i^*\left(m\right)$ is everywhere continuous. We know that $\lim_{m \to 1} \rho_1^*\left(m\right) = \psi_1^* < \zeta_1$, and that $\lim_{m \to \infty} \rho_1^*\left(m\right) = 0$. Note that given the normalization $\left(\zeta_1 + \zeta_2\right)/2 = 0$, $\zeta_1 < 0$. Therefore, by the intermediate value theorem, there exists some $\tilde{m} \in (1, \infty)$ such that $\rho_1^*\left(\tilde{m}\right) = \mu_1$. The same trivially holds for ρ_2^*. □

Of course, in practice we do not know $\tilde{m}$, even in this simple case; Proposition 2 is merely an existence result. However, Propositions 1 and 2 jointly motivate the use of FCM, by demonstrating that choosing $m > 1$ has the potential to recover the true parameters in this simple problem.

A.2 Asymptotic properties of FCP

In the text, we argue that FCP can be viewed as a GMM problem and provide consistency and asymptotic normality results for $\hat{\theta}$ in Theorems 1 and 2. We state results tailored to our empirical setting, where $y \in \mathbb{R}$, a cross-sectional structure. However, all results also apply for panel structure, where y is a $T \times 1$ vector, $y \in \mathbb{R}^T$. We state results here for that more general model for completeness.

We first restate the assumed model in this more general setting

$$y_i = \sum_{g=1}^{G^*} \mathbf{1}\left[i \in g\right] \theta_g x_i + \varepsilon_i, \quad i = 1, \ldots, N, \tag{11}$$

where $y_i \in \mathbb{R}^T$, $x_i \in \mathbb{R}^K$, (y_i, x_i) are i.i.d. according to the probability measure Π on y, x (denoting the conditional for y by $\Pi_{y|x}$ and the marginal for x by Π_x), $E\left[\varepsilon_i \mid x_i, i \in g\right] = 0$, and θ_g is a $T \times K$ matrix. Note that the K dimensions of x_i may contain observations of the same characteristic, repeated across time.

Define the FCP objective function as

$$J_m^{reg}\left(\Pi, \theta\right) = \int \int \sum_{g=1}^{G} \mu_g^{reg}\left(y \mid x; \theta\right)^m \left\|y - \theta_g x\right\|^2 \Pi_{y|x}\left(dy \mid x\right) \Pi\left(dx\right), \tag{12}$$

where $\theta \in \Theta \subset R^{G \times T \times K}$ and

$$\mu_g^{reg}\left(y \mid x; \theta\right) = \left(\sum_{h=1}^{G} \frac{\left\|y - \theta_g x\right\|^{2/(m-1)}}{\left\|y - \theta_h x\right\|^{2/(m-1)}} \right)^{-1}, g = 1, \ldots, G.$$

We first formally state Assumption 1, outlined in the main text, which presents standard grouped OLS assumptions:

Assumption 1. *(Grouped OLS Assumptions)*

1. *Observations (y_i, x_i) are generated according to (6) and jointly i.i.d. from Π,*
2. *$E\left[\varepsilon_i \mid x_i, i \in g\right] = 0 \; \forall g = 1, \ldots, G$,*
3. *The second moments of y and x are finite under Π,*
4. *The regressors x are not collinear, rank $\left(\int xx' \Pi_x\left(dx\right)\right) = K$,*
5. *G is finite.*

Following the development of Yang and Yu (1992), we begin by proving the equivalence of the WLS objective J_m^{reg} to an objective function subsuming weights into the loss function, L_m^{reg}.

Proposition 3. *(Equivalence) Under Assumption 1, $J_m^{reg}\left(\Pi, \theta\right) = L_m^{reg}\left(\Pi, \theta\right)$, where*

$$L_m^{reg}\left(\Pi, \theta\right) = \int \int \left(\sum_{g=1}^{G} \left\|y - \theta_g x\right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}\left(dy \mid x\right) \Pi_x\left(dx\right); \tag{13}$$

a minimizer $\theta^* \in \Theta$ *of* $L_m^{reg}(\Pi, \theta)$ *is also a minimizer of* $J_m^{reg}(\Pi, \theta)$ *over* Θ *and weights* μ^{reg}.

Proof. The first point follows from simple algebra and the definition of μ_g^{reg}. In particular,

$$\begin{aligned} J_m^{reg}(\Pi, \theta) &= \int\int \sum_{g=1}^{G} \mu_g^{reg,m}(y \mid x; \theta) \left\|y - \theta_g x\right\|^2 \Pi_{y|x}(dy \mid x) \Pi_x(dx) \\ &= \int\int \sum_{g=1}^{G} \left(\sum_{h=1}^{G} \frac{\left\|y - \theta_g x\right\|^{2/(m-1)}}{\left\|y - \theta_h x\right\|^{2/(m-1)}} \right)^{-m} \left\|y - \theta_g x\right\|^2 \Pi_{y|x}(dy \mid x) \Pi_x(dx) \\ &= \int\int \left(\sum_{g=1}^{G} \left\|y - \theta_g x\right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x) \Pi_x(dx), \end{aligned}$$

which is the formulation of $L_m^{reg}(\Pi, \theta)$.

Parallel to the development of Yang and Yu (1992) for the cluster means case, the minimization problem of J_m^{reg} is equivalent to that of L_m^{reg}. Note that Lemmata 1 and 2 from Yang and Yu (1992) hold trivially in this setting since they make no reference to the form of the square error function; combining them as in that paper completes the proof. □

Proposition 3 demonstrates that we can turn our attention from the WLS objective function $J_m^{reg}(\Pi, \theta)$ to the compact formulation in $L_m^{reg}(\Pi, \theta)$, without weights. Under additional regularity conditions, we now establish the existence of a solution to the FCP problem in Equation (12).

Proposition 4. *(Existence) If Assumption 1 holds, then for any* $G = 1, 2, \ldots$, *there exists a solution* θ^* *such that*

$$L_m^{reg}(\Pi, \theta^*) = \inf_{\theta} L_m^{reg}(\Pi, \theta).$$

Proof. The proof is a straightforward extension of the proof of Theorem 1 in Yang and Yu (1992). Define $a(G) = \inf_\theta L_m^{reg}(\Pi, \theta)$ obtained for G groups. If Π is degenerate at some set of $T \times K$ matrices $\tilde{\theta}_g$, then $a(G) = 0$ and $\theta^* = \{\tilde{\theta}_1, \ldots \tilde{\theta}_G\}$. Therefore we can restrict our attention to non-degenerate Π. When $G = 1$, (13) reduces to

$$\int\int \left\|y - \theta x\right\|^2 \Pi_{y|x}(dy \mid x) \Pi_x(dx),$$

which is the standard OLS objective function, which has the familiar solution $\theta_1^* = E[yx'] E[xx']^{-1}$ (the slightly unusual form accommodates y being $T \times 1$, T possibly greater than 1). Consider $G = 2$. Denote θ_{21} as the parameters for group 1 with $G = 2$, and let $\theta_{21} = \theta_1^*$, with

θ_{22} arbitrary. Then

$$\begin{aligned} a(2) &\leq \int\int \left(\sum_{g=1}^{2} \left\| y - \theta_{2g} x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x) \Pi_x(dx) \\ &< \int\int \left(\sum_{g=1}^{1} \left\| y - \theta_{g} x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x) \Pi_x(dx) \qquad (14) \\ &= a(1) < \infty, \end{aligned}$$

where the second inequality is strict since Π is not degenerate and $1 - m < 0$. Since $a(2) < \infty$, there exists $\theta^{(r)}(2) = \left(\theta_{21}^{(r)}, \theta_{22}^{(r)}\right)$ such that as $r \to \infty$,

$$\int\int \left(\sum_{g=1}^{2} \left\| y - \theta_{2g}^{(r)} x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x) \Pi_x(dx) \to a(2).$$

We want to show that $\left\{ \theta^{(r)}(2), r \geq 1 \right\}$ is bounded. Suppose the statement is false, so there exists a subsequence $\theta_{21}^{(r_j)}$ such that $\left\| \theta_{21}^{(r_j)} \right\|$ goes to infinity. Then

$$\begin{aligned} a(2) &= \lim_{r_j \to \infty} \int\int \left(\sum_{g=1}^{2} \left\| y - \theta_{2g}^{(r_j)} x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x) \Pi_x(dx) \\ &\geq \liminf_{r_j \to \infty} \int\int \left(\sum_{g=1}^{2} \left\| y - \theta_{2g}^{(r_j)} x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x) \Pi_x(dx) \\ &\geq \int\int \liminf_{r_j \to \infty} \left(\sum_{g=1}^{2} \left\| y - \theta_{2g}^{(r_j)} x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x) \Pi_x(dx) \\ &\geq \int\int \liminf_{r_j \to \infty} \left\| y - \theta_{22}^{(r_j)} x \right\|^{2} \Pi_{y|x}(dy \mid x) \Pi_x(dx) \\ &\geq a(1), \end{aligned}$$

where the second inequality follows from Fatou's Lemma and the third uses the fact that $\left\| \theta_{21}^{(r_j)} \right\|$ goes to infinity. The result contradicts (14). Thus, $\left\{ \theta^{(r)}(2), r \geq 1 \right\}$ is bounded and there exist $\theta_{21}^*, \theta_{22}^*$ such that $\theta^{(r)}(2)$ converges to $\theta^*(2) = (\theta_{21}^*, \theta_{22}^*)$ along a subsequence, say r_j. Then for all $\delta > 0$, there exists r_0 such that for all $r_j > r_0$, $\left\| \theta^{(r_j)}(2) - \theta^*(2) \right\| \leq \delta$.

Thus

$$
\begin{aligned}
&\left(\sum_{g=1}^{2}\left\|y-\theta_{2g}^{(r_j)}x\right\|^{-2/(m-1)}\right)^{1-m}\\
&\leq \max_{1\leq i\leq 2}\left\|y-\theta_{2g}^{(r_j)}x\right\|^{2}\\
&\leq \left\|y-\theta_{21}^{(r_j)}x\right\|^{2}+\left\|y-\theta_{22}^{(r_j)}x\right\|^{2}\\
&\leq \left(\|y\|+\left\|\theta_{21}^{(r_j)}x\right\|\right)^{2}+\left(\|y\|+\left\|\theta_{22}^{(r_j)}x\right\|\right)^{2}\\
&\leq \left(\|y\|+\left\|\theta_{21}^{(r_j)}\right\|\|x\|\right)^{2}+\left(\|y\|+\left\|\theta_{22}^{(r_j)}\right\|\|x\|\right)^{2}\\
&\leq \left(\|y\|+\|\theta_{21}^{*}\|\,\|x\|+\|\delta\iota_{T\times K}\|\,\|x\|\right)^{2}+\left(\|y\|+\|\theta_{21}^{*}\|\,\|x\|+\|\delta\iota_{T\times K}\|\,\|x\|\right)^{2},
\end{aligned}
$$

where the third inequality follows from the triangle inequality, the fourth follows from Cauchy-Schwarz, and the last line follows from the triangle inequality and the fact that $\delta \geq \left\|\theta^{(r_j)}(2)-\theta^{*}(2)\right\|$ implies $\left|\theta_{2g,tk}^{(r_j)}\right| \leq \left|\theta_{2g,tk}^{*}\right|+\delta$ for all g,t,k where t indexes dimensions of y and k indexes dimensions of x, which then implies $\left\|\theta_{2g}^{(r_j)}\right\| \leq \left\|\theta_{2g}^{*}\right\|+\|\delta\iota_{T\times K}\|$. By Assumption 1, the last line provides a bound in expectation for the left hand side. Finally, since the last line establishes a bounding function for $\left(\sum_{g=1}^{2}\left\|y-\theta_{2g}^{(r_j)}x\right\|^{-2/(m-1)}\right)^{1-m}$, the dominated convergence theorem shows that as r_j tends to infinity,

$$
\begin{aligned}
a(2) &= \lim_{r_j\to\infty}\int\int\left(\sum_{g=1}^{2}\left\|y-\theta_{2g}^{(r_j)}x\right\|^{-2/(m-1)}\right)^{1-m}\Pi_{y|x}(dy\mid x)\,\Pi_{x}(dx)\\
&= \int\int\left(\sum_{g=1}^{2}\left\|y-\theta_{2g}^{*}x\right\|^{-2/(m-1)}\right)^{1-m}\Pi_{y|x}(dy\mid x)\,\Pi_{x}(dx).
\end{aligned}
$$

This establishes that the infimum $a(2)$ is indeed obtained at $\theta^{*}(2)$, the limit of $\theta^{(r)}(2)$ for subsequence r_j (which exists). A similar argument can then be made sequentially for $G = 3,4,\ldots$, so by mathematical induction, the theorem is therefore true for all $G = 1,2,\ldots$. □

The relationship between θ^{*} and the true parameter, θ_0, generating the data in Equation (11) merits further discussion. θ^{*} is a *pseudo-true* parameter, which may not be in

general be equivalent to θ_0, since it optimizes an objective function that is not perfectly aligned with the the true data generating process (DGP). J_m^{reg} puts non-zero weight on individuals being members of groups other than their (unknown) true group. HKM also recovers pseudo-true parameters in fixed$-T$ environments, as noted in the supplemental appendix to Bonhomme and Manresa (2015), since group membership cannot be consistently estimated, even though the objective function appropriately represents the DGP. The tuning parameter m trades off two factors influencing θ^*: as $m \to 1$, J_m^{reg} approaches the true DGP, but as m increases, the potential impact of miss-assigned individuals on optimal parameters also increases as the weights approach binary values. Under two special limiting cases, however, $\theta^* = \theta_0$. First, as the degree of separation of the groups diverges, $\sum_{h=1}^{G} \frac{\|y-\theta_{g_0}x\|^{2/(m-1)}}{\|y-\theta_h x\|^{2/(m-1)}} \to 1$,where g_0 is the true group, so the FCP objective function L_m^{reg} converges to its HKM counterpart, and both recover the truth. Second, as $m \to 1$, under additional weak dependence assumptions (like those in Bonhomme and Manresa (2015)), the weight on the true group will likewise converge to unity asymptotically in T. While neither of these conditions describes our dataset, or indeed most macroeconomic datasets, our extensive simulations calibrated to our empirical setting show that θ^* can still be very close to θ_0. Moreover, even if the population values for the group parameters themselves may not align precisely with the true values, we show that the distribution of MPCs recovered matches the true distribution very well, on average.

We now show that FCP can be solved as a GMM problem.

Proposition 5. *(Moments) The solution θ^* satisfies the moment equations*

$$E\left[\left(\sum_{h=1}^{G}\frac{\|y_i-\theta_g x_i\|^{2/(m-1)}}{\|y_i-\theta_h x_i\|^{2/(m-1)}}\right)^{-m}\left(y_{it}-\theta_{g,(t)}x_i\right)x_i\right]=0 \text{ for } g=1,\ldots,G \text{ and } t=1,\ldots,T, \tag{15}$$

where t indexes dimensions of y_i and (t) rows of θ_g; FCP is a GMM problem.

Proof. We start by differentiating the integrand of $L_m^{reg}(\Pi, \theta)$ with respect to $\theta_{g,tk}$:

$$
\begin{aligned}
&\frac{\partial}{\partial \theta_{g,tk}} \left(\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2/(m-1)} \right)^{1-m} \\
&= (1-m) \left(\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2/(m-1)} \right)^{-m} \frac{\partial}{\partial \theta_{g,tk}} \sum_{g=1}^{G} \left\| y - \theta_g x \right\|^{-2/(m-1)} \\
&= (1-m) \left(\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2/(m-1)} \right)^{-m} \frac{-2}{m-1} \left\| y - \theta_g x \right\|^{(1+m)/(1-m)} \frac{\partial}{\partial \theta_{g,tk}} \left\| y - \theta_g x \right\| \\
&= 2 \left(\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2/(m-1)} \right)^{-m} \left\| y - \theta_g x \right\|^{-(1+m)/(m-1)} \frac{y_t - \theta_{g,(t)} x}{\left\| y - \theta_g x \right\|} (-x_k) \\
&= -2 \left(\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2/(m-1)} \right)^{-m} \left\| y - \theta_g x \right\|^{-2m/(m-1)} \left(y_t - \theta_{g,(t)} x \right) x_k \\
&= -2 \left(\sum_{h=1}^{G} \frac{\left\| y - \theta_g x \right\|^{2/(m-1)}}{\left\| y - \theta_h x \right\|^{2/(m-1)}} \right)^{-m} \left(y_t - \theta_{g,(t)} x \right) x_k,
\end{aligned}
$$

where $\theta_{g,(t)}$ denotes the row of θ_g corresponding to outcome y_t. Note that since these partial derivatives are continuous in θ (by inspection; see also Yang (1994) Lemma 2), the integrand is (continuously) differentiable in θ (Spivak (1971) Theorem 2.8). Moreover, $\left(\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2/(m-1)} \right)^{1-m}$ is Lebesgue integrable for each θ as

$$
\left(\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2/(m-1)} \right)^{1-m} \leq \sum_{h=1}^{G} \left\| y - \theta_h x \right\|^{-2(1-m)/(m-1)} = \sum_{h=1}^{G} \left\| y - \theta_h x \right\|^2
$$

since $1 - m < 0$ and

$$
\begin{aligned}
\sum_{h=1}^{G} \left\| y - \theta_h x \right\|^2 &\leq \sum_{h=1}^{G} \left(\|y\| + \|\theta_h x\| \right)^2 \\
&\leq \sum_{h=1}^{G} \left(\|y\| + \|\theta_h\| \, \|x\| \right)^2 \\
&= \sum_{h=1}^{G} \|y\|^2 + 2 \|\theta_h\| \, \|x\| \, \|y\| + \|\theta_h\|^2 \|x\|^2, \qquad (16)
\end{aligned}
$$

which is integrable by Assumption 1. Moreover, (16) establishes a bounding function for the integrand in terms of θ. From these conditions, the dominated convergence theorem allows the interchange of differentiation and integration:

$$
\begin{aligned}
\frac{\partial L_m^{reg}(\Pi,\theta)}{\partial \theta_{g,tk}} &= \frac{\partial}{\partial \theta_{g,tk}} \int\int \left(\sum_{g=1}^{G} \left\| y - \theta_g x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x)\, \Pi_x(dx) \\
&= \int\int \left(\frac{\partial}{\partial \theta_{g,tk}} \sum_{g=1}^{G} \left\| y - \theta_g x \right\|^{-2/(m-1)} \right)^{1-m} \Pi_{y|x}(dy \mid x)\, \Pi_x(dx) \\
&= E\left[\left(\frac{\partial}{\partial \theta_{g,tk}} \sum_{g=1}^{G} \left\| y_i - \theta_g x_i \right\|^{-2/(m-1)} \right)^{1-m} \right] \\
&= E\left[-2 \left(\sum_{h=1}^{G} \frac{\left\| y_i - \theta_g x_i \right\|^{2/(m-1)}}{\left\| y_i - \theta_h x_i \right\|^{2/(m-1)}} \right)^{-m} \left(y_{it} - \theta_{g,(t)} x_i \right) x_{i,k} \right],
\end{aligned}
$$

where we henceforth replace the Lebesgue integrals with expectations. Stacking the conditions vertically for row t of θ_g yields the $K \times 1$ vector

$$
\frac{\partial L_m^{reg}}{\partial \theta_{g,(t)}'} = E\left[-2 \left(\sum_{h=1}^{G} \frac{\left\| y_i - \theta_g x_i \right\|^{2/(m-1)}}{\left\| y_i - \theta_h x_i \right\|^{2/(m-1)}} \right)^{-m} \left(y_{it} - \theta_{g,(t)} x_i \right) x_i \right].
$$

Proceeding likewise across $t = 1, \ldots, T$ and for $g = 1, \ldots, G$ yields $G \times T \times K$ conditions which θ^* must satisfy,

$$
E\left[\left(\sum_{h=1}^{G} \frac{\left\| y_i - \theta_g x_i \right\|^{2/(m-1)}}{\left\| y_i - \theta_h x_i \right\|^{2/(m-1)}} \right)^{-m} \left(y_{it} - \theta_{g,(t)} x_i \right) x_i \right] = 0, \text{ for } g = 1, \ldots, G, t = 1, \ldots, T,
$$

since θ^* minimizes $L_m^{reg}(\Pi, \theta)$. These $G \times T \times K$ equations constitute moment conditions for the $G \times T \times K$ free parameters in θ. Thus, the objective function (13) constitutes a just-identified GMM problem. □

Proposition 5 has two important implications. First, reframing FCP as a GMM problem allows the asymptotic properties of estimators $\hat{\theta}$ to be derived using standard theory. Second, it justifies a numerical approach to solve the moment equations (15), in addition to the standard iterative procedure used in the FCM literature. The moment equations in Proposition 5 can easily accommodate regressors with common coefficients across groups, $\theta_{g,tk} = \theta_{h,tk}$, or additionally across dimensions of y_i, $\theta_{g,tk} = \theta_{h,sk}$. The former applies to the coefficients on time dummies in our model, for example. In this case, it is

straightforward to show that the corresponding moment condition is

$$E\left[\left(\sum_{h=1}^{G}\left\|y_i-\theta_h x_i\right\|^{-2/(m-1)}\right)^{-m}\sum_{g=1}^{G}\left\|y_i-\theta_g x_i\right\|^{-2m/(m-1)}\left(y_{it}-\theta_{g,(t)}x_i\right)x_{ik}\right]=0$$

and in the case of common coefficients across both g and t,

$$E\left[\left(\sum_{h=1}^{G}\left\|y_i-\theta_h x_i\right\|^{-2/(m-1)}\right)^{-m}\sum_{t=1}^{T}\sum_{g=1}^{G}\left\|y_i-\theta_g x_i\right\|^{-2m/(m-1)}\left(y_{it}-\theta_{g,(t)}x_i\right)x_{ik}\right]=0.$$

The moment conditions in Equation (15) have natural sample counterparts that can be used to define the estimator $\hat{\theta}$, as defined in Equation (9). Under the additional conditions in Assumption 2, detailed in the text, $\hat{\theta}$ is consistent for θ^*. Assumption 2.1 is an identification condition. In clustering models, if identification holds, it does so up to a labeling of the groups g, which can always be permuted. The FCM literature (and HKM when T is finite, e.g., Pollard (1981, 1982)) always assumes the uniqueness of θ^*. While it appears intractable to characterize primitive conditions under which identification holds in the non-linear form of Equation (13), the condition is closely linked to OLS identification. With known membership, Assumption 1 suffices for OLS to uniquely identify the true parameters, and for relatively small m, the objective (13) is a perturbation around the OLS objective. In a formal sense, identification in clustering models when group membership cannot be recovered remains a topic for future work.

Assumption 2 adds standard assumptions for consistency of the GMM estimator, noted in the main text:

Assumption 2. *(Consistency)*

1. *θ^* is the unique solution to $E\left[\eta\left(\theta, y_i, x_i\right)\right]=0$ (up to ordering of the groups),*
2. *Θ is compact.*

Theorem 1 shows that $\hat{\theta}$ is consistent for θ^*, the population solution to Equation 7.

Theorem 1. *(Consistency) Under Assumptions 1-2, $\hat{\theta} \xrightarrow{p} \theta^*$ as $N \to \infty$.*

Proof. By Assumption 1, (y_i, x_i) are i.i.d.. By Assumption 2, θ^* uniquely satisfies $\eta\left(\theta, y_i, x_i\right)$. As noted in the proof of Corollary 5, the moment functions $\eta\left(\theta, y_i, x_i\right)$ are continuous for all $\theta \in \Theta$. Next, we show that the moments are bounded in expectation for all $\theta \in \Theta$ (the dominance condition). Observe that $\left(\sum_{h=1}^{G}\frac{\left\|y-\theta_g x\right\|^{2/(m-1)}}{\left\|y-\theta_h x\right\|^{2/(m-1)}}\right)^{-m}$ is bounded between zero

and one (the supremum of the summation is infinity as the residuals $y - \theta_h x$, $h \neq g$ go to zero and the infimum is 1 as $y - \theta_h x$, $h \neq g$ go to infinity). So

$$\begin{aligned} E\left[\sup_{\theta\in\Theta}\|\eta\left(\theta, y_i, x_i\right)\|\right] &\leq E\left[\sup_{\theta\in\Theta}\sup_{g}\left\|\left(y_i - \theta_g x_i\right)x_i'\right\|\right] \\ &= E\left[\sup_{\theta\in\Theta}\sup_{g}\left\|y_i x_i' - \theta_g x_i x_i'\right\|\right] \\ &\leq E\left[\sup_{\theta\in\Theta}\sup_{g}\left\|y_i x_i'\right\| + \left\|\theta_g x_i x_i'\right\|\right] \\ &\leq E\left[\sup_{\theta\in\Theta}\sup_{g}\|y_i\|\,\|x_i\| + \left\|\theta_g\right\|\|x_i\|\,\|x_i\|\right] \\ &< \infty, \end{aligned}$$

where the third inequality follows from the triangle inequality, the fourth from Cauchy-Schwarz, and the final follows from Assumption 1. These points jointly satisfy the requirements of standard GMM arguments, (e.g., Newey and McFadden (1994), p. 2121—2, Hayashi (2011) Proposition 7.7), so $\hat{\theta} \xrightarrow{p} \theta^*$. □

To characterize the asymptotic distribution, we require further technical conditions that are standard for the asymptotic normality of GMM estimators:

Assumption 3. *(Asymptotic Normality)*

1. θ^* *is in the interior of* Θ,
2. $H = E\left[\frac{\partial\eta\left(\theta, y_{i,}, x_i\right)}{\partial vec(\theta)'}\right]$ *is full rank,*
3. $E\left[\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial\eta\left(\theta, y_{i,}, x_i\right)}{\partial vec(\theta)'}\right\|\right] < \infty$ *in a neighborhood* $\mathcal{N}$ *of* θ^*,
4. $E\left[\eta\left(\theta^*, y_{i,}, x_i\right)\eta\left(\theta^*, y_{i,}, x_i\right)'\right]$ *is positive definite.*

Theorem 2 gives the limiting distribution of $\hat{\theta}$:

Theorem 2. *(Asymptotic Normality) Under Assumptions 1 - 3,*

$$\sqrt{N}\left(vec\left(\hat{\theta}\right) - vec\left(\theta^*\right)\right) \xrightarrow{d} \mathcal{N}\left(0, H^{-1}VH^{-1}\right),$$

where

$$V = E\left[\eta\left(\theta, y_i, x_i\right)\eta\left(\theta, y_i, x_i\right)'\right],$$

and H is the Hessian of Equation (13).

Proof. First, we provide expressions for H to establish continuous differentiability of $\eta\left(\theta, y_i, x_i\right)$ in θ. We focus on the cross-sectional case here ($T = 1$, so θ_g is a $K \times 1$ vector) for the sake of brevity and in keeping with our empirical focus, but provide fully general expressions for panel data below. Partition the blocks of H as

$$H = \begin{bmatrix} H_{11} & \cdots & H_{1g} & \cdots & H_{1G} \\ \vdots & \ddots & & & \vdots \\ H_{g1} & & H_{gg} & & H_{gG} \\ \vdots & & & \ddots & \vdots \\ H_{G1} & \cdots & H_{Gg} & \cdots & H_{GG} \end{bmatrix},$$

where $H_{gh} = \frac{\partial^2 L_m^{reg}}{\partial\theta_g \partial\theta_h'}$,with $H_{gh} = H_{hg}'$ by symmetry of the Hessian. For the case where all coefficients are group-specific, it can be shown that

$$\begin{aligned} H_{gg} &= E\left[x_i x_i' \left\{\frac{-2m}{m-1} A_i^{-m-1}\left(e_{i,g}\right)^2 C_{i,g}^2 + \frac{m+1}{m-1} A_i^{-m} C_{i,g}\right\}\right] \\ H_{gh} &= E\left[x_i x_i' \left\{\frac{-2m}{m-1} A_i^{-m-1} C_{i,h} e_{i,h} e_{i,g} C_{i,g}\right\}\right], h \neq g, \end{aligned}$$

where $e_{i,g} = y_i - \theta_g x_i$, $A_i = \Sigma_{g=1}^G \left\|e_{i,g}\right\|^{-2/(m-1)}$, $C_{ig} = \left\|e_{i,g}\right\|^{-2m/(m-1)}$. We also provide expressions for additional elements of the Hessian when there are covariates with common coefficients across groups, such that $\theta_{g,k} = \theta_{h,k} \equiv \theta_{\star,k}, h \neq g$. In this case,

$$\begin{aligned} &\frac{\partial^2 L_m^{reg}}{\partial\theta_{\star,k}\partial\theta_{\star,k}} E\left[x_{i,k}^2 \left\{\frac{-2m}{m-1} A_i^{-m-1} B_i^2 + \frac{m+1}{m-1} A_i^{-m} \sum_{g=1}^G C_{i,g}\right\}\right] \\ &\frac{\partial^2 L_m^{reg}}{\partial\theta_{\star,k}\partial\theta_{\star,l}} = E\left[x_{i,k} x_{i,l} \left\{\frac{-2m}{m-1} A_i^{-m-1} B_i^2 + \frac{m+1}{m-1} A_i^{-m} \sum_{g=1}^G C_{i,g}\right\}\right] \\ &\frac{\partial^2 L_m^{reg}}{\partial\theta_{\star,k}\partial\theta_{g,l}} = E\left[x_{i,k} x_{i,l} \left\{\frac{-2m}{m-1} A_i^{-m-1} C_{i,g} e_{i,g} B_i + \frac{m+1}{m-1} A_i^{-m} C_{i,g}\right\}\right], \end{aligned}$$

where $B_i = \sum_{g=1}^G \left[e_{i,g} C_{i,g}\right]$. By inspection, all elements of these Hessians are continuous in θ, since $e_{i,g}$, A_i^{-m}, A_i^{-m-1}, C_{ig}, B_i are continuous in θ, and all elements of H are continuous functions of these objects.

Next, we establish asymptotic normality of $\frac{1}{\sqrt{N}} \sum_{i=1}^N \eta\left(\theta, y_i, x_i\right)$. Since y_i, x_i are assumed to be jointly i.i.d., $\eta\left(\theta, y_i, x_i\right)$ is i.i.d. across observations, so by the Lindeberg-Levy

central limit theorem,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta\left(\theta^{*},y_{i},x_{i}\right)\xrightarrow{d}\mathcal{N}\left(0,V\right),$$

where $V=E\left[\eta\left(\theta^{*},y_{i},x_{i}\right)\eta\left(\theta^{*},y_{i},x_{i}\right)'\right]$ is assumed to be positive definite in Assumption 3.4.

Combining these two results with the conditions of Assumption 3, the standard conditions for asymptotic normality of a GMM estimator are satisfied (e.g., Hayashi (2011) Proposition 7.10). Since the weighting matrix is the identity matrix (the problem is just-identified),

$$\sqrt{N}\left(vec\left(\hat{\theta}\right)-vec\left(\theta^{*}\right)\right)\xrightarrow{d}\mathcal{N}\left(0,H^{-1}VH^{-1}\right).$$

□

With this final result, it is possible to conduct inference on the population solution θ^{*}, which provides the points of the distribution of coefficients. The Hessian, H, is characterized above for our empirical setting, where y is a scalar ($T=1$), with the possibility of common coefficients across groups. For the sake of generality, we also report expressions for the panel data case ($T>1$). Partial derivatives are then given by

$$\begin{aligned}
\frac{\partial^{2}L_{m}^{reg}}{\partial\theta_{g,kt}\partial\theta_{g,lt}} &= E\left[x_{i,k}\left\{-mA^{-m-1}\left(\frac{2}{m-1}\left\|e_{i,g}\right\|^{-2m/(m-1)}e_{i,g,t}x_{i,l}\right)e_{i,g,t}C_{i,g}\right.\right.\\
&\quad\left.\left.-A^{-m}x_{i,l}C_{i,g}+A^{-m}e_{i,g,t}\left(\frac{2m}{m-1}\left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}}e_{i,g,t}x_{i,l}\right)\right\}\right]\\
&= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A^{-m-1}C_{i,g}^{2}e_{i,g,t}^{2}-A^{-m}C_{i,g}+\frac{2m}{m-1}A^{-m}C_{i,g}^{\frac{2m-l}{m}}e_{i,g,t}^{2}\right\}\right]\\
\frac{\partial^{2}L_{m}^{reg}}{\partial\theta_{g,kt}\partial\theta_{g,ls}} &= E\left[x_{i,k}\left\{-mA^{-m-1}\left(\frac{2}{m-1}\left\|e_{i,g}\right\|^{-2m/(m-1)}e_{i,g,s}x_{i,l}\right)e_{i,g,t}C_{i,g}\right.\right.\\
&\quad\left.\left.+A^{-m}\times0\times C_{i,g}+A^{-m}e_{i,g,t}\left(\frac{2m}{m-1}\left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}}e_{i,g,s}x_{i,l}\right)\right\}\right]\\
&= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A^{-m-1}C_{i,g}^{2}e_{i,g,t}e_{i,g,s}+\frac{2m}{m-1}A^{-m}C_{i,g}^{\frac{2m-l}{m}}e_{i,g,t}e_{i,g,s}\right\}\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_m^{reg}}{\partial\theta_{g,kt}\partial\theta_{h,lt}} &= E\left[x_{i,k}\left\{-mA^{-m-1}\left(\frac{2}{m-1}\left\|e_{i,h}\right\|^{-2m/(m-1)}e_{i,h,t}x_{i,l}\right)e_{i,g,t}C_{i,g}\right.\right.\\
&\quad\left.\left.+A^{-m}\times 0\times C_{i,g}+A^{-m}e_{i,g,t}\times 0\right\}\right]\\
&= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A^{-m-1}C_{i,g}C_{i,h}e_{i,g,t}e_{i,h,t}\right\}\right], h\neq g\\
\frac{\partial^2 L_m^{reg}}{\partial\theta_{g,kt}\partial\theta_{h,ls}} &= E\left[x_{i,k}\left\{-mA^{-m-1}\left(\frac{2}{m-1}\left\|e_{i,h}\right\|^{-2m/(m-1)}e_{i,h,s}x_{i,l}\right)e_{i,g,t}C_{i,g}\right.\right.\\
&\quad\left.\left.+A^{-m}\times 0\times C_{i,g}+A^{-m}e_{i,g,t}\times 0\right\}\right]\\
&= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A^{-m-1}C_{i,g}C_{i,h}e_{i,g,t}e_{i,h,s}\right\}\right], h\neq g.
\end{aligned}$$

With common coefficients across groups, additional partial derivatives with respect to the common coefficients must be obtained. For this purpose, let $B_{it}=\sum_{g=1}^{G}\left[e_{i,g,t}C_{i,g}\right]$. Then the relevant derivatives are given by

$$\begin{aligned}
\frac{\partial^2 L_m^{reg}}{\partial\theta_{\star,kt}\partial\theta_{\star,lt}} &= E\left[x_{i,k}\left\{\frac{-2m}{m-1}A^{-m-1}\sum_{g=1}^{G}\left(\left\|e_{i,g}\right\|^{-2m/(m-1)}e_{i,g,t}x_{i,l}\right)B_{it}\right.\right.\\
&\quad\left.\left.+A^{-m}\sum_{g=1}^{G}\left(-x_{i,l}C_{i,g}+e_{i,g,t}\frac{2m}{m-1}\left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}}e_{i,g,t}x_{i,l}\right)\right\}\right]\\
&= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A_i^{-m-1}B_{it}^2+A_i^{-m}\sum_{g=1}^{G}\left(\frac{2m}{m-1}C_{i,g}^{\frac{2m-1}{m}}e_{i,g,t}^2-C_{i,g}\right)\right\}\right]\\
\frac{\partial^2 L_m^{reg}}{\partial\theta_{\star,kt}\partial\theta_{\star,ls}} &= E\left[x_{i,k}\left\{\frac{-2m}{m-1}A^{-m-1}\sum_{g=1}^{G}\left(\left\|e_{i,g}\right\|^{-2m/(m-1)}e_{i,g,s}x_{i,l}\right)B_{it}\right.\right.\\
&\quad\left.\left.+A^{-m}\sum_{g=1}^{G}\left(0\times C_{i,g}+e_{i,g,t}\frac{2m}{m-1}\left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}}e_{i,g,s}x_{i,l}\right)\right\}\right]\\
&= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A_i^{-m-1}B_{is}B_{it}+A_i^{-m}\sum_{g=1}^{G}\left(\frac{2m}{m-1}C_{i,g}^{\frac{2m-1}{m}}e_{i,g,s}e_{i,g,t}\right)\right\}\right]
\end{aligned}$$

$$\frac{\partial^2 L_m^{reg}}{\partial\theta_{\star,kt}\partial\theta_{g,lt}} = E\left[x_{i,k}\left\{\frac{-2m}{m-1}A^{-m-1}\left\|e_{i,g}\right\|^{-2m/(m-1)}e_{i,g,t}x_{i,l}B_{it}\right.\right.$$
$$\left.\left. + A^{-m}\left(-x_{i,l}C_{i,g} + e_{i,g,t}\frac{2m}{m-1}\left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}}e_{i,g,t}x_{i,l}\right)\right\}\right]$$
$$= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A_i^{-m-1}C_{i,g}e_{i,g,t}B_{it} + A_i^{-m}\left(\frac{2m}{m-1}C_{i,g}^{\frac{2m-1}{m}}e_{i,g,t}^2 - C_{i,g}\right)\right\}\right]$$
$$\frac{\partial^2 L_m^{reg}}{\partial\theta_{\star,kt}\partial\theta_{g,ls}} = E\left[x_{i,k}\left\{\frac{-2m}{m-1}A^{-m-1}\left\|e_{i,g}\right\|^{-2m/(m-1)}e_{i,g,s}x_{i,l}B_{it}\right.\right.$$
$$\left.\left. + A^{-m}\left(0\times C_{i,g} + e_{i,g,t}\frac{2m}{m-1}\left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}}e_{i,g,s}x_{i,l}\right)\right\}\right]$$
$$= E\left[x_{i,k}x_{i,l}\left\{\frac{-2m}{m-1}A_i^{-m-1}C_{i,g}e_{i,g,s}B_{it} + A_i^{-m}\sum_{g=1}^{G}\left(\frac{2m}{m-1}C_{i,g}^{\frac{2m-1}{m}}e_{i,g,s}e_{i,g,t}\right)\right\}\right].$$

A.3 Properties of FCP-IV

In this section, we replicate the results of the FCP implementation of the OLS regression in the previous section for FCP-IV. For the purposes of this development, we assume cross-sectional data, $T = 1$. Let x^e denote the endogenous regressors of interest, and let ω denote additional controls. Consider a homogeneous first-stage regression

$$x_i^e = \gamma' z_i + \tau'\omega_i + u_i, \tag{17}$$

for K_z instruments z with $K_z \geq K_e$, where K_e is the number of endogenous regressors x^e. We henceforth consider the just-identified case with a single endogenous regressor, $K_z = K_e = 1$, for economy of notation and our empirical setting, but the results can be trivially extended to allow for additional dimensions, overidentification, and an arbitrary weight matrix W in the first-stage. We also assume that x^e has heterogeneous coefficients. Denote $\tilde{x}_i^e = \gamma z_i + \tau'\omega_i$, the predicted first-stage values, and $\tilde{x}_i = \begin{pmatrix} \tilde{x}_i^e & \omega' \end{pmatrix}'$, the vector of predicted endogenous regressors and exogenous controls (so $\tilde{x}^e$ is ordered first). Using these values, we define the FCP second-stage as

$$J_m^{IV} = E\left[\sum_{g=1}^{G}\mu_g^{IV,m}\left(y_i \mid \tilde{x}_i; \theta^{IV}\right)\left\|y_i - \theta_g^{IV'}\tilde{x}_i\right\|^2\right],$$

where

$$\mu_g^{IV}\left(y_i \mid \tilde{x}_i; \theta^{IV}\right) = \left(\sum_{h=1}^{G} \frac{\left\|y_i - \theta_g^{IV\prime}\tilde{x}_i\right\|^{2/(m-1)}}{\left\|y_i - \theta_h^{IV\prime}\tilde{x}_i\right\|^{2/(m-1)}}\right)^{-1}, g = 1, \ldots, G.$$

Assumption 4. *(Two-stage least squares)*

1. $\left(y_i, x_i^e, w_i, z_i\right)$ *are i.i.d. with probability measure* $\tilde{\Pi}$ *and* $E\left[\varepsilon_i z_i \mid i \in g\right] = 0, \forall g = 1, \ldots, G,$
2. *The second moments of* $y, \tilde{x}, x^e, \omega,$ *and* z *are finite under* $\tilde{\Pi}$*:*

$$E\left[y_i^2\right] < \infty,\ E\left[\tilde{x}_i^2\right] < \infty,\ E\left[\tilde{x}_i y_i\right] < \infty, E\left[x_i^e\right] < \infty,$$
$$E\left[\begin{pmatrix} z_i \\ \omega_i \end{pmatrix}\begin{pmatrix} z_i \\ \omega_i \end{pmatrix}'\right] < \infty, E\left[\begin{pmatrix} z_i \\ \omega_i \end{pmatrix} x_i^e\right] < \infty,$$

3. *Additionally, neither* $\tilde{x}$ *nor* $\begin{pmatrix} z \\ \omega \end{pmatrix}$ *is collinear,*

$$rank\left(E\left[\tilde{x}_i \tilde{x}_i'\right]\right) = K,$$
$$rank\left(E\left[\begin{pmatrix} z_i \\ \omega_i \end{pmatrix}\begin{pmatrix} z_i \\ \omega_i \end{pmatrix}'\right]\right) = K.$$

Assumption 4 stipulates TSLS assumptions, extended to grouped regression, and regularity conditions on the data. The relevance condition is incorporated in point 3.

Corollary 1. *Define* $L_m^{IV} = E\left[\left(\sum_{g=1}^{G}\left\|y - \theta_g^{IV}\tilde{x}\right\|^{-2/(m-1)}\right)^{1-m}\right]$. *If* γ *and* τ *are known and Assumption 4 holds, then*

1. L_m^{IV} *is equivalent to* J_m^{IV},
2. *There exists a solution to* L_m^{IV}, $\theta^{*,IV}$.

Proof. The results follow immediately from Propositions 3 and 4, simply replacing x with $\tilde{x}$. □

Corollary 1 establishes the existence of a solution to the FCP problem, $\theta^{*,IV}$, which minimizes L_m^{IV}. L_m^{IV} is identical to the regression objective function, just evaluated for $\tilde{x}$ instead of x. Thus, it has first-order-conditions given by $\rho\left(\theta^{IV}, y_i, \tilde{x}_i\right)$. Let $\kappa\left(\gamma, \tau, x_i^e, z_i, \omega_i\right) =$

$E\left[\begin{pmatrix} z_i \\ \omega_i \end{pmatrix}\left(x_i^e - \gamma z_i - \tau\omega_i\right)\right]$, the standard OLS moment conditions corresponding to Equation (17) and let $\xi\left(\theta^{IV}, \gamma, \tau, y_i, x_i^e, z_i, \omega_i\right) \equiv \xi\left(\theta^{IV}, \gamma, \tau, \cdot\right)$ be the stacked vector of moment equations $\begin{pmatrix} \kappa\left(\gamma, \tau, x_i^e, z_i, \omega_i\right)', & \rho\left(\theta^{IV}, y_{i,}, \tilde{x}_i\right)' \end{pmatrix}'$. Denote the parameter vector combining both first and second stage coefficients as $v = \begin{pmatrix} vec\left(\theta^{IV}\right)', & \gamma', & \tau', \end{pmatrix}'$. Define $\hat{v}$ as the estimated parameter vector solving the sample analogue of $\xi\left(\theta^{IV}, \gamma, \tau, \cdot\right)$. As in the regression model, these moments constitute the basis for a GMM interpretation of the FCP-IV model. Some additional assumptions are needed to characterize the asymptotic properties of $\hat{v}$. Denote as v^* the vector of true parameters from the first stage, γ_0, τ_0, and $\theta^{*,IV}$.

Assumption 5. *(Consistency and asymptotic normality) Additionally,*

1. *v is in the interior of* Y*;* Y *is compact.*
2. *$\theta^{*,IV}$ is unique,*
3. *$\Gamma = E\left[\frac{\partial \xi(v^*, \cdot)}{\partial v'}\right]$ is full rank,*
4. *$E\left[\sup_{v \in \mathcal{N}} \left\| \frac{\partial \xi(v, \cdot)}{\partial v'} \right\|\right] < \infty$ in a neighborhood $\mathcal{N}$ of v^*,*
5. *$E\left[\xi\left(v, \cdot\right) \xi\left(v, \cdot\right)'\right]$ is positive definite.*

Theorem 3. *Under Assumptions 4-3,*

1. *$\hat{v}$, is consistent for v^*,*
2. *$\sqrt{N}\left(\hat{v} - v^*\right) \xrightarrow{d} \mathcal{N}\left(0, \Gamma^{-1} V^{IV} \Gamma'^{-1}\right)$, where*

$$V^{IV} = E\left[\xi\left(v, \cdot\right) \xi\left(v, \cdot\right)'\right].$$

Proof. The proof largely follows from those of Theorems 1 and 2. For consistency, first note that γ_0 and τ_0 are unique solutions to the set of first-stage moments $\kappa\left(\gamma, \tau, x_i^e, z_i, \omega_i\right) = E\left[\begin{pmatrix} z_i \\ \omega_i \end{pmatrix}\left(x_i^e - \gamma z_i - \tau\omega_i\right)\right]$ corresponding to the OLS problem (17) by Assumption 4. v^* is thus the unique solution to the stacked moments $\xi\left(v\right)$. Since the second stage is just-identified (so the corresponding moments can always be set to zero at the optimum $\theta^{*,IV}$ regardless of the value the first stage parameters take), the second stage has no influence on the first stage coefficients. It is immediate that $\kappa\left(v\right)$ is continuous in v; as noted in the proof of Proposition 5, the moment conditions $\eta\left(\theta, y_i, x_i\right)$ are continuous for all

$\theta \in \Theta$. Since $\tilde{x}$ is continuous in γ, τ, this means that $\xi\left(\theta^{IV}, y_i, \tilde{x}_i\left(\gamma, \tau\right)\right)$ is continuous in v. The moments are bounded in expectation for all $v \in \mathrm{Y}$ by duplicating the argument in the proof of Theorem 1 for the second stage under Assumption 4 and observing that the boundedness of the first-stage moments follows immediately from Assumptions 4 and 5. These points jointly satisfy the requirements of standard GMM arguments, (e.g., Newey and McFadden (1994), p. 2121—2, Hayashi (2011) Proposition 7.7), so $\hat{v} \xrightarrow{p} v^*$.

For asymptotic normality, we first provide expressions for Γ to establish the continuous differentiability of $\xi\left(v\right)$ in v. Partition the blocks of Γ as

$$\Gamma = \begin{bmatrix} \Gamma_{\gamma\gamma} & \Gamma_{\gamma\tau} & \Gamma_{\gamma 1} & \cdots & \Gamma_{\gamma g} & \cdots & \Gamma_{\gamma G} \\ \Gamma_{\tau\gamma} & \Gamma_{\tau\tau} & \Gamma_{\tau 1} & \cdots & \Gamma_{\tau g} & \cdots & \Gamma_{\tau G} \\ \Gamma_{1\gamma} & \Gamma_{1\tau} & \Gamma_{11} & \cdots & \Gamma_{1g} & \cdots & \Gamma_{1G} \\ \vdots & \vdots & \vdots & \ddots & & & \vdots \\ \Gamma_{g\gamma} & \Gamma_{g\tau} & \Gamma_{g1} & & \Gamma_{gg} & & \Gamma_{gG} \\ \vdots & \vdots & \vdots & & & \ddots & \vdots \\ \Gamma_{G\gamma} & \Gamma_{G\tau} & \Gamma_{G1} & \cdots & \Gamma_{Gg} & \cdots & \Gamma_{GG} \end{bmatrix},$$

Note that since the moment conditions are no longer derived as the gradient of a single objective function, Γ is no longer a Hessian, and no longer symmetric, in particular the blocks linking the first and second stages. For the case where all second stage coefficients are group-specific,

$$\begin{aligned}
\Gamma_{\gamma\gamma} &= E\left[z_i^2\right] \\
\Gamma_{\tau\gamma} &= E\left[\omega_i z_i\right] \\
\Gamma_{\tau\tau} &= E\left[\omega_i \omega_i^{'}\right] \\
\Gamma_{\gamma g} &= 0 \\
\Gamma_{\tau g} &= 0 \\
\Gamma_{gg} &= E\left[\tilde{x}_i \tilde{x}_i' \left\{ \frac{-2m}{m-1} A_i^{-m-1} \left(e_{i,g}\right)^2 C_{i,g}^2 + \frac{m+1}{m-1} A_i^{-m} C_{i,g} \right\}\right] \\
\Gamma_{gh} &= E\left[\tilde{x}_i \tilde{x}_i' \left\{ \frac{-2m}{m-1} A_i^{-m-1} C_{i,h} e_{i,h} e_{i,g} C_{i,g} \right\}\right], h \neq g,
\end{aligned}$$

with the elements of $\Gamma_{g\gamma}$ and $\Gamma_{g\tau}$ given by

$$\begin{aligned}
\frac{\partial^2 L_m^{IV}}{\partial \theta_{g,1}^{IV} \partial \gamma} &= E\left[\left\{\frac{-2m}{m-1} A^{-m-1} \sum_{h=1}^{G} \left(\left\|e_{i,h}\right\|^{-2m/(m-1)} e_{i,h}\theta_{h,1}^{IV}\right) C_{i,g} e_{i,g} \tilde{x}_i^e \right.\right. \\
&\quad \left.\left. + A^{-m} \frac{2m}{m-1} \left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}} e_{i,g}\theta_{g,1}^{IV} e_{i,g}\tilde{x}_i^e - A^{-m} C_{i,g}\theta_{g,1}^{IV}\tilde{x}_i^e + A^{-m} C_{i,g} e_{i,g}\right\} z_i\right] \\
&= E\left[A^{-m} C_{i,g}\left\{\frac{-2m}{m-1} A^{-1} \sum_{h=1}^{G}\left(C_{i,h} e_{i,h}\theta_{h,1}^{IV}\right) e_{i,g}\tilde{x}_i^e + \frac{m+1}{m-1}\theta_{g,1}^{IV}\tilde{x}_i^e + e_{i,g}\right\} z_i\right] \\
\frac{\partial^2 L_m^{IV}}{\partial \theta_{g,k}^{IV} \partial \gamma} &= E\left[\left\{\frac{-2m}{m-1} A^{-m-1} \sum_{h=1}^{G} \left(\left\|e_{i,h}\right\|^{-2m/(m-1)} e_{i,h}\theta_{h,1}^{IV}\right) C_{i,g} e_{i,g} \tilde{x}_{ik} \right.\right. \\
&\quad \left.\left. + A^{-m} \frac{2m}{m-1} \left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}} e_{i,g}\theta_{g,1}^{IV} e_{i,g}\tilde{x}_{ik} - A^{-m} C_{i,g}\theta_{g,1}^{IV}\tilde{x}_{ik}\right\} z_i\right] \\
&= E\left[A^{-m} C_{i,g}\tilde{x}_{ik}\left\{\frac{-2m}{m-1} A^{-1} \sum_{h=1}^{G}\left(C_{i,h} e_{i,h}\theta_{h,1}^{IV}\right) e_{i,g} + \frac{m+1}{m-1}\theta_{g,1}^{IV}\right\} z_i\right], k > 1 \\
\frac{\partial^2 L_m^{IV}}{\partial \theta_{g,1}^{IV} \partial \tau'} &= E\left[\left\{\frac{-2m}{m-1} A^{-m-1} \sum_{h=1}^{G} \left(\left\|e_{i,h}\right\|^{-2m/(m-1)} e_{i,h}\theta_{h,1}^{IV}\right) C_{i,g} e_{i,g} \tilde{x}_i^e \right.\right. \\
&\quad \left.\left. + A^{-m} \frac{2m}{m-1} \left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}} e_{i,g}\theta_{g,1}^{IV} e_{i,g}\tilde{x}_i^e - A^{-m} C_{i,g}\theta_{g,1}^{IV}\tilde{x}_i^e + A^{-m} C_{i,g} e_{i,g}\right\} \omega_i'\right] \\
&= E\left[A^{-m} C_{i,g}\left\{\frac{-2m}{m-1} A^{-1} \sum_{h=1}^{G}\left(C_{i,h} e_{i,h}\theta_{h,1}^{IV}\right) e_{i,g}\tilde{x}_i^e + \frac{m+1}{m-1}\theta_{g,1}^{IV}\tilde{x}_i^e + e_{i,g}\right\} \omega_i'\right] \\
\frac{\partial^2 L_m^{IV}}{\partial \theta_{g,k}^{IV} \partial \tau'} &= E\left[\left\{\frac{-2m}{m-1} A^{-m-1} \sum_{h=1}^{G} \left(\left\|e_{i,h}\right\|^{-2m/(m-1)} e_{i,h}\theta_{h,1}^{IV}\right) C_{i,g} e_{i,g} \tilde{x}_{ik} \right.\right. \\
&\quad \left.\left. + A^{-m} \frac{2m}{m-1} \left\|e_{i,g}\right\|^{\frac{-4m+2}{m-1}} e_{i,g}\theta_{g,1}^{IV} e_{i,g}\tilde{x}_{ik} - A^{-m} C_{i,g}\theta_{g,1}^{IV}\tilde{x}_{ik}\right\} \omega_i'\right] \\
&= E\left[A^{-m} C_{i,g}\tilde{x}_{ik}\left\{\frac{-2m}{m-1} A^{-1} \sum_{h=1}^{G}\left(C_{i,h} e_{i,h}\theta_{h,1}^{IV}\right) e_{i,g} + \frac{m+1}{m-1}\theta_{g,1}^{IV}\right\} \omega_i'\right], k > 1
\end{aligned}$$

where $e_{i,g} = y_i - \theta_g^{IV}\tilde{x}_i$, $A_i = \Sigma_{g=1}^{G}\left\|e_{i,g}\right\|^{-2/(m-1)}$, $C_{ig} = \left\|e_{i,g}\right\|^{-2m/(m-1)}$ and we have exploited the fact that $\left\|e_{i,g}\right\|^2 = e_{i,g}^2$ since $T = 1$. We also provide expressions for elements of Γ that change when there are controls in ω_k with common coefficients across groups,

such that $\theta^{IV}_{g,k} = \theta^{IV}_{h,k} \equiv \theta^{IV}_{\star,k}, h \neq g$. In this case,

$$\begin{aligned}
\frac{\partial^2 L^{IV}_m}{\partial \theta^{IV}_{\star,k} \partial \theta^{IV}_{\star,k}} &= E\left[\tilde{x}^2_{i,k} A_i^{-m}\left\{\frac{-2m}{m-1} A_i^{-1} B_i^2 + \frac{m+1}{m-1}\sum_{g=1}^{G} C_{i,g}\right\}\right] \\
\frac{\partial^2 L^{IV}_m}{\partial \theta^{IV}_{\star,k} \partial \theta^{IV}_{\star,l}} &= E\left[\tilde{x}_{i,k}\tilde{x}_{i,l} A_i^{-m}\left\{\frac{-2m}{m-1} A_i^{-1} B_i^2 + \frac{m+1}{m-1}\sum_{g=1}^{G} C_{i,g}\right\}\right] \\
\frac{\partial^2 L^{IV}_m}{\partial \theta^{IV}_{\star,k} \partial \theta^{IV}_{g,l}} &= E\left[\tilde{x}_{i,k}\tilde{x}_{i,l} A_i^{-m}\left\{\frac{-2m}{m-1} A_i^{-1} C_{i,g} e_{i,g} B_i + \frac{m+1}{m-1} C_{i,g}\right\}\right], \\
\frac{\partial^2 L^{IV}_m}{\partial \theta^{IV}_{\star,k} \partial \gamma} &= E\left[\tilde{x}_{i,k} z_i A_i^{-m}\left\{\frac{-2m}{m-1} A_i^{-1}\sum_{g=1}^{G}\left(C_{i,g} e_{i,g} \theta^{IV}_{g,1}\right) B_i + \frac{m+1}{m-1}\sum_{g=1}^{G} C_{i,g}\theta^{IV}_{g,1}\right\}\right] \\
\frac{\partial^2 L^{IV}_m}{\partial \theta^{IV}_{\star,k} \partial \tau'} &= E\left[\tilde{x}_{i,k} A_i^{-m}\left\{\frac{-2m}{m-1} A_i^{-1}\sum_{g=1}^{G}\left(C_{i,g} e_{i,g} \theta^{IV}_{g,1}\right) B_i + \frac{m+1}{m-1}\sum_{g=1}^{G} C_{i,g}\theta^{IV}_{g,1}\right\}\omega_i'\right]
\end{aligned}$$

where $B_i = \sum_{g=1}^{G}\left[e_{i,g} C_{i,g}\right]$. By inspection, Γ is continuous in v, since $e_{i,g}$, A_i^{-m}, A_i^{-m-1}, C_{ig}, B_i are continuous in θ^{IV}, and all elements of Γ are continuous functions of these objects.

Next, we establish the asymptotic normality of $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi\left(v\right)$. Since $y_i, x_i^e, \omega_i, z_i$ are assumed to be jointly i.i.d., $\xi\left(v, y_i, x_i^e, \omega_i, z_i\right)$ is i.i.d. across observations, so by the Lindeberg-Levy central limit theorem,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi\left(v^*, \cdot\right) \xrightarrow{d} \mathcal{N}\left(0, V^{IV}\right),$$

where $V^{IV} = E\left[\xi\left(v^*, \cdot\right)\xi\left(v^*, \cdot\right)'\right]$ is assumed to be positive definite in Assumption 5.5.

Combining these two results with the additional conditions of Assumption 5, the standard conditions for asymptotic normality of a GMM estimator are satisfied (e.g.,Hayashi (2011) Proposition 7.10). Since the weighting matrix is the identity (we assumed the problem is just-identified),

$$\sqrt{N}\left(\hat{v} - v^*\right) \xrightarrow{d} \mathcal{N}\left(0, \Gamma^{-1} V^{IV} \Gamma'^{-1}\right).$$

□

A.4 Relationship between FCP and finite mixture models

In the main text, we note that FCP is closely related to finite mixture models, including Gaussian mixture models or Bayesian mixture models, for example. Such models have

previously been used to characterize latent heterogeneity in economic relationships, for example by Boz et al. (2019). Bonhomme and Manresa (2015) (Section 2.1 and Appendix S5) discuss this point for HKM, and show formal equivalence between HKM and finite mixture models, where the weights in the mixture model take the form of the binary HKM weights. The same is true for FCP. The objective function $J_m^{reg}(\Pi, \theta)$, (Equation (12)) has precisely the form of the M-step of a Gaussian mixture model, where the weights would instead be derived in the E-step based on the likelihood. Thus, the distinction between FCP and conventional mixture models can be seen as a choice of weights.

When the likelihood is known, as well as any relationship between heterogeneity and observables, mixture models will identify the true parameters even in the face of uncertain group membership (and Cramer-Rao efficiency results apply). If the true likelihood is unknown, such approaches will only recover pseudo-true parameters, much like HKM and FCP. In this case, the econometrician has to choose between alternative sets of misspecified weights. However, Bezdek (1973) shows that the FCP weights are the optimal choice based on the formulation in Equation (4), given no knowledge of the likelihood. Thus, when the likelihood is known, there is a strong argument to use a mixture model based upon it, but when the likelihood is unknown, there are firm theoretical grounds to opt for FCP instead.

B Specification details

In this section, we discuss two considerations underlying our chosen specification, (2). First, we present details on the gap statistic, which we use to select the number of groups, G. Second, we describe the role that additional observable controls play in our model.

B.1 Selecting G using the Gap Statistic

For a given G, it is easy to apply the FCP algorithm as described above. However, G is unknown. To choose G, we extend the "gap statistic" of Tibshirani et al. (2001) to the regression setting. In the Tibshirani et al. (2001) setting, the researcher has data on some characteristic for observation i, y_i, for each $i = 1, \ldots, N$. They define the within-group residual sum of squares $W_{ss}(G)$ under G groups as:

$$W_{ss}(G) = \sum_{g \in G} \frac{1}{2N_g} \sum_{i',i \in g} d_{i'i}$$

where $d_{i'i}$ is the Euclidean distance between observations i and i' and N_g is the number of observations belonging to cluster g. The gap statistic identifies the number of clusters by comparing the average within group sum of squares under G groups ($W_{ss}(G)$) to the *expected* within group sum of squares under G groups under the assumption that no clustering is present in the data (the "reference distribution"). Formally, the gap statistic is:

$$\text{Gap}(G) = E_{ref}\left[log\left(W_{SS}(G)\right)\right] - log\left(\bar{W_{SS}}(G)\right) \tag{18}$$

The reference distribution is obtained by sampling the outcomes from a uniform distribution on an interval $[a, b]$ where a and b are chosen as the maximum and minimum observed values of the outcomes in the sample. Ideally, G is chosen such that the gap statistic is maximized. The basic idea is to normalize the within group sum of squares ($W_{SS}(G)$) curve by what one would expect to obtain for the $W_{SS}(G)$ in a sample in which no clusters are present, and any improvement in the $W_{SS}(G)$ as G increases is simply due to fitting noise. In practice, Tibshirani et al. (2001) find that the gap statistic may display local maxima when the data are not well-separated or sub-clusters are present, and it is advisable to inspect the "gap curve" as opposed to mechanically choosing its maximum. The gap statistic is also closely related to the approach described by Dunn (1977) for FCM, in which the partition entropy is normalized by its value in a reference distribution.

The objective W_{SS} above is tailored to the cluster mean setting of Tibshirani et al. (2001). We opt to use our regression objective, $J_m^{reg}(\theta(G))$ as a natural alternative: the (sample analog of) the weighted sum of squared residuals from Equation (12). The selection procedure for G then proceeds as follows:

1. Run a homogeneous regression (one that assumes no group heterogeneity). Call the residuals from the regression $\hat{\epsilon}_i$.

2. Generate B samples of simulated outcomes for each observation using the homogeneous coefficients, with an error that is uniformly distributed from the minimum to the maximum of $\hat{\epsilon}_i$.

3. Run the FCP algorithm described in Section 3.2 on each of the B samples for each $G \leq \bar{G}$, where $\bar{G}$ is some upper bound. For each sample and each G, compute the weighted sum of squared residuals, the sample analog of Equation (12).

4. Run the FCP algorithm on the actual data and compute the same object for $G \leq \bar{G}$.

5. Choose $G \leq \bar{G}$ corresponding to the maximum value of the Gap statistic from Equation (18) that is statistically significantly greater than those for $G - 1$ and $G + 1$ and

displays a positive gradient.

The final point operationalizes the advice in Tibshirani et al. (2001) to not simply choose the maximum of the gap statistic. In particular, the requirement that the chosen G represents a *statistically significant* increase over all previous estimates through $G-1$ avoids spuriously increasing G when doing so does not offer a significantly better representation of the data. We introduce the positive gradient requirement to ensure we identify important sub-clusters and to address the possible non-monotonicity noted by Tibshirani et al. (2001).[40]

Gap statistic performance

We also assess the performance of the gap statistic in selecting the correct number of groups. For this simulation, we draw 50 samples using as true structure $G = 5$ (using the same approach detailed in Section 3.3) and resampling from the empirical distribution of errors for each group to preserve its properties. We find that in all 50 samples, the gap statistic correctly selects $G = 5$ as the preferred specification. This evidence extends the original finding of Tibshirani et al. (2001) that the gap statistic performs well for mean clustering problems in our regression context.

B.2 Inclusion of controls in the objective function

In this section, we highlight the role that controls play in our model. In particular, we show that the baseline specification in Equation (19) without controls included can be seen as a non-parametric alternative to a specification with controls as the number of groups, G, increases.

Consider two models. Suppose the true model has the form

$$\Delta C_j = \beta' W_j + \sum_{g \in G} \left(\phi_g \mathbf{1}\left[j \in g\right] R_j + \alpha_g \mathbf{1}\left[j \in g\right] \right) + \epsilon_j, j = 1, \ldots, N. \tag{19}$$

Suppose the econometrician estimates a simpler model, omitting the controls:

$$\Delta C_j = \sum_{\tilde{g} \in \tilde{G}} \left(\widetilde{\phi}_{\tilde{g}} \mathbf{1}\left[j \in \tilde{g}\right] R_j + \widetilde{\alpha}_{\tilde{g}} \mathbf{1}\left[j \in \tilde{g}\right] \right) + \widetilde{\epsilon}_j, j = 1, \ldots, N. \tag{20}$$

[40] This is particularly important in our empirical setting, where we group both on intercepts and MPCs, but are interested only in the latter, since it is conceivable that the first level of clustering might only represent level heterogeneity in consumption changes. We additionally find that – in simulations in which the data are not well-separated – without this requirement the gap statistic has a tendency to erroneously favor homogeneous models.

In general, this second model could be susceptible to omitted variable bias (in particular, if $\tilde{G} = G$). However, this need not be the case if $\tilde{G}$ is allowed to vary. This is because the model in Equation (19) can be rewritten in the form of Equation (20), with $\tilde{G} \geq G$. To see this, consider a simple example where W_j is a scalar binary regressor and $G = 2$. Then if $\tilde{G} = 4$, there are four cases to consider, based on two "true" groups with heterogeneous parameters, and two levels of W_j within each group. Then the following relationships exist between $\{\phi_g, \alpha_g\}_{g=1,2}$ and $\{\tilde{\phi}_{\tilde{g}}, \tilde{\alpha}_{\tilde{g}}\}_{\tilde{g}=1,\ldots,4}$ (where the labels of groups are arbitrary):

expanded group	true group	control value	slope	intercept
$\tilde{g} = 1$	$g = 1$	$W_j = 0$	$\tilde{\phi}_1 = \phi_1$	$\tilde{\alpha}_1 = \alpha_1$
$\tilde{g} = 2$	$g = 1$	$W_j = 1$	$\tilde{\phi}_2 = \phi_1$	$\tilde{\alpha}_2 = \alpha_1 + \beta$
$\tilde{g} = 3$	$g = 2$	$W_j = 0$	$\tilde{\phi}_3 = \phi_2$	$\tilde{\alpha}_3 = \alpha_2$
$\tilde{g} = 4$	$g = 2$	$W_j = 1$	$\tilde{\phi}_4 = \phi_2$	$\tilde{\alpha}_4 = \alpha_2 + \beta$

The effect of W_j is absorbed entirely into the fixed effects $\tilde{\alpha}_{\tilde{g}}$, which now vary with an individual's W_j. The true values of ϕ_g and α_g are still recovered, as well as values of α_g shifted by β, provided $\tilde{G}$ is chosen correctly.

This argument can be extended to allow for effects of W_j other than simple level shifts. For example, if the true model has the additional interaction term $\phi_g^W \mathbf{1}[j \in g] R_j W_j$, then the $\tilde{\phi}_{\tilde{g}}$'s recovered would incorporate ϕ_g^W just like the expressions for $\tilde{\alpha}_{\tilde{g}}$ above incorporate β. The argument also extends to non-binary controls. For example, a discrete regressor taking k values would expand a $G-$group model to a $k \times G$ group model. Admittedly, extending the argument to a continuous regressor introduces a computational challenge in practice, but in our setting, available controls are generally discrete. Finally, the argument generalizes in the same way when W_j is a vector and not a simple scalar.

There remains a question over whether estimating such a model accurately recovers the heterogeneity in ϕ_g, or rather overestimates heterogeneity as G grows to $\tilde{G}$. In our view, $\tilde{G}$ more accurately represents the true heterogeneity in the underlying data, seeing as it incorporates any differences in MPCs arising from observable controls, W_j, as opposed to only residual heterogeneity after partialling out W_j. Individuals still have meaningfully different MPCs, even if that difference is explained by observable characteristics. We can then, of course, investigate the relationship between the recovered $\tilde{\phi}_{\tilde{g}}$ and W_j *ex post*, as we do in Section 4.3. We seek to characterize the full heterogeneity of MPCs, as opposed to the conditional heterogeneity of MPCs, as for instance in Kaplan et al. (2014), Fagereng et al. (2016), Johnson et al. (2006), Parker et al. (2013), and Crawley and Kuchler (2018).

A further advantage of estimating Equation (20) as opposed to Equation (19) is that it allows the relationship between W_j and ΔC_j to be completely non-parametric. Including W_j as in Equation (19) assumes the term enters linearly; including an interaction with $\mathbf{1}[j \in g] R_j$ likewise assumes a functional form. However, estimating a separate set of parameters $\tilde{\phi}_{\tilde{g}}, \tilde{\alpha}_{\tilde{g}}$ for each $\tilde{g} \in \tilde{G}$ takes no stance on the parametric structure relating W_j to C_j. On this basis, as well as the desire to recover the full degree of heterogeneity in MPCs, we proceed using specifications based on Equation (20) as our baseline. We opt to include a minimal set of covariates common to the literature, W_j, including time dummies, age, age squared, and changes in household membership, and explore the relationship between additional covariates and consumption behavior in Section 4.3.

C Simulation results

In this section, we compare the computational performance of FCP implemented via numerical minimization to an iterative approach, HKM, and an EM implementation of a Gaussian mixture model. We then document in full the simulation study sketched in the text. Finally, we provide an illustrative example of how quantile regression estimates of the MPC depend on the correlation of MPC heterogeneity with other forms of heterogeneity.

C.1 Computational tractability and details

Our approach is computationally tractable, entailing a sizable improvement compared to alternative techniques. First, we find that the numerical minimization motivated by GMM implies a reduction in the computational time to solve the algorithm relative to the iterative procedure, while also improving precision. More generally, in our setting, the FCP approach proves somewhat slower than a simple version of HKM, but considerably faster than the full HKM algorithm and than an EM implementation of the Gaussian mixture model.

The structure of our problem is such that we can apply standard non-linear minimization routines to the objective function, and make use of analytical gradients to further enhance performance. This strategy is substantially faster than an iterative procedure that sequentially optimizes weights and parameters, as is the norm for FCM. The improvements in computational speed increase with the chosen number of groups. With 10 groups, for instance, our GMM approach is more than twice as fast as its iterative

counterpart.[41]

The FCP algorithm is also faster than HKM implementations that are feasible in our setting. In the simulations shown in the following subsection, we consider two versions of the algorithms proposed by Bonhomme and Manresa (2015). "Algorithm 1", directly from that paper, starts by randomly allocating observations to a given number of groups, estimates the model, and iteratively reallocates observations while estimating the model until convergence. This strategy is faster than FCP, but has weaker performance in simulations, as we discuss below. It takes about 6 seconds to solve our FCP algorithm with 5 groups, for a single start value. Algorithm 1 takes instead about 2 seconds per start value using the same machine (MacBook Pro, 2.7Ghz Intel Core with turbo boost up to 3.8Ghz, 16GB memory). Nevertheless, we find that, even with 1,000 starting values, results are not particularly stable for Algorithm 1. For FCP, we find that 500 starting values are more than enough to obtain consistent and stable results.

Bonhomme and Manresa (2015) propose a variable neighborhood search algorithm, "Algorithm 2", which, in their framework, improves simulation performance and is often necessary to avoid local minima. In our relatively large cross-sectional dataset, this algorithm is infeasible.[42] We therefore consider an intermediate version, which we label "Algorithm 1.5". This uses the best result of Algorithm 1 (with 500 starting values) as the start value of a variable neighborhood approach, excluding the local search component, however. This algorithm repeats the assignment procedure relocating n randomly selected individuals and proceeds iteratively until the objective function stops improving. It increases n by 1 up to n_{max}, until the objective decreases. When this happens, it restarts n and repeats the procedure for b outer iterations. Jumps of size n allow escape from local solutions trapped in valleys. In the simulations shown in the following subsection, we set both n_{max} and b to 100. We also set a stopping rule of 30 iterations which exits the algorithm if 30 consecutive b iterations do not improve the objective function.[43] Within every variable neighborhood iteration, Algorithm 2 systematically checks all re-assignments of individual observations across groups, updating group assignment when the objective function decreases. This step took more than 1 week to perform for only 1 starting value in our dataset. Algorithm 1.5, instead, adds only a further 9 minutes to Algorithm 1. Its nature, however, does not allow parallelization beyond the start values in the first step.

[41] We solve the model, for one set of start values, in 17 seconds, compared with 40 seconds required to solve the model iteratively.

[42] The infeasibility stems from the fact that one step of the algorithm – local search – requires looping over all observations (~17,000 in our case) sequentially in every iteration, a step which cannot be parallelized.

[43] In unreported results we increased these parameters and find that HKM performance improves only mildly, at the expense of a dramatic increase in computational time.

Our FCP approach also improves upon an expectation-maximization (EM, Dempster et al. (1977)) implementation of a Gaussian mixture model. In the simulation results for this estimator below, we set start values by randomly assigning each individual to a group, with uniform probabilities, and calculating group-specific parameters based on this binary assignment. We then proceed through the standard EM algorithm until the log likelihood function converges. For $G = 5$, the algorithm takes over 45seconds to converge for a single start value. This computational time is much closer to that of an iterative implementation of the FCP approach, likely because both are iterative in nature; our FCP implementation via GMM is 7 times as fast as the EM implementation of the Gaussian mixture model. For each Monte Carlo sample in our simulations, we run the EM algorithm for 500 starting values, and select the solution that maximizes the log likelihood across the random starting values.

C.2 Simulation comparison of clustering approaches

To evaluate the performance of FCP in our setting, we conduct simulations calibrated to our empirical data. We simulate data according to Equation (6) using the empirical parameter estimates, taking an individual's true group membership as the group with maximal weight. We consider three specifications: $G = 5$ (our empirical baseline), with Gaussian errors ε_i (with group-specific variance); $G = 5$, with ε_i resampled from each group's empirical errors; and $G = 10$, with Gaussian errors (with group-specific variance). For each specification, we estimate the model using FCP ($m = 2.35$, our empirical choice), HKM extended to a regression framework (using Algorithm 1.5), and a Gaussian mixture model. For each estimator, we construct the mean and RMSE for each MPC, the share of observations misclassified (by maximal weight), and the median estimated CDF of MPCs, across 500 simulations. For FCP we also report rejection rates of the true parameters based on the asymptotic results provided in Theorem 2.

We start by presenting the results for $G = 5$ and group-specific Gaussian errors. Table 5 reports the results across 500 Monte Carlo samples. The top panel reports mean point estimates for each MPC for each estimator. FCP recovers values very close to the truth, on average. The values estimated by HKM are reasonably close to the DGP, although generally somewhat further from the truth than FCP. In contrast, the EM implementation of the Gaussian mixture model performs poorly, with all estimated values quite different to the true parameter values. The second panel reports the RMSE for each estimate. They are reasonably small for both FCP and HKM; the RMSE is substantially higher for those groups with relatively few observations. The Gaussian mixture model does not display

this clear pattern with some RMSEs quite large. Finally, we report the share of individuals misclassified (based on their maximal weight for FCP and the Gaussian mixture model). Again, FCP and HKM perform similarly, while the Gaussian mixture model does consistently worse.

Table 5: Simulation results, Gaussian errors, $S = 500$

	Truth	FCP	HKM	Gaussian mixture
Point Estimates	0.756	0.769	0.716	3.588
	0.382	0.342	0.344	0.256
	0.253	0.264	0.253	0.107
	0.709	0.713	0.638	-0.076
	0.566	0.560	0.549	0.615
RMSE		0.429	0.367	9.250
		0.118	0.116	0.270
		0.044	0.047	0.559
		0.097	0.121	2.908
		0.411	0.363	0.361
Share Misclassified		0.096	0.098	0.162

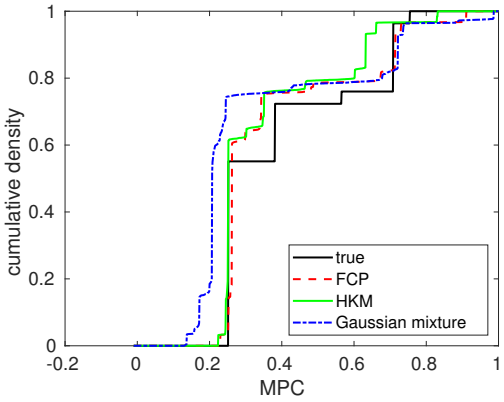
Notes: MPCs estimated on simulated data generated according to (6), $G = 5$, with group assignment based on maximal weights and Gaussian errors with group-specific variance. MPC estimates are labeled by selecting the permutation that minimizes the Euclidean distance between the estimated fixed effects and MPCs and the true values. The first panel reports the mean MPC for FCP, HKM, and an EM implementation of the Gaussian mixture model. The second panel reports the RMSE for each MPC. The final panel reports the share of observations misassigned based on maximal estimated weights. 500 Monte Carlo samples.

While these results illustrate the ability of the algorithms to recover point estimates, they do not directly demonstrate the ability of each algorithm to fit the MPC distribution. To this end, we compute the CDF of the estimated MPC distribution for each estimator for each sample (censored to the 0-1 interval), and report the median of these CDFs in Figure 8. These results show that FCP is able to recover the full distribution of MPCs remarkably well, not just point estimates. Focusing on the median CDF as opposed to mean estimates, however, does favor the Gaussian mixture model, since it minimizes the role of many large estimated outliers. The mean MPC is 0.40 in the data generating process, and is 0.38 for FCP (based on individuals' maximal weights), 0.36 for HKM, and 0.39 for the Gaussian mixture model, averaging across simulations.[44]

The second specification samples errors from the group-specific empirical distribution, which results in data that are much less well-separated. Table 6 reports parameter estimates, RMSE and misclassification across models. While no estimator precisely recovers the truth on average, FCP comes by far the closest. On average, MPC point estimates differ by 0.12 at most. In contrast, HKM comes *no closer* than 0.20 to any true MPC. Deviations are much larger still for the Gaussian mixture model. As above, the RMSE results

[44]Values are basically unchanged considering the medians instead.

Figure 8: Empirical CDFs of the MPC distribution: G = 5, Gaussian errors



Notes: 500 samples. Medians of the estimated CDF of the MPC distribution across samples, with individuals assigned based on their maximal weights.

are less consistent. In particular, FCP underperforms for the least populated groups. In terms of misclassification, FCP performs the best across models, and even better than with Gaussian errors.

Figure 9 plots the median CDF of MPCs for each estimator. The median CDF for FCP is largely aligned with the true DGP. HKM, in contrast, systematically wrongly assigns three quarters of the observations to MPCs close to 0, which is not in the support of true data generating process. The Gaussian mixture model solution instead counterfactually spreads out the MPC distribution, well beyond the censored distribution plotted below. For example, the median share of negative MPCs across simulations is 40%. The average mean MPC across simulations (based on maximal weights) is 0.41 for FCP, almost equal to the true value, 0.40. The low MPCs estimated by HKM, and the large share of negative MPC for the Gaussian mixture model, lead both models to underestimate the mean MPC at 0.10 and 0.22 respectively.

Table 7 reports results for the $G = 10$ specification with Gaussian noise. As more groups are introduced, clusters naturally become less well-separated, posing a sterner challenge to the estimators. Remarkably, even in this poorly-separated data, baseline FCP (with $m = m^*$ optimized for the $G = 5$ specification) continues to estimate the true parameters reasonably accurately. In contrast, both HKM and the Gaussian mixture model misestimate most of the MPCs. These results are further supported by the RMSE, reported in the second panel; FCP recovers the true parameters quite precisely, despite the challenges posed by the additional clusters. The misclassification rates further accentuate the advantage of FCP in this setting with additional groups.

Figure 10 plots the median CDF of MPCs for each estimator. It shows that the misestimation of some MPCs by HKM and the Gaussian mixture model is not confined to a

Table 6: Simulation results, empirical errors, $S = 500$

	Truth	FCP	HKM	Gaussian mixture
Point Estimates	0.756	0.853	0.407	0.982
	0.382	0.317	-0.040	0.197
	0.253	0.253	0.040	-0.270
	0.709	0.758	0.339	0.712
	0.566	0.685	0.383	0.707
RMSE		0.901	0.807	4.711
		0.298	0.554	0.547
		0.102	0.239	0.719
		0.179	0.397	1.397
		0.888	0.627	0.576
Share Misclassified		0.037	0.067	0.576

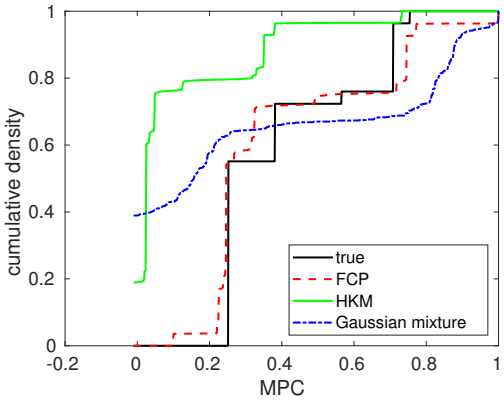
Notes: MPCs estimated on simulated data generated according to (6), $G = 5$, with group assignment based on maximal weights and resampled empirical errors. MPC estimates are labeled by selecting the permutation that minimizes the Euclidean distance between the estimated fixed effects and MPCs and the true values. The first panel reports the mean MPC for FCP, HKM, and an EM implementation of the Gaussian mixture model. The second panel reports the RMSE for each MPC. The final panel reports the share of observations misassigned based on maximal estimated weights. 500 Monte Carlo samples.

small share of observations. In contrast, FCP recovers the true distribution of MPCs very closely, on average. The Gaussian mixture model wrongly estimates a non-negligible share of observations with MPC above 1, while HKM estimates more than 10% of MPCs to be negative. The average mean MPC is close to the truth for all estimators, but this reflects these spurious tails cancelling out for HKM and the Gaussian mixture model. However, those estimators fail to match other moments of the distribution, illustrated in particular in the right panel, which reports the non-censored CDFs.

Finally, Table 8 reports the size of nominal 5% tests for the the MPCs estimated using FCP for all three specifications Size is very good for the $G = 5$ model with Gaussian noise; the largest distortion is about 2%. With empirical noise, some larger distortions (up to 7%) appear, but these are associated with the parameters for which the point estimates exhibit appreciable bias. Finally, for the $G = 10$ specification, the tests are again well-sized for most MPCs, with a larger deviation from the nominal size for a single MPC.

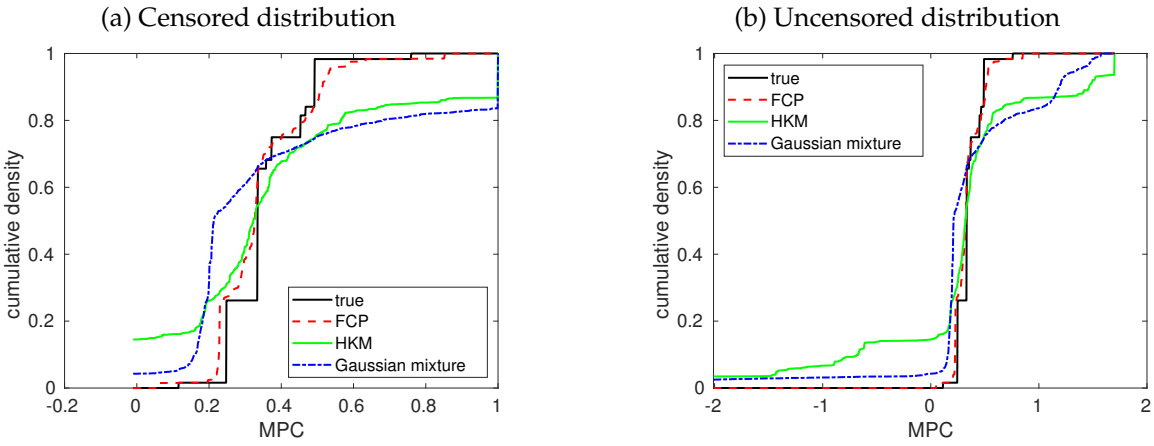
In summary, both FCP and HKM perform quite well for a small number of groups and Gaussian noise, in terms of the mean and RMSE of point estimates, misclassification of individuals and recovering the distribution of MPCs. For a small number of groups and empirical noise, FCP demonstrates a clear advantage in our empirical setting in terms of the mean of point estimates and recovering the distribution of MPCs. For a large number of groups, the gap in performance between FCP and the other estimators is sizable for all metrics. The limiting distribution provided in Theorem 2 provides a good approximation, yielding tests that are generally well-sized, suggesting that inference on the recovered

Figure 9: Empirical CDFs of the MPC distribution: G = 5, empirical errors



Notes: 500 samples. Medians of the estimated CDF of the MPC distribution across samples, with individuals assigned based on their maximal weights.

Figure 10: Empirical CDFs of the MPC distribution: G = 10, Gaussian errors



Notes: 500 samples. Medians of the estimated CDF of the MPC distribution across samples, with individuals assigned based on their maximal weights. The left panel censors the distributions on the interval $[-0.2, 1]$; the right panel reports the uncensored distributions.

MPC distribution is reliable.

C.3 The role of correlated heterogeneity in quantile regression

In this section, we illustrate the challenges faced when using quantile regression to recover the MPC distribution in two ways. First, we present an intuitive example to illustrate how the recovered MPC distribution is impacted by heterogeneity in additional parameters. Next, we use quantile regression to estimate the MPC distribution in the full simulation setting described in the preceding section. We find that quantile regression systematically estimates a spurious left tail for the MPC distribution, with a considerable mass of households having MPCs below the lowest true MPC and even below zero.

Table 7: Simulation results, Gaussian errors, $G^* = 10$, $S = 500$

	Truth	FCP	HKM	Gaussian mixture
Point Estimates	0.761	0.780	1.649	-1.945
	0.359	0.387	0.544	0.629
	0.454	0.464	-0.375	0.260
	0.336	0.313	0.224	0.361
	0.249	0.231	0.211	0.205
	0.335	0.330	0.444	0.730
	0.493	0.502	0.908	-0.961
	0.373	0.425	0.414	0.603
	0.467	0.401	0.346	0.602
	0.117	0.168	0.165	0.165
RMSE		0.427	9.519	10.918
		0.236	0.545	0.418
		0.176	2.282	1.004
		0.067	0.658	0.253
		0.039	0.575	0.282
		0.031	1.020	0.904
		0.050	3.124	3.276
		0.146	0.770	0.371
		0.391	3.258	0.969
		0.163	0.199	0.155
Share Misclassified		0.138	0.454	0.719

Notes: MPCs estimated on simulated data generated according to (6), $G = 10$, with group assignment based on maximal weights and Gaussian errors with group-specific variance. MPC estimates are labeled by selecting the permutation that minimizes the Euclidean distance between the estimated fixed effects and MPCs and the true values. The first panel reports the mean MPC for FCP, HKM, and an EM implementation of the Gaussian mixture model. The second panel reports the RMSE for each MPC. The final panel reports the share of observations misassigned based on maximal estimated weights. 500 Monte Carlo samples.

Quantile regression with multiple dimensions of heterogeneity

In this section, we provide an intuitive example of how the heterogeneity in MPCs recovered by quantile regression depends on the correlation of MPC heterogeneity with other forms of heterogeneity. We consider a simple setting, where there are two possible fixed effect values, $\alpha_j \in \{-10,000, 10,000\}$ (the order of magnitude of our estimated fixed effects), and two MPCs, $\phi_j = \{0.20, 0.70\}$ (approximately the high and low points of our baseline distribution). We draw non-zero rebate values $R_j \sim N\left(900, 100^2\right)$, centered at the median in our data. We then generate data according to

$$\Delta C_j = \alpha_j + \phi_j R_j + \epsilon_j, j = 1, \ldots, N,$$

where $\epsilon_j \sim N\left(0, 1000^2\right)$, somewhat lower than the estimated noise in the data. We set $N = 100,000$, with 17.5% of observations receiving a rebate, as in our data, with $R_j = 0$

Table 8: Size of tests for FCP parameter estimates

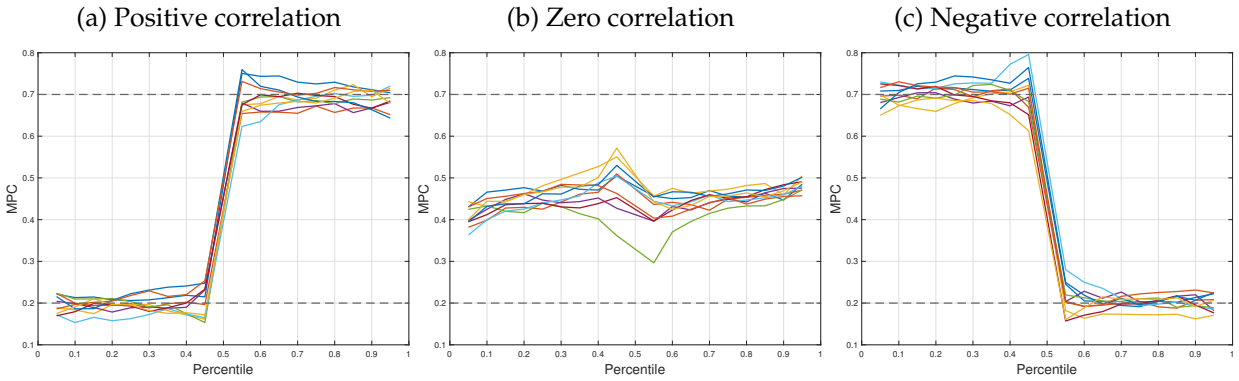
MPC	0.756	0.382	0.253	0.709	0.566					
$G = 5$, Gaussian	4.8	5.4	4.4	5.4	7.6					
$G = 5$, empirical	10.2	8.4	4.6	3.6	12.0					
MPC	0.761	0.359	0.454	0.336	0.249	0.335	0.493	0.373	0.468	0.117
$G = 10$, Gaussian	11.4	3.4	8.2	7.6	7.4	4.8	5.6	6.4	17.2	8.2

Notes: Rejection rates of nominal 5% tests of true MPC values based on FCP estimates using the asymptotic variance derived in Theorem 2 for each of the three simulation specifications, 500 Monte Carlo samples.

for the others.

We assume α_j and ϕ_j take each value with 50% probability. We consider three possible relationships between α_j and ϕ_j. First, we assume that they are perfectly positively correlated, so $(\alpha_j, \phi_j) \in \{(-10,000, 0.20), (10,000, 0.70)\}$, with equal probabilities. Next, we assume that fixed effects and MPCs have zero correlation. Thus, $(\alpha_j, \phi_j) \in \{(-10,000, 0.20), (10,000, 0.70), (-10,000, 0.70), (10,000, 0.20)\}$, with equal probabilities. Finally, we assume that they are perfectly negatively correlated, so $(\alpha_j, \phi_j) \in \{(10,000, 0.20), (-10,000,$ with equal probabilities. For each specification, we draw 10 samples, estimate the model using quantile regression for every fifth percentile, and plot the estimated MPC distributions in Figure 11. The first panel shows that when the fixed effects and MPCs are positively correlated, the MPCs are well estimated; half of the distribution is associated with an MPC around 0.20, and half with an MPC around 0.70. Because the fixed effects dominate the conditional distribution, and the MPCs are correlated with the fixed effects, the lower MPC aligns with the lower half of the distribution. In the second panel, there is zero correlation between fixed effects and MPCs. Since the percentile of the distribution to which each observation corresponds is driven largely by the fixed effect, the two MPCs occur with approximately equal frequency at each percentile, so a value near the average MPC is estimated *at each percentile*. Finally, the third panel shows that when fixed effects and MPCs are negatively correlated, the MPCs are again well-estimated, as in the first panel. However, this time the high MPC corresponds to the lower half of the distribution, since it aligns with the lower fixed effect. These results show that if there is heterogeneity in other parameters besides the MPC, the relationship between such heterogeneity and the MPC will impact the econometrician's ability to recover the distribution of MPCs using quantile regression.

Figure 11: The role of correlated heterogeneity in quantile regression



Notes: Figure 11 plots the estimated MPCs from quantile regression for every fifth percentile for 10 samples of simulated data for three specifications. In each specification, both fixed effects and MPCs take two possible values. In the first panel, fixed effects and MPCs are perfectly positive correlated, in the second they have zero correlation, and in the third they are perfectly negatively correlated. The dashed lines represent the two true MPC values.

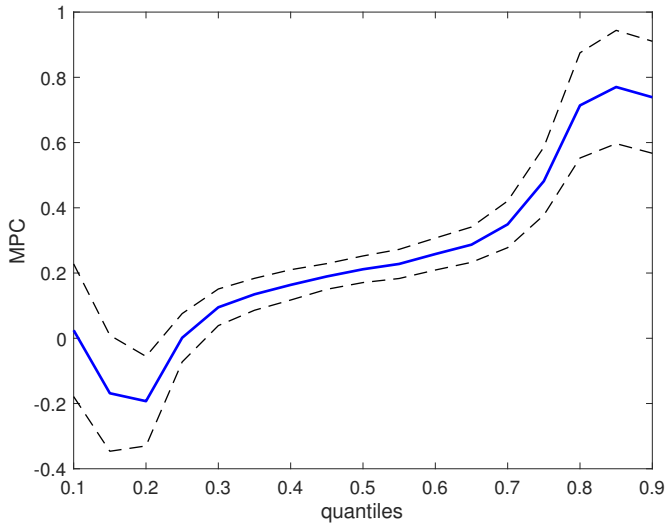
Simulation results for quantile regression

The preceding example illustrates the potential pitfalls of quantile regression in a highly simplified setting. We now simulate data from our full empirically-calibrated specifications described in Section C.2 and estimate the distribution of MPCs using quantile regression, at 5-percentile increments from 0.1 to 0.9, as in Misra and Surico (2014). Since quantile regression does not compute group-specific parameters, but rather parameters at given quantiles, we present results differently. To summarize the distribution across 500 samples, we compute the mean MPC at each quantile and plot it in Figure 12, along with 68% confidence intervals (for the $G = 5$ Gaussian errors case, but results are similar for other specifications). The lowest MPC used to generate the data is 0.25, but quantile regression estimates a large mass of MPCs near or below zero, about 25%. In fact, only for the 60th percentile and beyond is the average MPC (across samples) above 0.25, the true lowest MPC. Quantile regression thus estimates a large spurious left tail for the MPC distribution in this setting. In contrast, the right tail of the distribution matches the DGP quite well, with the average MPC at the highest quantiles close to the true highest MPC, 0.76.

We can also present the results in the form of a CDF, as we did for clustering approaches in Section C.2. Figure 13 reports the true CDF, along with the median CDF estimated across samples for both quantile regression and FCP. The results mirror those of the previous section and of Figure 12: FCP traces the true CDF quite well, on average, while quantile regression exhibits a large left tail, with considerable mass below zero, and over 50% of the distribution below the true lowest MPC.

Finally, we also assess the extent to which quantile regression miss-assigns households

Figure 12: Quantiles of the MPC distribution



Notes: 500 samples. Average MPC at each quantile from 0.1 to 0.9 at 5 percentile intervals, with 68% confidence intervals computed with the standard deviation of the quantile-specific MPC across samples. $G = 5$ and Gaussian errors.

Table 9: Distribution of absolute errors of estimated MPCs

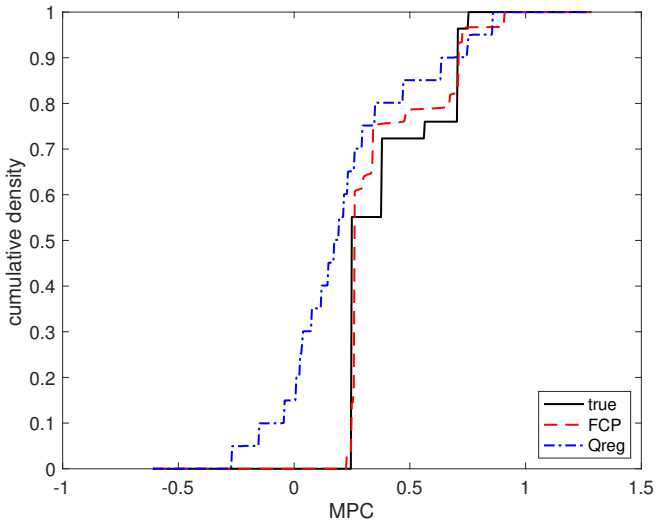
	average	p10	p25	p50	p75	p90
FCP	0.10	0.02	0.03	0.04	0.11	0.22
EM	0.28	0.04	0.06	0.07	0.20	0.35
HKM	0.10	0.03	0.04	0.04	0.12	0.22
QR	0.27	0.03	0.07	0.18	0.45	0.71

Notes: 500 samples. Average quantiles of the absolute error in households' estimated MPCs across samples. $G = 5$ and Gaussian errors.

to MPCs, since this will impact subsequent attempts to recover the determinants of the MPC. Table 9 presents the results for quantile regression, as well as all three clustering approaches. We compute the absolute difference of a household's estimated MPC (based on the quantile assignment rule of Misra and Surico (2014) for quantile regression and the maximal weight for clustering approaches) from its true MPC. We report the average quantiles of this error distribution. FCP and HKM are roughly comparable at all points across the distribution (although FCP outperforms HKM for other specifications, as before). Both the Gaussian mixture model and quantile regression perform considerably worse, but quantile regressions generate a tail of very large MPC errors.

Overall, we find evidence that quantile regression mischaracterizes the MPC distribution in this setting, in particular estimating a spurious left tail, with many MPCs below the true minimum, and even below zero. We provide an intuitive illustration for why that may be the case, based on the role of multiple dimensions of heterogeneity in the data.

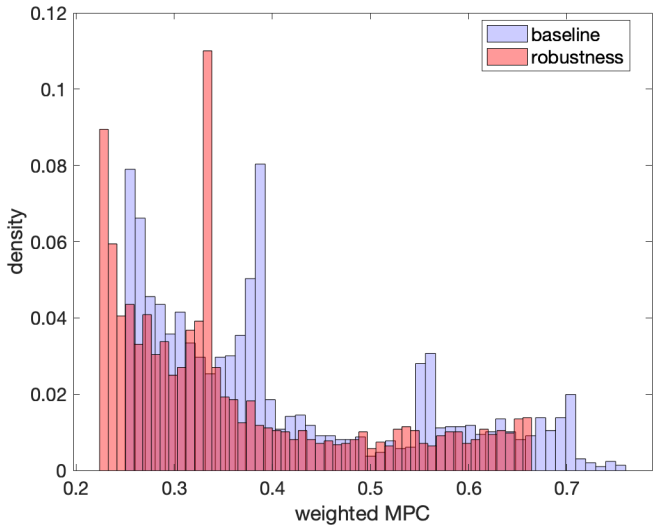
Figure 13: Empirical CDFs of the MPC distribution: quantile regression



Notes: 500 samples. Medians of the estimated CDF of the MPC distribution across samples, with individuals assigned based on their maximal weights for FCP. $G = 5$ and Gaussian errors. For quantile regressions, we assign observations to estimated MPCs as in Misra and Surico (2014): a household is assigned to each quantile τ(and associated MPC) if $[y - X\alpha(\tau)] \leq 0$ and $[y - X\alpha(\tau - 0.05)] > 0$, where y is the dependent variable, $X$is the set of right-hand side variables, and α the set of quantile-specific estimated coefficients.

Ultimately, our results are based on simulations, since we do not know the true distribution of MPCs in the data, but these simulations are calibrated to represent the data for our empirical setting.

Figure 14: The MPC distribution with heterogeneous coefficients on controls



Notes: Estimated MPC distributions for the baseline total expenditures specification as in Figure 1, here shown with blue histogram. The red histogram shows the weighted MPC distributions estimated in a specification that allows coefficients on household-level controls to be group specific.

D Supplemental empirical results

In this section we report additional empirical results. First, we allow for heterogeneity in all coefficients on household-specific variables in Equation (2), which leaves the estimated distribution of MPCs largely unchanged. Second, we estimate the MPC distribution using HKM and a Gaussian mixture model, obtaining results broadly similar to our baseline, albeit quantitatively different. Third, we report additional details on the MPC distribution for several specifications. Fourth, we document further results on the relationship between MPC heterogeneity and observable characteristics. Finally, we provide details on the interpretation of the MPC in this setting.

D.1 Heterogeneous coefficients on household characteristics

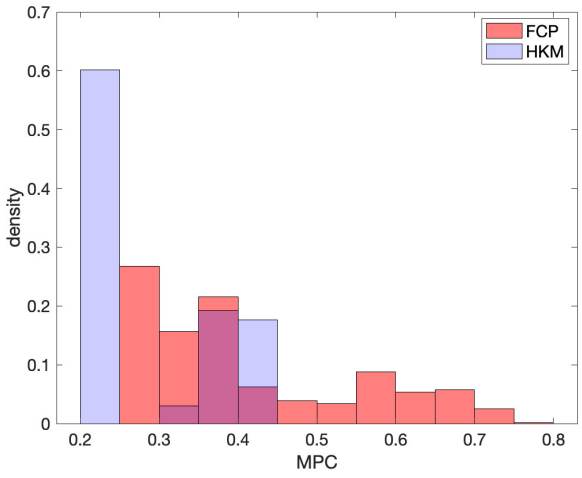
It is natural to wonder if there is also a role for heterogeneity in the coefficients on additional household characteristics included as controls. In Equation (2), this amounts to interacting W_j with the group dummies and allowing β to vary across group. We allow heterogeneity in all household characteristics, but maintain homogeneous coefficients on time dummies. Figure 14 plots the estimated MPC distributions for total expenditures allowing for this heterogeneity against our baseline distribution; the distribution is very similar across the two specifications.

D.2 The MPC distribution under alternative estimators

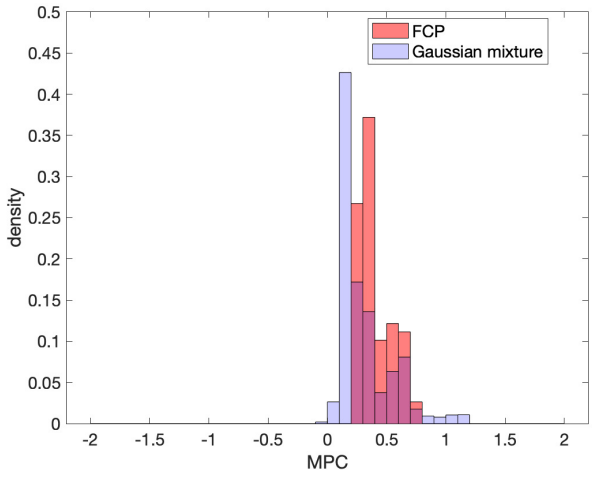
To assess the robustness of the MPC distribution to the estimator used, we additionally estimate MPCs using HKM and a Gaussian mixture model implemented using the EM algorithm. Recall that based on the simulation results detailed in Section C.2, the performance of these estimators tends to be considerably poorer in our setting. Figure 15 overlays the distributions recovered using each estimator. For HKM, we set the same number of groups as in FCP. For the Gaussian mixture model, the BIC chooses 6 groups. The histograms show that the general shape of the distribution is similar across estimators, confirming our baseline findings. However, the dispersion of MPCs is quite different. HKM predicts a much more concentrated distribution, and misses the right tail. In other words, it implies that there are no households that consume a large fraction of their rebate. The Gaussian mixture model, on the other hand, generates much higher dispersion, and even some negative MPCs. We document this tendency to counterfactually stretch the distribution beyond its true domain in the simulation study in Supplement C.2.

Figure 15: The MPC distribution under alternative estimators

(a) FCP and HKM



(b) FCP and Gaussian mixture

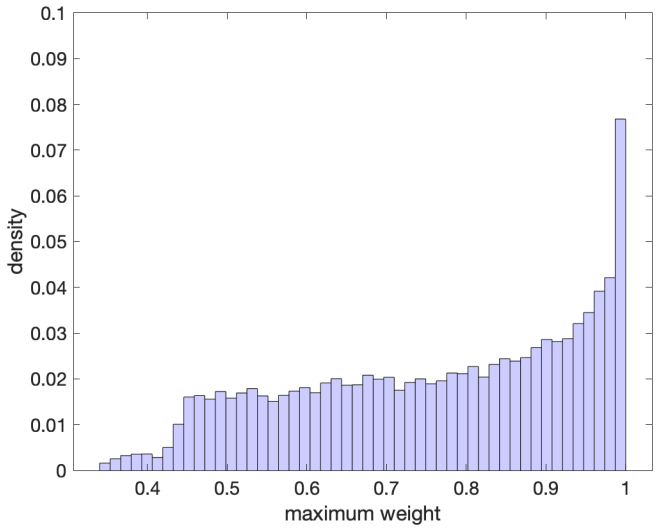


Notes: Estimated MPC distributions for the baseline total expenditures specification as in Figure 1. In both panels, the red distributions display the weighted MPCs as computed in the baseline FCP model. In the left panel, we overlay the distribution of HKM MPCs. In the right panel, the distribution of weighted MPCs obtained estimating the Gaussian mixture model with the EM algorithm.

D.3 The MPC distribution: additional results

Figure 16 depicts the distribution of estimated individual maximal weights. Some households' assignments are estimated with near certainty (those with a maximum weight of

Figure 16: Estimated distribution of households' maximal weights



Notes: Figure 16 plots a histogram of estimated maximal weights for each household for total expenditures, defined as in Parker et al. (2013).

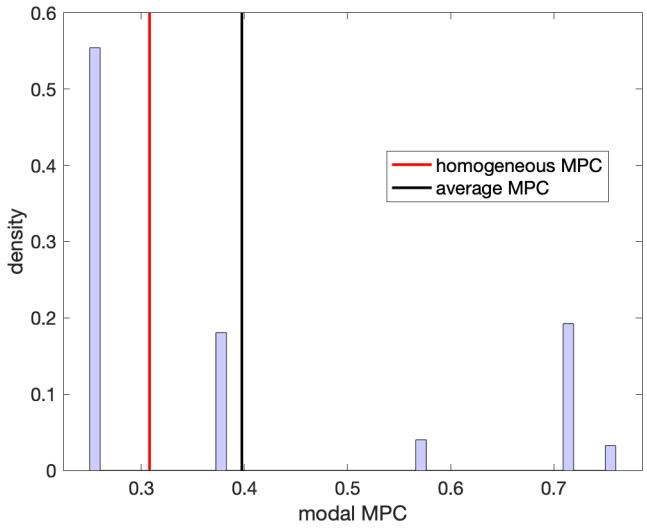
near unity), but is clearly quite different from the binary assignment imposed by HKM. Our weighted approach therefore allows us to recover the smoother distribution shown in Figure 1. Figure 17, instead, shows the distribution of MPCs for total expenditures associated with individuals' maximal weights.

We further confirm the reliability of our estimated MPC distribution as follows. We draw 250 samples via bootstrap with replacement. For each sample, we estimate our FCP algorithm (using the baseline specification). Table 10 shows how, on average, the bootstrapped samples generate quantiles of the weighted MPC distribution that are very close to those of the empirical distribution shown in Figure 1. This demonstrates that even though some MPC estimates may not be individually statistically significant, there is relatively little sampling uncertainty surrounding the overall shape of the distribution. Similar results hold if we look at the distribution of MPCs associated with individuals' maximal weights.

Figure 18 reports the 68% and 90% confidence bands of our estimated MPCs for total expenditures, estimated via TSLS and dropping the households that never receive a rebate.

Finally, we repeat the analysis in Section 4.1 and show that nearly all nondurable MPCs are statistically different from each other when we estimate a heterogeneous WLS model taking the FCP weights as given, as explained in the notes for Table 1.

Figure 17: Estimated distribution of MPCs out of the tax rebate: maximal weights



Notes: Figure 17 plots a histogram of estimated MPCs associated with maximal weights for total expenditures, defined as in Parker et al. (2013). The homogeneous MPC (red line) is estimated assuming homogeneous response to the tax rebate, as in Parker et al. (2013). The black line shows the average MPC associated with individuals' maximal weights in our sample.

Table 10: Average moments of the MPC distribution across bootstrapped samples

	Mean	p10	p25	p50	p75	p90
Data	0.41	0.26	0.30	0.38	0.53	0.64
Bootstrap	0.41	0.24	0.28	0.36	0.50	0.66
	(0.18)	(0.14)	(0.13)	(0.16)	(0.20)	(0.22)

Notes: Table 10 shows various statistics of the distribution of weighted MPCs in the data (first row) and in a bootstrap exercise. pxx denotes the xx^{th} percentile. "Data" refers to Figure 1, while the row labeled "Bootstrap" reports the average of each moment across 250 bootstrapped samples. The last row reports the standard deviation of each moment across bootstrapped samples.

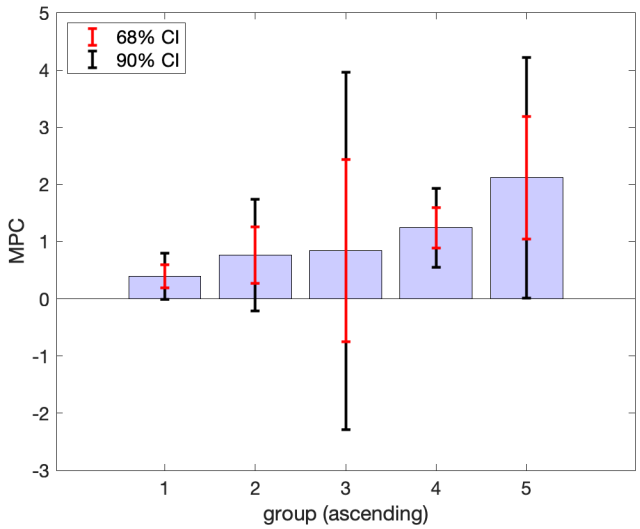
D.4 What drives MPC heterogeneity: additional results

Figure 19 graphically displays the correlation of MPCs for durable and nondurable goods.

We have shown in Section 4.3 that some household characteristics individually correlate with the MPC distribution. Here, we analyze whether the linear correlation with age and liquid wealth hides some non-linear pattern. Figure 21a suggests a positive and convex relationship between the weighted MPC and log liquid wealth. The relationship looks instead concave with respect to age, as shown in Figure 21b.

Neither relationship is robust to the inclusion of a set of controls, as shown in Section 4.3. In Table 12 we show that the joint drivers of nondurable MPCs are the same as those uncovered for the MPC for total expenditures in Table 3. Considering the APC for nondurable expenditures delivers a threefold increase in the magnitude of the relationship with the MPC. Turning to the MPC for durable goods, we find almost no significant cor-

Figure 18: Estimated MPCs out of the Tax Rebate: TSLS



Notes: Figure 18 depicts the estimated MPCs for total expenditures (defined as in Parker et al. (2013)) each of the $G = 5$ groups using the TSLS specification. 90% confidence intervals (black lines) and 68% confidence intervals (red lines) are depicted as vertical lines. The confidence intervals are constructed using the analytical formulas given in Theorem 2.

relations. Due to the concentrated nature of the distribution, we examine observations with a durables MPC of about one quarter. While APC and non-salary income remain positively correlated with the marginal propensity to consume, this relationship is statistically insignificant.

In Figure 21, we show that "rich-spenders" (i.e. households with high APC and high total income) have high MPC for nondurable expenditures too.

Finally, we explore nonlinear effects of observable predictors of the MPC distribution. We estimate a multinomial logit model, using the MPC associated with an individual's maximal weight as the dependent variable. Table 13 reports the estimation output, whereas Figure 22 plots the marginal effects for the APC.[45] The results confirm that, conditional on other predictors, non-salary income and the APC are the most significant predictors of the marginal propensity to consume, even when we allow for nonlinear effects.

[45]We report only the results for the MPC distribution estimated via WLS and an array of predictors that excludes liquid wealth. Results are virtually unchanged if we look at the TSLS MPC or include liquidity.

Table 11: Test for MPC equality: non-durables

	0.03	0.15	0.22	0.24	0.49
0.03	8.61				
	(0.00)				
0.15	73.40	310.24			
	(0.00)	(0.00)			
0.22	134.78	27.27	332.57		
	(0.00)	(0.00)	(0.00)		
0.24	33.72	6.90	0.29	50.82	
	(0.00)	(0.01)	(0.59)	(0.00)	
0.49	167.41	100.52	60.28	29.71	212.48
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: The table shows $F-$statistics from pairwise two-sided Wald tests of equality across MPCs (the diagonals shows tests of equality with zero). FCP weights are taken as given. $p-$values are reported in parentheses.

D.5 Rebate coefficient versus MPC

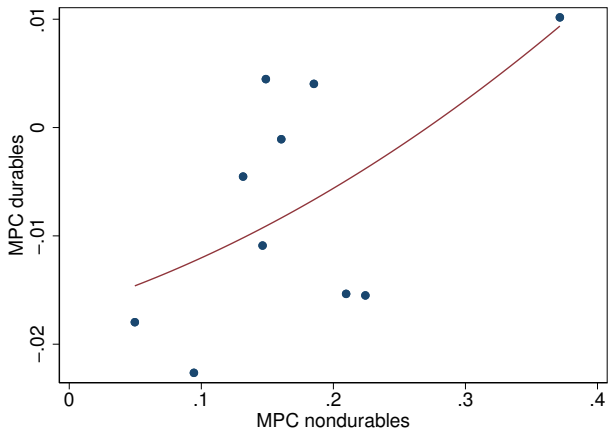
As discussed by Kaplan and Violante (2014), ϕ may not correctly measure the marginal propensity to consume out of a transitory income shock, but is instead better thought of as a "rebate coefficient". This is because the control group of non-recipients in period t is made of three groups: (i) households that never receive the rebate, (ii) households that have not yet received a rebate, but may anticipate receiving the rebate in the future, and (iii) households that have already received the rebate. The second group might display a positive MPC out of news of the rebate, biasing the estimated rebate coefficient θ downward. Similarly, the third group might also have a positive lagged MPC out of the rebate, further contributing to a downward bias. Following Kaplan and Violante (2014), we modify the specification in Equation (1) by introducing the lag of the rebate variable R_j^{lag} so that the estimated rebate coefficient can be interpreted as an MPC:

$$\Delta C_j = \beta' W_j + \phi R_j + \phi^{\text{lag}} R_j^{\text{lag}} + \alpha + \epsilon_j \tag{21}$$

By absorbing the lagged consumption response, this modification accounts for the fact that, in the baseline specification, the control group includes households that received the rebate in the past, and whose consumption response might be persistent.[46] We then interact the rebate, its lagged value, and the constant with the group indicators $\mathbf{1}[j \in g]$,

[46]This is true as long as the persistent effect of the rebate lasts strictly less than four quarters. Moreover, we assume that the policy is fully anticipated by all households. In an intermediate information case in which, for instance, the policy enters the agents' information set after the receipt of the first rebate, this specification cannot fully account for anticipatory effects often labelled as the MPC out of news.

Figure 19: The correlation of MPCs across consumption goods



Notes: Figure 19 shows a binscatter of household MPC estimates for durables against nondurables. Each dot shows the average weighted MPC for durable goods for each decile of the distribution of weighted MPCs for nondurable goods. The red line shows the quadratic fit.

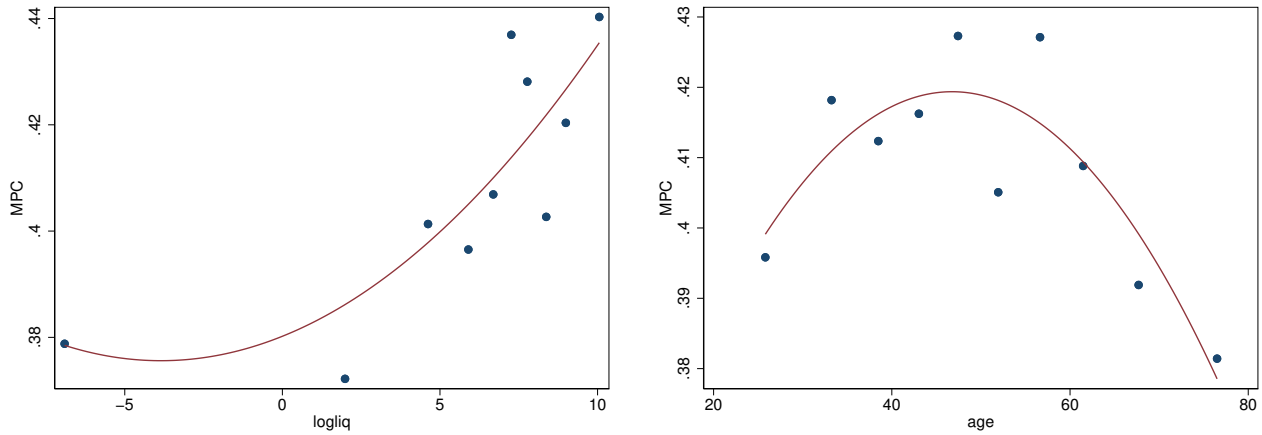
and use FCP algorithm to estimate the vector of coefficients $\left\{\phi_g, \phi_g^{\text{lag}}, \alpha_g, \beta\right\}$. In Figure 23, we show that the distribution of weighted MPCs is very similar to the one estimated in the baseline specification.

To address the fact that some households never receive the rebate (and thus may be meaningfully different from those who do receive the rebate), we drop households who do not receive a rebate within the sample period that we cover. 40% of the observations in the sample are associated with households that do not receive a rebate in this time period. Figure 24 shows the distribution of the weighted MPCs in this subsample. Our results are very similar to those shown in Figure 1, with a slight rightward shift of the distribution. Indeed, Parker et al. (2013) also estimate a larger homogeneous rebate coefficient in this subsample.

E Homogeneous effects and average heterogeneous effects

Across specifications, we find that the estimated MPC from a homogeneous regression is often lower than the average MPC from our heterogeneous models. In general, homogeneous effects need not equal the average from heterogeneous specifications. In this section, we explain why, and provide a decomposition to help understand what factors drive this discrepancy in our data. This issue is related to Słoczyński (2020), who studies this problem for OLS in a binary treatment setting.

Figure 20: Marginal propensities to consume: liquid wealth and age



(a) Liquid wealth (b) Age

Notes: Figure 20 plots binscatters showing the average weighted MPC for total expenditures for each decile of the distribution of lagged log liquid wealth (left panel) and for each decile of the distribution of age of the reference person in the household (right panel). Log liquid wealth takes log(0.001) when liquid wealth is 0 or negative. The red line shows the quadratic fit.

First, consider the univariate regression

$$y_j = \alpha_{g(j)} + \phi_{g(j)} x_j + u_j, j = 1, \ldots, N, \tag{22}$$

where $g(j)$ indicates individual j's group membership and $g = 1, \ldots, G$. For this discussion, we assume that $g(j)$ is known. Let $X_j = \begin{bmatrix} 1 & x_j \end{bmatrix}'$. In the population, the expressions for the heterogeneous parameters of group $\tilde{g}$ are given by

$$\begin{pmatrix} \alpha_{\tilde{g}} \\ \phi_{\tilde{g}} \end{pmatrix} = E\left[X_j X_j' \mid g(j) = \tilde{g}\right]^{-1} E\left[X_j y_j \mid g(j) = \tilde{g}\right]$$

Let $\pi_{\tilde{g}}$ denote the share of the population belonging to group $\tilde{g}$. Now suppose instead the econometrician estimates Equation (22) as a homogeneous regression, pooling the data across groups. The expression for these parameters is:

$$\begin{pmatrix} \alpha_{homo} \\ \phi_{homo} \end{pmatrix} = E\left[X_j X_j'\right]^{-1} E\left[X_j y_j\right],$$

Figure 21: The relationship between MPCs, APCs, and income: nondurable expenditures

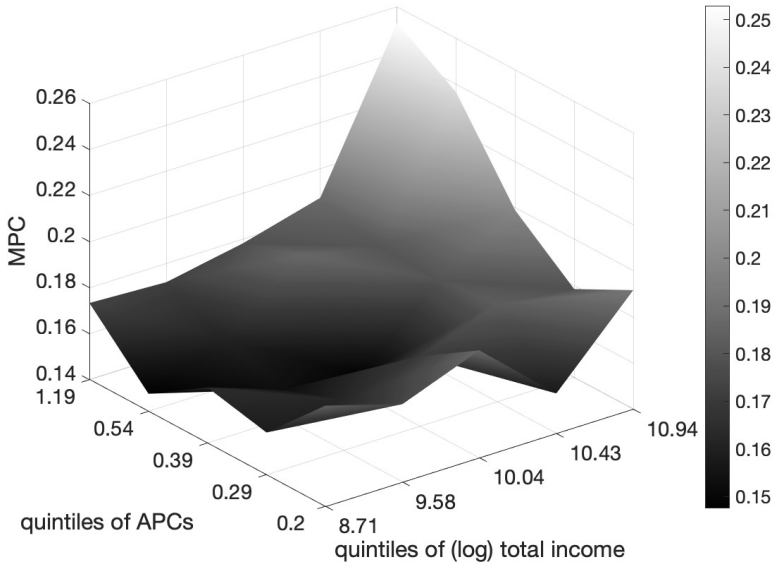


Figure 21 shows the average weighted MPC for nondurable expenditures for pairs of quintiles of APC and log total income. The color-bar on the right represents the MPC.

where

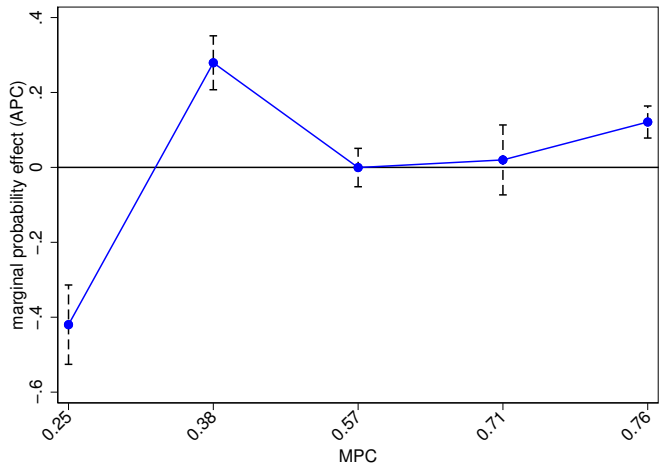
$$E\left[X_j X_j'\right]^{-1} = \left(\sum_{g=1}^{G} \pi_g E\left[X_j X_j' \mid g\left(j\right) = g\right]\right)^{-1},$$

$$E\left[X_j y_j\right] = \sum_{g=1}^{G} \pi_g E\left[X_j\left(\alpha_g + \phi_g x_j\right) \mid g\left(j\right) = g\right].$$

The numerator is indeed a weighted average of the group moments, where the weights are the population shares in each group. However, simple algebra shows that in general $\begin{pmatrix} \alpha_{homo} \\ \phi_{homo} \end{pmatrix} = \sum_{g=1}^{G} \pi_g \begin{pmatrix} \alpha_g \\ \phi_g \end{pmatrix}$ if and only if $E\left[X_j X_j'\right] = E\left[X_j X_j' \mid g\left(j\right) = \tilde{g}\right] \forall \tilde{g}$. Thus, deviations of the homogeneous estimates from the average of heterogeneous effects follow from differences in the distribution of right-hand-side variables across groups.

We provide an analytical decomposition of which moments of X_j drive these deviations. In this setting where group membership is known, the formulas are exact. To obtain these expressions, we simply compute the second row of $E\left[X_j X_j'\right]^{-1}$, multiply it

Figure 22: Marginal and average propensities to consume: multinomial logit



Notes: Figure 22 shows the marginal probability effect of a household's APC, measured as mean lagged consumption relative to lagged total income, on the MPC for total expenditures associated with an individual's maximal weight. Marginal effects are estimated by the estimated regression in Table 13.

by $\sum_{g=1}^{G} \pi_g E\left[X_j\left(\alpha_g + \phi_g x_j\right) \mid g\left(j\right) = g\right]$, and collect terms. In particular, this yields:

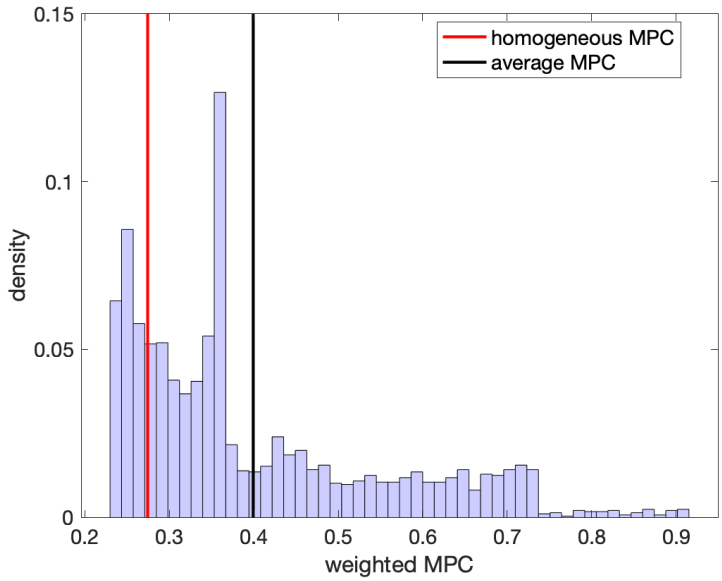
$$\phi_{homo} - \sum_{g=1}^{G} \pi_g \phi_g = \sum_{g=1}^{G} \pi_g \alpha_g \frac{E\left[x_i \mid g\left(j\right) = g\right] - E\left[x_j\right]}{\sigma_x^2} + \sum_{g=1}^{G} \pi_g \phi_g \left(\frac{E\left[x_j^2 \mid g\left(j\right) = g\right] - E\left[x_j\right] E\left[x_j \mid g\left(j\right) = g\right]}{\sigma_x^2} - 1 \right), \tag{23}$$

where σ_x^2 is the unconditional variance of x. It is clear from Equation (23) that if the group-specific moments of x_j are equal to their unconditional counterparts, both terms are equal to zero. These expressions can shed light on the source of the discrepancy between homogeneous and the average of the estimated heterogeneous effects. The first term will be equal to zero if the mean of x is group invariant. The second term will be equal to zero if the variance of x is group invariant. Even if moments vary across groups, the first term will be zero if the group means of x are uncorrelated with the fixed effects α_g, and the second will be zero if the group variances are uncorrelated with the coefficients ϕ_g. Crucially, this decomposition presumes that groups $g\left(j\right)$ are known.

These analytical expressions hold for the simple univariate regression (22). However, we are more generally interested in multivariate regressions, as in our empirical setting, for example:

$$y_i = \alpha_{g(j)} + \beta_{g(i)} x_j + \delta'_{g(j)} W_j + u_j, j = 1, \ldots, N,$$

Figure 23: Estimated distribution of MPCs out of the tax rebate: control for lagged responses



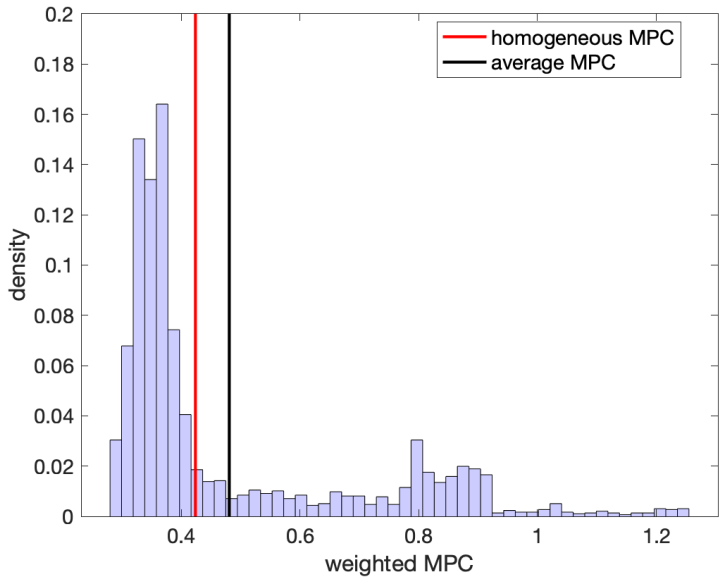
Notes: Total expenditures, defined as in Parker et al. (2013). The homogeneous MPC (red line) is estimated assuming a homogeneous response to the tax rebate, as in Parker et al. (2013). For each household we compute the weighted MPC. The black line shows the average weighted MPC in our sample.

where W_j is $K_W \times 1$. It is possible to extend the analytical expressions (23) to accommodate one or two additional regressors, but the expressions become exponentially more complicated in the number of regressors. However, we can still compute a decomposition like that in (23). Now, denote $\tilde{X}_j = \begin{pmatrix} X_j \\ W_j \end{pmatrix}$. A general decomposition can be computed as

$$\begin{aligned} \phi_{homo} - \sum_{g=1}^{G} \pi_g \phi_g = & \sum_{g=1}^{G} \pi_g \alpha_g E\left[\tilde{X}_j \tilde{X}_j'\right]_{2\cdot}^{-1} E\left[\tilde{X}_j\right], \\ & + \sum_{g=1}^{G} \pi_g \phi_g \left(E\left[\tilde{X}_j \tilde{X}_j'\right]_{2\cdot}^{-1} E\left[\tilde{X}_j x_j\right] - 1 \right) \\ & + \sum_{k=1}^{K_W} \sum_{g=1}^{G} \pi_g \delta_g E\left[\tilde{X}_j \tilde{X}_j'\right]_{2\cdot}^{-1} E\left[\tilde{X}_j W_{jk}\right] \end{aligned} \tag{24}$$

where $E\left[\tilde{X}_j \tilde{X}_j'\right]_{2\cdot}^{-1}$ denotes the second row of $E\left[\tilde{X}_j \tilde{X}_j'\right]^{-1}$. The terms in (24) provide a similar decomposition into the components driving the deviation between homogeneous and average heterogeneous parameter estimates. Of course, the last term can be further decomposed into the elements of the summation over k to provide a separate contribution

Figure 24: Estimated distribution of MPCs out of the tax rebate: only rebate recipients



Notes: Total expenditures, defined as in Parker et al. (2013). The homogeneous MPC (red line) is estimated assuming a homogeneous response to the tax rebate, as in Parker et al. (2013). For each household we compute the weighted MPC. The black line shows the average weighted MPC in our sample. The homogeneous response and average heterogeneous response are much closer than in other specifications.

for each covariate W_k.

Importantly, these expressions all assume that group membership is binary. If group membership is non-binary, they can still provide a heuristic decomposition of the discrepancy between homogeneous and heterogeneous estimates. We thus take Equation (24) to our data in two different ways. First, we generate binary group assignment for each individual by setting their maximal weight equal to one and all other weights equal to zero. We then estimate both homogeneous and average heterogeneous parameters for this model, the difference between which will be exactly explained by Equation (24), the terms of which we compute based upon this specification. Second, we compute the terms of the decomposition using our FCP parameters, but the same binary weights as before to compute the group-specific moments. We conduct this exercise for both our baseline and IV specifications.

Table 14 reports the results. $\bar{\phi}$ denotes the average MPC based on the binary membership described above, and $\bar{\phi}_\mu$ denotes the average MPC using the FCP group weights, μ. By definition, the three decomposition terms sum to the parameter deviation for the specifications using parameters estimated based on binary group assignment. However, turning to the FCP specifications, the sums of the decomposition terms are reasonably close to the overall deviation. This heuristic approach captures most of the deviation

since, for many individuals, the final weights are close to binary, see Figure 16. Across specifications, we find that the term due to variation in the covariance of regressors explains virtually none of the deviation. We find that the term due to the covariance of the rebate with regressors explains very little of the deviation. The vast majority of the deviation between homogeneous and average heterogeneous effects instead comes from variation in the mean of regressors with the fixed effects across groups. Further decomposing this term into the specific covariates associated with the mean of each regressor, we find that variation in the means of certain time dummies and the age variable across groups is most responsible for the difference in the estimates.

At first glance, variation in the moments of regressors across groups might appear to compromise identification. However, correlation of regressors or their moments with group membership does not violate identification. Rather, that is only the case if, beyond the group-based heterogeneity incorporated in the model, x_i is correlated with the error term. The point is particularly clear for the IV specification, where treatment is effectively binary. Recall that the average treatment effect, written in terms of potential outcomes $y_j(\cdot)$ for binary treatment Z_j is defined as:

$$E\left[y_j(1) - y_j(0) \mid W_j\right]. \tag{25}$$

In a primitive sense, our estimand is the object:

$$E\left[y_j(1) \mid Z_j = 1, W_j\right] - E\left[y_j(0) \mid Z_j = 0, W_j\right]. \tag{26}$$

Given the nature of our instrument, the random timing at which rebates are received, $E\left[y_j(1) \mid Z_j = 1, W_i\right] - E\left[y_j(1) \mid Z_j = 0, W_j\right]$ (and similarly for $y_j(0)$), and the distribution of W_j is also independent of Z_j. Thus, (25) and (26) are equivalent. Of course, there is no assumption that $y_j(\cdot)$ is independent of W_j – if that were the case, there would be no reason to include the controls. However, this framework may provide motivation to use an IV approach as opposed to OLS, since using x_j, the dollar value of the rebate as treatment, leaves open the possibility that x_j is correlated with potential outcomes $y_j(\cdot)$.

Table 12: Explanatory variables: weighted MPC for nondurables and durables

	(1)	(2)	(3)	(4)
dummy for no salary	-0.053	-0.075	-0.576*	-0.359
	(0.091)	(0.075)	(0.275)	(0.232)
log salary income	-0.003	-0.005	-0.035	-0.021
	(0.006)	(0.005)	(0.018)	(0.015)
log non-salary income	0.021*	0.026***	0.005	0.011
	(0.008)	(0.006)	(0.023)	(0.016)
mortgage interest/income	-0.004	0.006	-0.000	0.046
	(0.024)	(0.023)	(0.086)	(0.079)
APC	0.024**	0.022**	0.005	-0.002
	(0.009)	(0.008)	(0.021)	(0.014)
homeowner dummy	0.011	0.008	-0.011	-0.010
	(0.010)	(0.008)	(0.028)	(0.021)
dummy for mortgage	-0.010	-0.010	-0.016	-0.013
	(0.010)	(0.008)	(0.030)	(0.023)
log liquid assets	-0.001		0.003	
	(0.002)		(0.004)	
dummy for liquid assets >= 0	-0.028		0.008	
	(0.025)		(0.056)	
N	739	1099	714	1062
R^2	0.034	0.040	0.044	0.025

Notes: Columns (1) and (2) use the weighted MPC for non-durables as dependent variable. The dependent variable in columns (3) and (4) is a dummy that takes 1 if the weighted durable MPC is above 0.25, which pertains to roughly 5% of the observations. All logged variables take a value of log(0.001) when the raw value is 0 or negative. Non-salary income is positive for all observations. Standard errors are robust to heteroskedasticity and reported in parentheses. We control for marriage dummies, education dummies, number of children, age and age squared; those coefficients are not reported. Age and its square are controls in our FCP estimation. While this does not pose an issue for the point estimates shown in this table, it could potentially affect inference. However, we repeat the same regressions, excluding age and and age squared, and find that the remaining coefficients are unaffected. , *, and *** denote significance at the 10, 5, and 1% levels respectively.

Table 13: Multinomial logit on maximal weight MPC for total expenditures

MPC	0.38	0.57	0.71	0.76
Dummy for no salary	-0.52	-5.65*	-3.41*	1.62
	(2.26)	(3.12)	(1.95)	(5.88)
Log salary income	0.03	-0.32	-0.20	0.45
	(0.15)	(0.20)	(0.13)	(0.41)
Log non-salary income	1.72***	1.12**	1.10***	3.96***
	(0.24)	(0.45)	(0.22)	(0.71)
Mortgage interest/income	-0.31	-1.07	1.45	-3.49*
	(0.88)	(2.56)	(0.81)	(2.09)
APC	2.94***	0.99	1.12***	7.38***
	(0.35)	(0.86)	(0.37)	(0.92)
Homeowner dummy	0.41	-0.53	0.74**	0.96
	(0.33)	(0.57)	(0.30)	(1.00)
Mortgage dummy	-0.25	0.15	-0.48*	-0.28
	(0.28)	(0.53)	(0.26)	(0.73)
Pseudo R^2		0.12		
Number of observations		1079		

Notes: Output from a single multinomial logit regression of MPCs on observable characteristics. The excluded base outcome is the lowest MPC, 0.25. All logged variables takes log(0.001) when the raw value is 0 or negative. Non-salary income is positive for all observations. *, ** and *** denote significance of the coefficients at 10, 5 and 1% levels respectively. Standard errors are in parentheses. Coefficients on additional controls are not reported and include age, age squared, marriage dummy, number of children and education dummies as in Table 3.

Table 14: Decomposition of deviation between homogeneous and average heterogeneous coefficients

Binary groups

term	mean	rebate	controls	sum	$\phi_{homo} - \bar{\phi}$
OLS	-0.13	0.02	0.00	-0.11	-0.11
IV	-0.12	-0.00	0.00	-0.13	-0.13

FCP

term	mean	rebate	controls	sum	$\phi_{homo} - \bar{\phi}$	$\phi_{homo} - \bar{\phi}_{\mu}$
OLS	-0.12	0.02	0.00	-0.10	-0.11	-0.09
IV	-0.11	-0.01	0.00	-0.11	-0.19	-0.22

Notes: Contributions to the deviation between the homogeneous MPC and the average weighted MPC are computed using the formula in (24). The binary groups results compare homogeneous MPC to the mean MPC estimated using WLS where the weights are binary, according to individuals' maximal weights. The FCP panel compares the homogeneous result to the mean MPC estimated in our baseline, where the mean is computed either based on binary assignment according to maximal weights, $\bar{\phi}$, or the weighted average using FCP weights, $\bar{\phi}_{\mu}$.

References

BEZDEK, J. C. (1973): "Fuzzy Mathematics in Pattern Classification," *Ph.D. dissertation, Cornell University*.

BONHOMME, S. AND E. MANRESA (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147–1184.

BOZ, E., G. GOPINATH, AND M. PLAGBORG-MÃŽLLER (2019): "Dollar Invoicing and the Heterogeneity of Exchange Rate Pass-Through," *AEA Papers and Proceedings*, 109, 527–32.

CRAWLEY, E. AND A. KUCHLER (2018): "Consumption Heterogeneity: Micro Drivers and Macro Implications," *Danish National Bank Working Paper 129*.

DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38.

DUNN, J. C. (1977): "Indices of partition fuzziness and the detection of clusters in large data sets," .

FAGERENG, A., M. B. HOLM, AND G. J. NATVIK (2016): "MPC Heterogeneity and Household Balance Sheets," Discussion Papers 852, Statistics Norway, Research Department.

HAYASHI, F. (2011): *Econometrics*, Princeton University Press.

JOHNSON, D. S., J. A. PARKER, AND N. S. SOULELES (2006): "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review*, 96, 1589–1610.

KAPLAN, G. AND G. L. VIOLANTE (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82, 1199–1239.

KAPLAN, G., G. L. VIOLANTE, AND J. WEIDNER (2014): "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, 45, 77–153.

MISRA, K. AND P. SURICO (2014): "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs," *American Economic Journal: Macroeconomics*, 6, 84–106.

NEWEY, W. K. AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," Elsevier, vol. 4 of *Handbook of Econometrics*, 2111 – 2245.

PARKER, J. A., N. S. SOULELES, D. S. JOHNSON, AND R. MCCLELLAND (2013): "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, 103, 2530–53.

POLLARD, D. (1981): "Strong Consistency of K-Means Clustering," *The Annals of Statistics*, 9, 135–140.

——— (1982): "A Central Limit Theorem for K-Means Clustering," *The Annals of Probability*, 10, 919–926.

SŁOCZYŃSKI, T. (2020): "Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights," *The Review of Economics and Statistics*, forthcoming, 1–27.

SPIVAK, M. (1971): *Calculus On Manifolds: A Modern Approach To Classical Theorems Of Advanced Calculus*, Avalon Publishing.

TIBSHIRANI, R., G. WALTHER, AND T. HASTIE (2001): "Estimating the Number of Clusters in a Data Set via the Gap Statistic," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63, 411–423.

YANG, M.-S. (1994): "On Asymptotic Normality of a Class of Fuzzy C-Means Clustering Procedures," *International Journal of General Systems*, 22, 391–403.

YANG, M.-S. AND K. F. YU (1992): "On Existence and Strong Consistency of a Class of Fuzzy C-Means Clustering Procedures," *Cybernetics and Systems*, 23, 583–602.