A New Approach to Estimating Heterogeneous Effects and the Distribution of the Marginal Propensity to Consume

Daniel J. Lewis Davide Melcangi Laura Pilossoph

Federal Reserve Bank of New York

November 5, 2019

These views are my own and do not represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

Heterogeneous effects

Economists are interested in how effects vary over individuals.

- Our example: the marginal propensity to consume (MPC)
- How do different individuals respond to a transfer?
- Currently a very important object for policy questions

Heterogeneous effects

Economists are interested in how effects vary over individuals.

- Our example: the marginal propensity to consume (MPC)
- How do different individuals respond to a transfer?
- Currently a very important object for policy questions

We could just split into groups and estimate effects...

- But the literature disagrees on determinants of MPC.
- Maybe the true determinants are unobservable.

We would like to classify and estimate "endogenously".

Clustering algorithms

One way to do this is clustering algorithms – widely studied in statistics

- Gaussian mixture models, K-means, etc.
- Most frequently, simply estimate group means.

Recently, Bonhomme and Manresa (2015) extend K-means to regression problems.

- Possibly heterogeneous slopes as well as fixed effects

Clustering fits our goal: group assignment is determined by (unobservable) relationship between outcome and independent variable alone.

A good fit for macro data?

K-means has two parts: estimate group assignment, estimate parameters.

- The first stage needs to be consistent.
- This requires a long panel structure, $T \to \infty$, and fixed type.

A good fit for macro data?

K-means has two parts: estimate group assignment, estimate parameters.

- The first stage needs to be consistent.
- This requires a long panel structure, $T \to \infty$, and fixed type.

In most economic settings, T is not very large...

- $T = 7$ in Bonhomme and Manresa (2015)

In fact, we often have cross sectional data, $T = 1$.

Bonhomme and Manresa (2015) have small-T asymptotics, but how does K-means do in practice?

Fuzzy C-means

The Fuzzy C-means algorithm (FCM) was introduced by Dunn (1973) in computer science.

- Instead of binary assignment, individuals are given weights.
- Membership no longer needs to be consistently estimated.
- Can be thought of as a non-parametric Bayesian interpretation.
- T does not need to tend to infinity.
- Uncertainty in group assignment with small T explicitly accommodated.
- Up to now, only used for clustering with mean parameter.

What we do

In this paper, we

- Extend FCM to regression (and IV) problems,
- Show iterative FCM equivalent to one-step GMM problem,
- Derive asymptotic distributions for FCM regressions,
- Show advantages of FCM in simulations,
- Recover the distribution of heterogeneous MPCs using the 2008 tax rebate,
- Assess the determinants of an individual's MPC.

Literature

- **K-means algorithm:** Pollard (1981, 1982); Bonhomme and Manresa (2015); Bonhomme et al. (2017); Lin and Ng (2012)
- **FCM:** Dunn (1973); Bezdek et al. (1984); Yang and Yu (1992); Yang (1994)
- **Heterogeneous treatment effects in machine learning:** Athey and Imbens (2016); Chernozhukov et al. (2018); Davis and Heller (2018)
- **MPC (heterogeneity) in the data:**
  - **(2001) 2008 tax rebate:** Johnson, Parker, and Souleles (2006); Parker et al (2013)
  - **Heterogeneity by observables:** Kaplan, Violante and Weidner (2014); Fagereng, Holm and Natvik (2018); Aguiar, Bils and Boar (2019); Crawley (2018); ...
  - **Reported preference:** Sahm, Shapiro and Slemrod (2010); Jappelli and Pistaferri (2014); Graziani, van der Klaauw and Zafar (2016); Fuster, Kaplan and Zafar (2018)
  - **Unconditional heterogeneity:** Misra and Surico (2014)
- **MPC (heterogeneity) in theory:**
  - Carroll and Kimball (1996); Kaplan and Violante (2014); ...
  - Macro transmission: Bilbiie (2008); Kaplan, Moll and Violante (2018); Auclert (2018); Patterson (2019)

Roadmap

1. Introduction
2. The FCM algorithm
3. Asymptotics (and other theory)
4. Simulations & performance
5. Heterogeneity in MPCs

The K-means algorithm

- The K means algorithm minimizes the objective function

$$L_1(P, \psi) = \int \min_g \|y - \psi_g\|^2 P(dy),$$

$y \in R^T$ are outcomes with probability measure P, $g \in \{1, 2, \ldots, G\}$ indexes groups, and $\psi \in R^{G \times T}$.

The K-means algorithm

- The K means algorithm minimizes the objective function

$$L_1(P, \psi) = \int \min_g \|y - \psi_g\|^2 P(dy),$$

  $y \in R^T$ are outcomes with probability measure P, $g \in \{1, 2, \ldots, G\}$ indexes groups, and $\psi \in R^{G \times T}$.

- "Hard" K-means (HKM) can be rewritten using weights:

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

where $w_g = 1\left[\|y - \psi_g\|^2 \leq \|y - \psi_h\|^2 \ \forall \ h \neq g\right]$.

The Fuzzy C-means algorithm

FCM generalizes HKM by adjusting the weight w_g:

$$J_m(P, \mu, \phi) = \int \sum_{g=1}^{G} \mu_g^m(y) \|y - \phi_g\|^2 P(dy),$$

and in particular setting optimal weights

$$\mu_g(y, \phi) = \left(\sum_{h=1}^{G} \frac{\|y - \phi_g\|^{2/(m-1)}}{\|y - \phi_h\|^{2/(m-1)}} \right)^{-1}, g = 1, \ldots, G,$$

where $m > 1$.

- m represents the degree of uncertainty over assignment.
- As $m \to 1$, $J_m(P, \mu, \phi) \to L_1(P, \psi)$: HKM limiting case!

From two steps to one

HKM and FCM are generally applied iteratively in practice:

1. Calculate weights using parameter estimates
2. Estimate parameters for each group
3. Repeat until convergence.

This makes it difficult to characterize asymptotic properties.

From two steps to one

HKM and FCM are generally applied iteratively in practice:

1. Calculate weights using parameter estimates
2. Estimate parameters for each group
3. Repeat until convergence.

This makes it difficult to characterize asymptotic properties.

Yang and Yu (1992) show FCM can be rewritten without weights:

$$
\begin{aligned}
J_m\left(P, \mu\left(\phi\right), \phi\right) &= L_m\left(P, \phi\right) \\
&= \int \left(\sum_{g=1}^{G} \left\| y - \phi_g \right\|^{-2/(m-1)} \right)^{1-m} P\left(dy\right).
\end{aligned}
$$

This is just another nonlinear objective function!

Our regression problem

- We study the model

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

where $y_i \in R^T$, $x_i \in R^k$, and θ_g is a $T \times k$ matrix.

Our regression problem

- We study the model

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

where $y_i \in R^T$, $x_i \in R^k$, and θ_g is a $T \times k$ matrix.

- Accordingly, we consider the objective function

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

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

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

Asymptotics (and other theory)

1. Introduction
2. The FCM algorithm
3. Asymptotics (and other theory)
4. Simulations & performance
5. Heterogeneity in MPCs

Equivalence

Our theoretical development extends Yang and Yu (1992); Yang (1994).

Assumption 1. *Observations* (y_i, x_i) *are generated according to (1), jointly i.i.d. with probability measure* Π*,* G *is finite, and* $E[\varepsilon_i \mid x_i] = 0$*.*

Theorem 1. (Equivalence) *Under Assumption 1,* $J_m^{reg}(\Pi, \mu^{reg}, \theta) = L_m^{reg}(\Pi, \theta)$*, where*

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

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

We henceforth focus on the one-step objective function L_m^{reg}.

Existence

Assumption 2. *The second moments of y and x are finite under Π, x is not collinear, and Θ is compact.*

Theorem 2. (Existence) *If Assumptions 1-2 hold, then for any $G = 1, 2, \ldots,$ there exists a solution θ^* such that*

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

How does θ^* related to a true value, θ_0 in equation (1)?

- Our objective function is technically misspecified, so θ^* is a pseudo-true parameter.
- We'll see these pseudo-true estimates can be better than estimates from HKM!
- HKM is *well-specified* but minimizer is pseudo-true for small T.

GMM

Theorem 3. (Moments) *The solution θ^* satisfies the moment equations*

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

for $g=1,\ldots,G$ and $t=1,\ldots,T$, where t indexes dimensions of y_i and rows of θ_g; FCM is a GMM problem.

- This means that we can apply standard GMM results to characterize the FCM algorithm.
- Define $\hat{\theta}$ as the solution to

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

Consistency

Assumption 3. *θ^* is the unique solution to $E\left[\rho\left(\theta, y_i, x_i\right)\right] = 0$.*

We can easily argue this in limiting cases (just OLS); otherwise, less obvious.

Theorem 4. (Consistency) *Under Assumptions 1-3, $\hat{\theta} \xrightarrow{p} \theta^*$ as $N \to \infty$.*

Asymptotic normality (1)

Assumption 4. *Additionally*

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

Theorem 5. (Asymptotic Normality) *Under Assumptions 1 - 4,* $\sqrt{N}\left(\hat{\theta} - \theta^*\right) \xrightarrow{d} \mathcal{N}\left(0, H^{-1} V H^{-1}\right)$, *where*

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

and H *is the Hessian of (2).*

Asymptotic normality (2)

- We provide explicit expressions for this Hessian.
- We also provide expressions for two special cases:
  - Coefficients are common across groups
  - Coefficients constant within groups across time when $T > 1$
- Our results never assume that the groups are well-separated - a challenge posed by Bonhomme and Manresa (2015).
- In our empirical application, it is doubtful that the groups are well-separated for higher G.

Two-stage least squares

We extend all of our results to TSLS. Let x^e be endogenous regressor, ω exogenous, with instrument z.

- Estimate homogeneous first stage:

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

- Use predicted values in FCM objective function: only exploit exogenous variation.
- Moments can be stacked into a one-stage problem to derive GMM asymptotics.
- Could also use FCM for first stage, stacking x^e into y.
- ML techniques have primarily been employed on *first-stage* (e.g., Chernozhukov et al (2010)), but we show a novel way they can be exploited in the *second-stage*.

Selecting G

Choosing G is a difficult problem in clustering in general.

- More groups, better fit! (perfect as $G \to N$)
- Most model selection approaches (BIC, AIC...) are based on large T.
- FCM objective does not align with a standard likelihood.

We adapt the non-parametric Gap Statistic of Tibshirani et al. (2001):

- Compute fit (within group weighted sum of squares) of $G \leq G^{max}$ group model in a homogeneous reference distribution.
- Find G that maximizes the difference in fit between reference distribution and actual data.

Gap Statistic "normalizes" for the fact that fit always improves with G since fit also improves in the reference distribution.

Taking stock

We've shown that results derived for FCM in the cluster mean case extend to regression problems, as well as IV.

- Analytical inference is easy.
  - formulas for both heterogeneous and homogeneous coefficients
- FCM can be implemented as GMM.
  - delivers computational advantage
- The approach is well-suited to any clustering problem.
  - short panels and cross-sectional data
- Asymptotic properties hold even when data is not well-separated.
  - pseudo-true parameters may differ from truth.

Simulations and performance

1. Introduction
2. The FCM algorithm
3. Asymptotics (and other theory)
4. Simulations & performance
5. Heterogeneity in MPCs

Computational improvements

We believe we are the first to exploit the Yang and Yu (1992) equivalence result to support a one-stage (GMM) estimator.

- This estimator is considerably faster than standard iterative procedure.
- This estimator finds global minima more reliably (fewer start values).

Computational improvements

We believe we are the first to exploit the Yang and Yu (1992) equivalence result to support a one-stage (GMM) estimator.

- This estimator is considerably faster than standard iterative procedure.
- This estimator finds global minima more reliably (fewer start values).

Bonhomme and Manresa (2015) propose two HKM algorithms.

- Their full algorithm (2) is infeasible in "medium data" as it scales rapidly with N.
- We adopt a mixed version of their simpler algorithm with a partial version of algorithm 2.

We characterize FCM as "flexible, feasible, and fast".

Simulation study

We calibrate our simulations to our empirical application:

- 2008 tax rebate: households receive rebate of value $\sim$\$1000 with random timing, $N \approx 17,000$.
  - we observe cross-section of households over 16-month period.
- Baseline: regress quarterly change in total consumption on rebate value and a constant, with group-specific coefficients.
  - common coefficients on time dummies, simple covariates
- We consider both $G = 5$ and $G = 10$.
- We consider Gaussian and empirical noise distributions.
- We examine distribution of point estimates and rejection rates of tests.

$G = 5$, Gaussian noise

Truth	FCM		HKM	
	$m = 1.8$	$m = 1.1$	Algo. 1	Algo. 1.5
0.651	0.650	0.650	0.648	0.647
0.423	0.422	0.420	0.424	0.424
0.245	0.245	0.248	0.247	0.247
0.516	0.514	0.516	0.515	0.515
0.289	0.291	0.291	0.289	0.288

Table: Mean point estimates across $S = 500$ samples

$G = 5$, Gaussian noise

Truth	FCM		HKM	
	$m = 1.8$	$m = 1.1$	Algo. 1	Algo. 1.5
0.651	0.054	0.054	0.053	0.053
0.423	0.025	0.025	0.026	0.026
0.245	0.017	0.016	0.017	0.017
0.516	0.023	0.023	0.025	0.024
0.289	0.052	0.052	0.054	0.053

Table: RMSE across $S = 500$ samples

$G = 5$, Gaussian noise

Truth	FCM		HKM	
	$m = 1.8$	$m = 1.1$	Algo. 1	Algo. 1.5
0.651	0.056	0.060	0.032	0.032
0.423	0.048	0.040	0.020	0.020
0.245	0.060	0.048	0.040	0.040
0.516	0.036	0.040	0.020	0.020
0.289	0.040	0.036	0.068	0.064
	0.036	0.044	0.016	0.012

Table: Rejection rates, nominal size 0.05, across $S = 500$ samples

$G = 5$, empirical noise

Truth	FCM		HKM	
	$m = 1.8$	$m = 1.1$	Algo. 1	Algo. 1.5
0.651	0.723	0.583	0.695	0.638
0.423	0.348	0.222	0.308	0.263
0.245	0.238	0.161	0.188	0.174
0.516	0.501	0.358	0.397	0.375
0.289	0.306	0.189	0.203	0.183

Table: Mean point estimates across $S = 500$ samples

$G = 5$, empirical noise

Truth	FCM		HKM	
	$m = 1.8$	$m = 1.1$	Algo. 1	Algo. 1.5
0.651	0.802	0.802	0.795	0.805
0.423	0.315	0.447	0.392	0.429
0.245	0.106	0.148	0.138	0.147
0.516	0.472	0.431	0.413	0.416
0.289	0.481	0.479	0.475	0.478

Table: RMSE across $S = 500$ samples

$G = 10$, Gaussian noise

Truth	FCM		HKM	
	$m = 1.8$	$m = 1.1$	Algo. 1	Algo. 1.5
0.844	0.802	0.748	-4.876	-0.125
0.986	1.025	4.090	13.122	7.429
0.795	0.797	0.802	-1.029	1.287
0.646	0.646	0.402	-0.344	0.605
0.496	0.490	0.370	0.287	0.306
0.468	0.478	0.500	0.741	0.354
0.496	0.479	0.319	0.397	0.839
0.268	0.267	0.339	0.381	0.348
0.340	0.341	0.492	0.067	0.164
0.257	0.265	0.506	0.440	0.483

Table: Mean point estimates across $S = 500$ samples

$G = 10$, Gaussian noise

Truth	FCM		HKM	
	$m = 1.8$	$m = 1.1$	Algo. 1	Algo. 1.5
0.844	0.685	5.888	10.373	9.024
0.986	0.803	18.294	20.452	16.586
0.795	0.032	3.833	6.864	5.243
0.646	0.021	2.549	4.379	3.289
0.496	0.119	1.665	1.896	1.772
0.468	0.139	3.520	2.215	2.142
0.496	0.151	1.739	2.139	2.047
0.268	0.017	0.327	0.475	0.597
0.340	0.039	0.840	1.332	1.371
0.257	0.061	0.403	0.418	0.459

Table: RMSE across $S = 500$ samples

Heterogeneity in MPCs

1. Introduction
2. The FCM algorithm
3. Asymptotics (and other theory)
4. Simulations & performance
5. Heterogeneity in MPCs

2008 tax rebate

- Receipt timing determined by final two digits of recipient's Social Security Number: random assignment.
- CEX data:
  - Quarterly interviews, rotating panel of around 8k households, at most 4 quarters each.
  - Quarterly consumption reported; rebate data added to the survey.
- Identification: compare spending of households that received payments in a given period to the spending of households that did not.

Econometric Specification: homogeneous MPC

Parker et al. (2013): 12-30 percent of rebate spent on average on nondurables, 50-90 on total consumption.

$$\Delta C_j = \beta' W_j + \theta R_j + \alpha + \epsilon_j$$

- R_j: amount of rebate received by household i at time t
- W_j: a vector of demographic controls and month dummies

Econometric Specification: heterogenous MPC

$$\Delta C_j = \beta' W_j + \sum_{g \in G} \theta_g I_{j \in g} R_j + \alpha + \epsilon_j$$

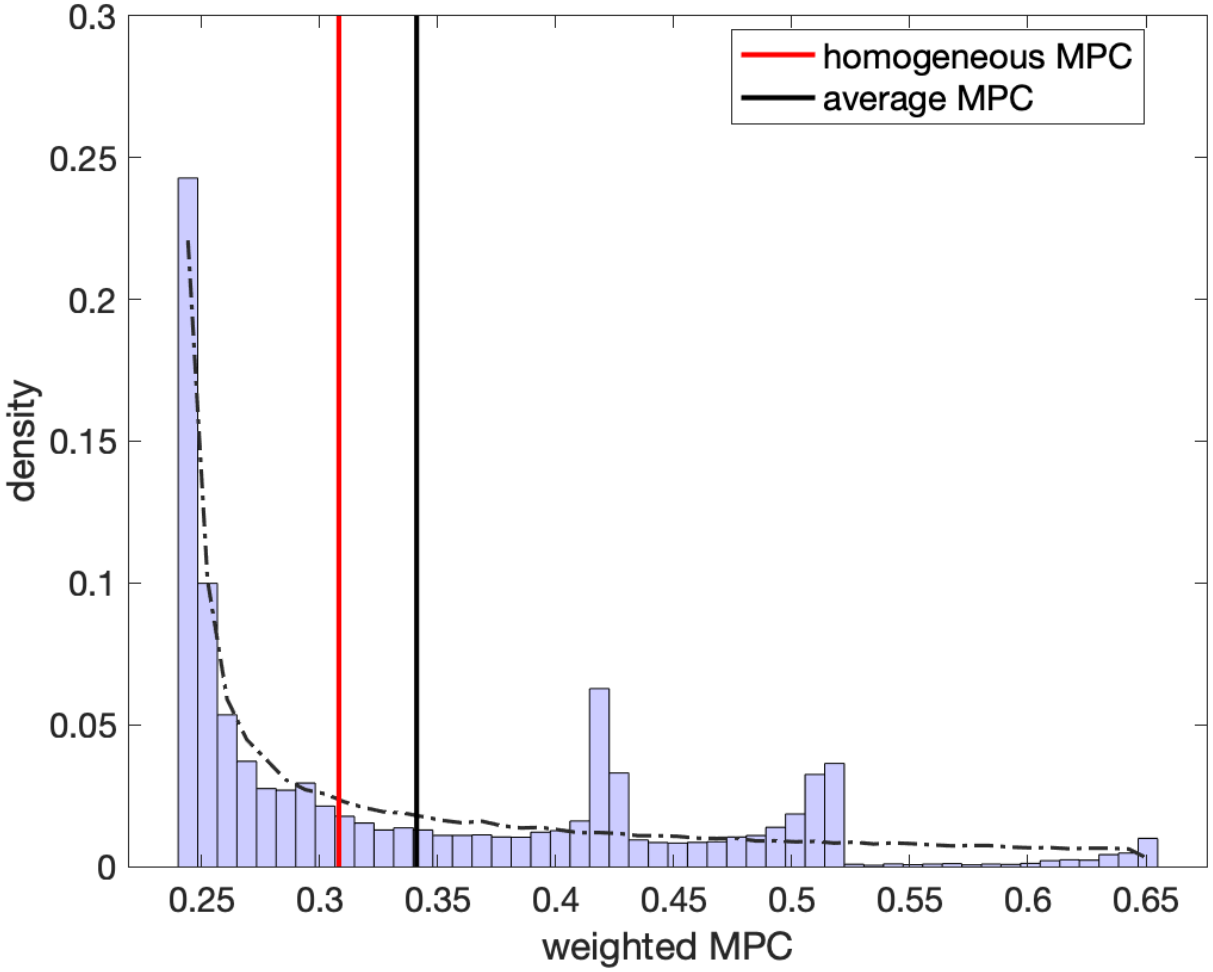
- R_j: amount of rebate received by household i at time t
- W_j: a vector of demographic controls and month dummies
- θ_g: MPC is group-specific
- $I_{j \in g}$: indicator function for observation j in group g

Econometric Specification: heterogenous MPC

$$\Delta C_j = \beta' W_j + \sum_{g \in G} (\theta_g I_{j \in g} R_j + \alpha_g I_{j \in g}) + \epsilon_j \tag{3}$$

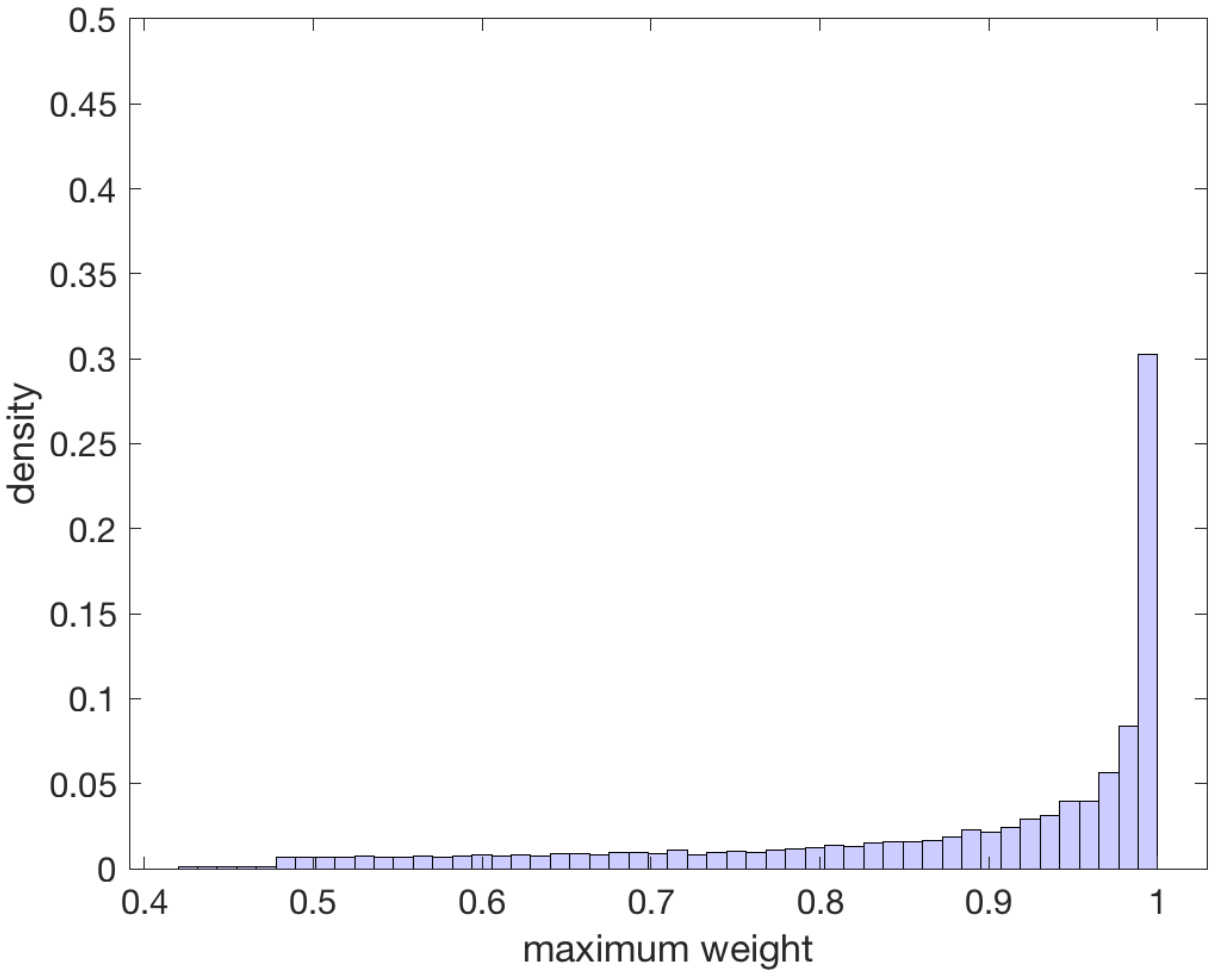
- R_j: amount of rebate received by household i at time t
- W_j: a vector of demographic controls and month dummies
- θ_g: MPC is group-specific
- $I_{j \in g}$: indicator function for observation j in group g
- α_g: group-specific fixed effects. Control for unobserved group-specific shocks that might spuriously correlate with MPC.
- And estimate this "hard" regression equation using FCM!

The MPC distribution: total consumption



gapstat mode

Distribution of maximum weights



What drives MPC heterogeneity?

Table 5: Explanatory variables: weighted MPC out of total expenditures

	OLS weighted MPC		TSLS weighted MPC	
	(I)	(II)	(III)	(IV)
Log salary income	0.002	-0.0003	0.003	0.001
	(0.02)	(0.001)	(0.030)	(0.002)
Log total income	0.036***	0.045***	0.081***	0.095***
	(0.008)	(0.006)	(0.015)	(0.012)
Mortgage interest to income ratio	0.086***	0.060**	0.141**	0.087*
	(0.032)	(0.027)	(0.061)	(0.052)
APC	0.044***	0.053***	0.111***	0.127***
	(0.008)	(0.007)	(0.016)	(0.013)
Outright homeowner dummy	0.006	0.008	0.006	0.009
	(0.011)	(0.010)	(0.021)	(0.019)
Mortgagor dummy	0.029***	0.023**	0.045**	0.035*
	(0.011)	(0.010)	(0.022)	(0.018)
Age	-0.002	-0.001	-0.004	-0.004
	(0.002)	(0.001)	(0.003)	(0.003)
Age-squared	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Married dummy	0.011	0.011	0.032**	0.028**
	(0.008)	(0.007)	(0.015)	(0.013)
Number of children	0.006	0.0001	0.006	-0.002
	(0.006)	(0.003)	(0.007)	(0.006)
Log liquid wealth	-0.001		-0.001	
	(0.001)		(0.002)	
R^2	0.13	0.13	0.16	0.15
Number of observations	723	1,079	723	1,079

What drives MPC heterogeneity?

Table 5: Explanatory variables: weighted MPC out of total expenditures

	OLS weighted MPC		TSLS weighted MPC	
	(I)	(II)	(III)	(IV)
Log salary income	0.002	-0.0003	0.003	0.001
	(0.02)	(0.001)	(0.030)	(0.002)
Log total income	0.036***	0.045***	0.081***	0.095***
	(0.008)	(0.006)	(0.015)	(0.012)
Mortgage interest to income ratio	0.086***	0.060**	0.141**	0.087*
	(0.032)	(0.027)	(0.061)	(0.052)
APC	0.044***	0.053***	0.111***	0.127***
	(0.008)	(0.007)	(0.016)	(0.013)
Outright homeowner dummy	0.006	0.008	0.006	0.009
	(0.011)	(0.010)	(0.021)	(0.019)
Mortgagor dummy	0.029***	0.023**	0.045**	0.035*
	(0.011)	(0.010)	(0.022)	(0.018)
Age	-0.002	-0.001	-0.004	-0.004
	(0.002)	(0.001)	(0.003)	(0.003)
Age-squared	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Married dummy	0.011	0.011	0.032**	0.028**
	(0.008)	(0.007)	(0.015)	(0.013)
Number of children	0.006	0.0001	0.006	-0.002
	(0.006)	(0.003)	(0.007)	(0.006)
Log liquid wealth	-0.001		-0.001	
	(0.001)		(0.002)	
R^2	0.13	0.13	0.16	0.15
Number of observations	723	1,079	723	1,079

What drives MPC heterogeneity?

Table 5: Explanatory variables: weighted MPC out of total expenditures

	OLS weighted MPC		TSLS weighted MPC	
	(I)	(II)	(III)	(IV)
Log salary income	0.002	-0.0003	0.003	0.001
	(0.02)	(0.001)	(0.030)	(0.002)
Log total income	0.036***	0.045***	0.081***	0.095***
	(0.008)	(0.006)	(0.015)	(0.012)
Mortgage interest to income ratio	0.086***	0.060**	0.141**	0.087*
	(0.032)	(0.027)	(0.061)	(0.052)
APC	0.044***	0.053***	0.111***	0.127***
	(0.008)	(0.007)	(0.016)	(0.013)
Outright homeowner dummy	0.006	0.008	0.006	0.009
	(0.011)	(0.010)	(0.021)	(0.019)
Mortgagor dummy	0.029***	0.023**	0.045**	0.035*
	(0.011)	(0.010)	(0.022)	(0.018)
Age	-0.002	-0.001	-0.004	-0.004
	(0.002)	(0.001)	(0.003)	(0.003)
Age-squared	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Married dummy	0.011	0.011	0.032**	0.028**
	(0.008)	(0.007)	(0.015)	(0.013)
Number of children	0.006	0.0001	0.006	-0.002
	(0.006)	(0.003)	(0.007)	(0.006)
Log liquid wealth	-0.001		-0.001	
	(0.001)		(0.002)	
R^2	0.13	0.13	0.16	0.15
Number of observations	723	1,079	723	1,079

Conclusion

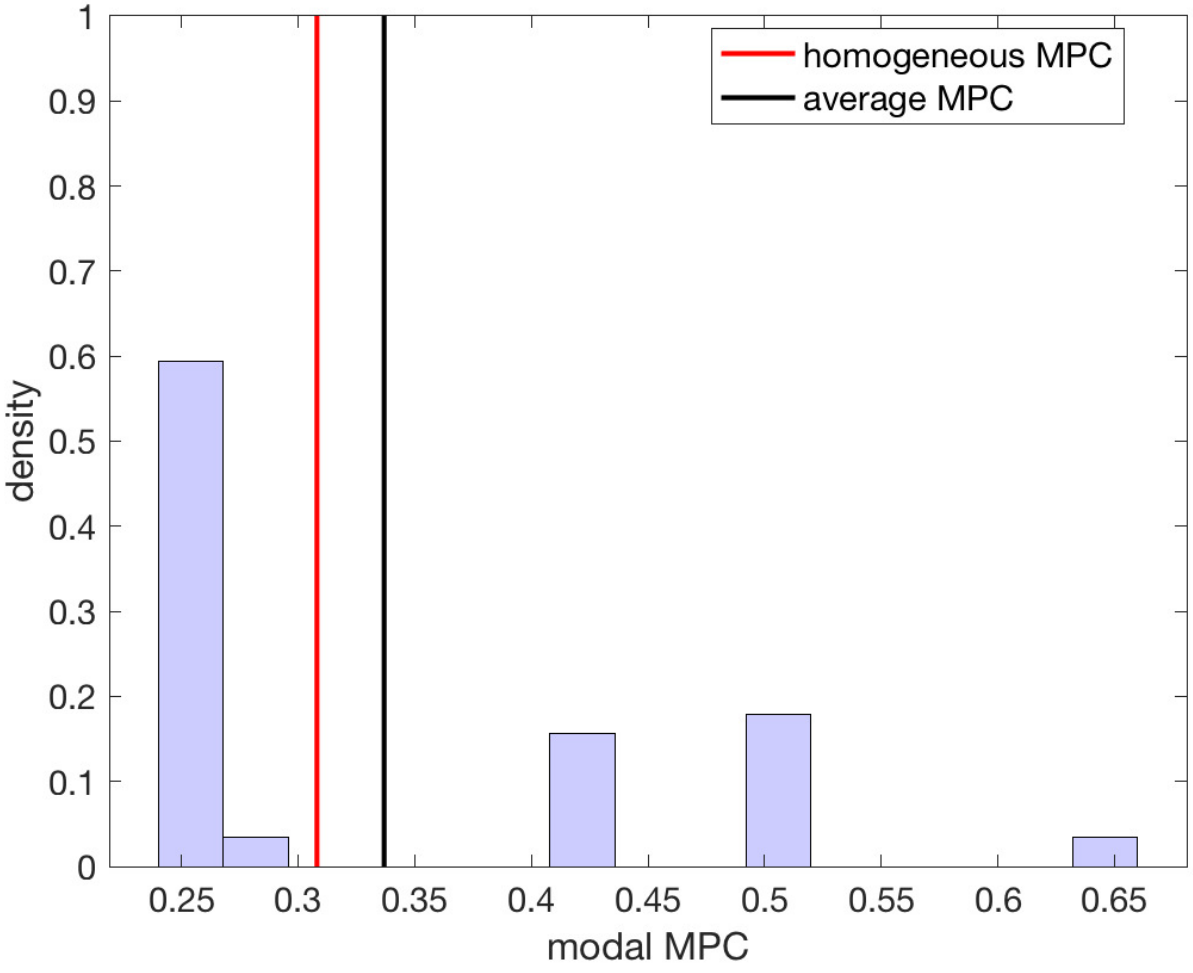
We find the FCM clustering approach to be better-suited to many economics contexts than alternatives

- We generalize it from cluster means to regression and IV.
- We derive desirable asymptotic properties.
- It performs very well in simulation.
- It has computational advantages.

We uncover considerable latent heterogeneity in the MPC.

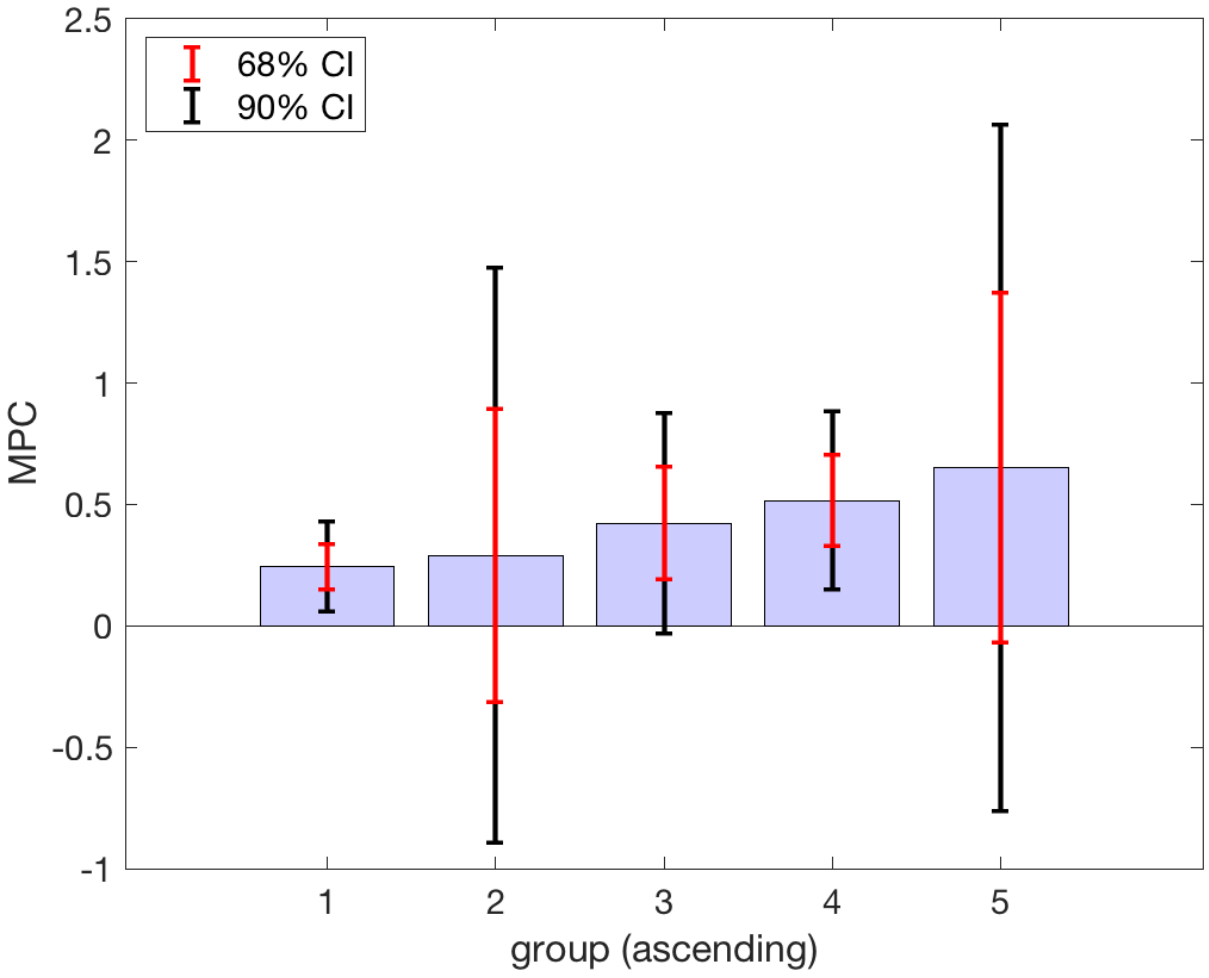
- We don't have to assume correlates to do so.
- Robust relationship with income, mortgages, past spending
- We find a positive correlation across spending categories.
- Recent theoretical work shows important policy implications.

The distribution of marginal propensity to consume

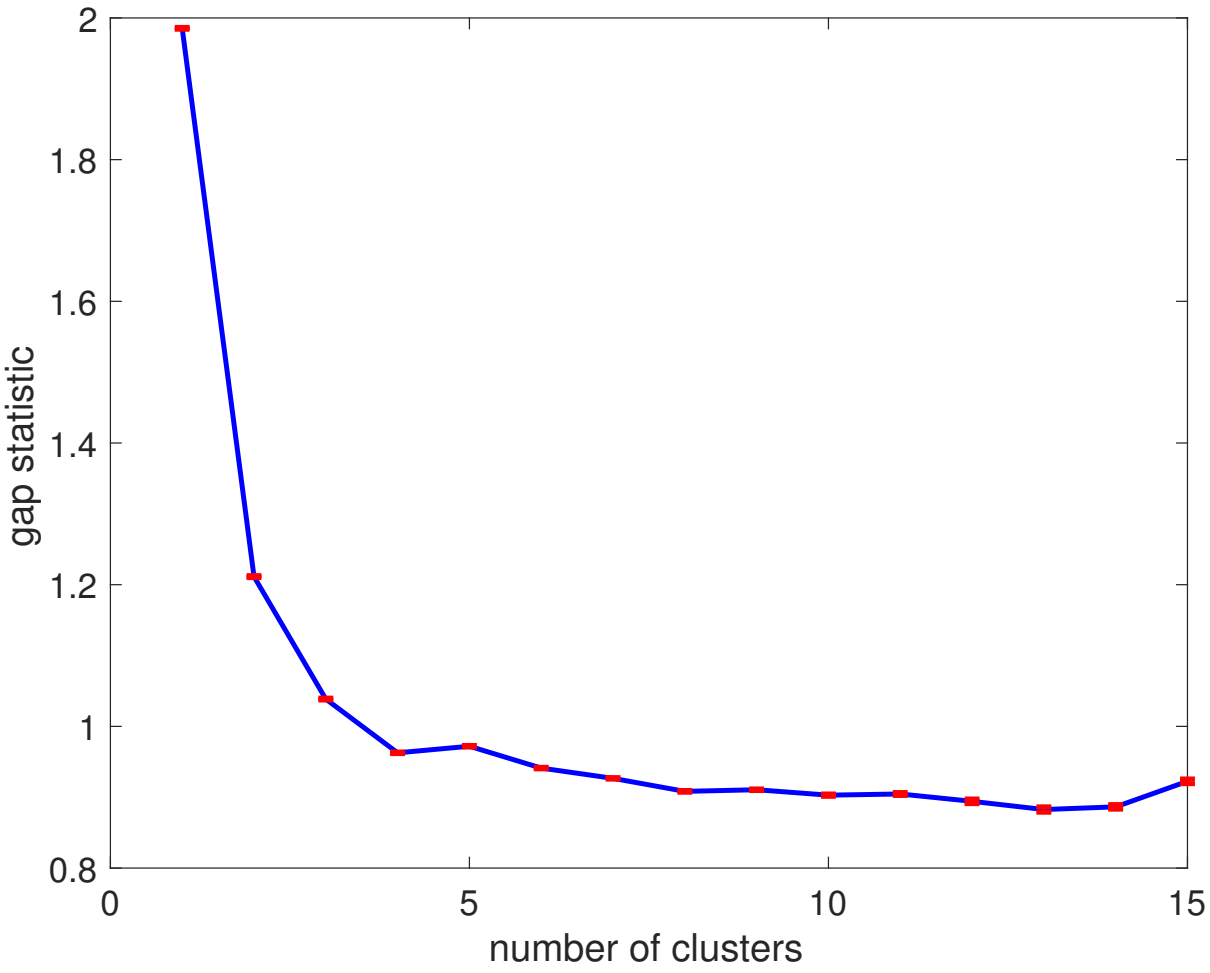


back

Inference on heterogeneous MPC

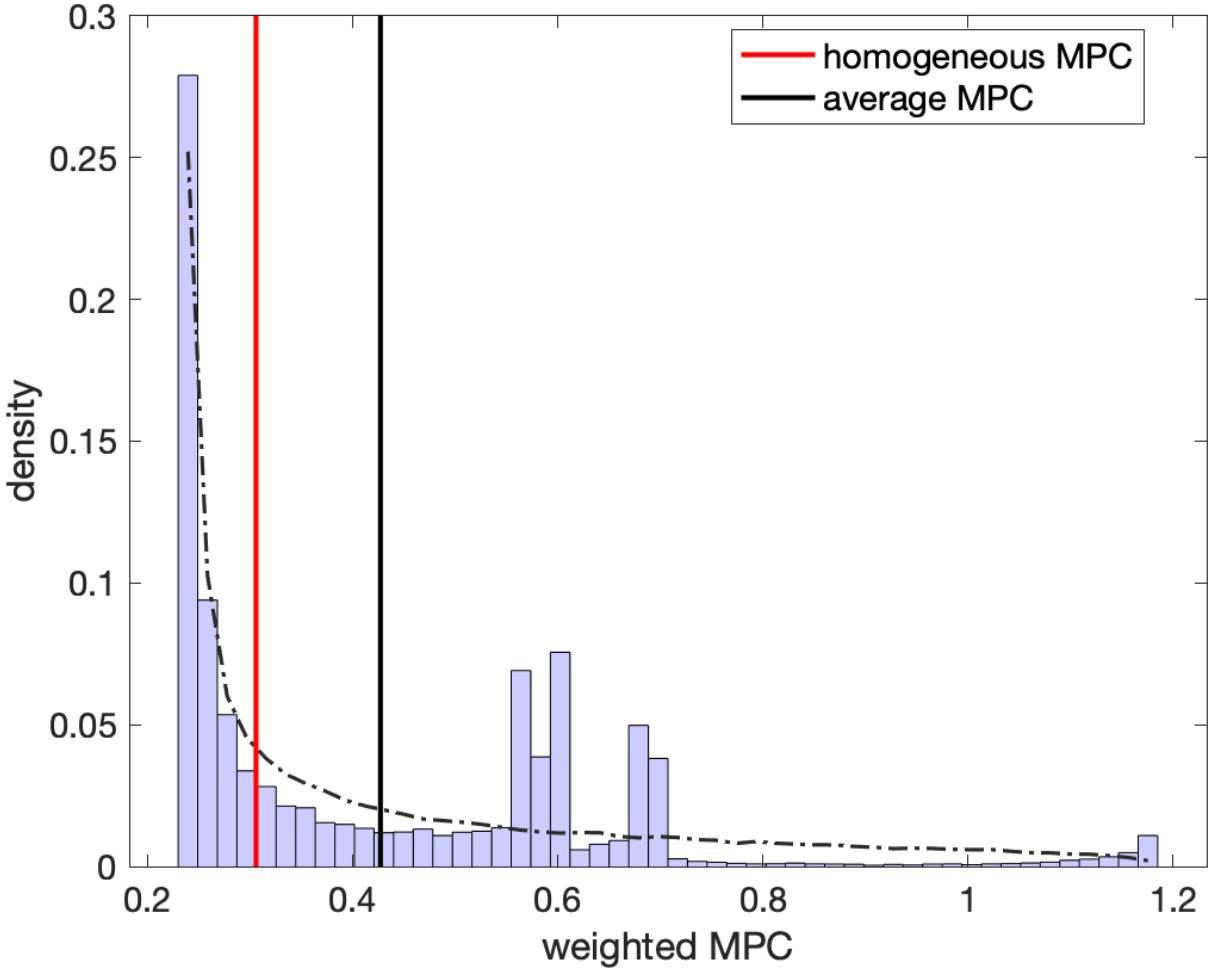


Gap statistic



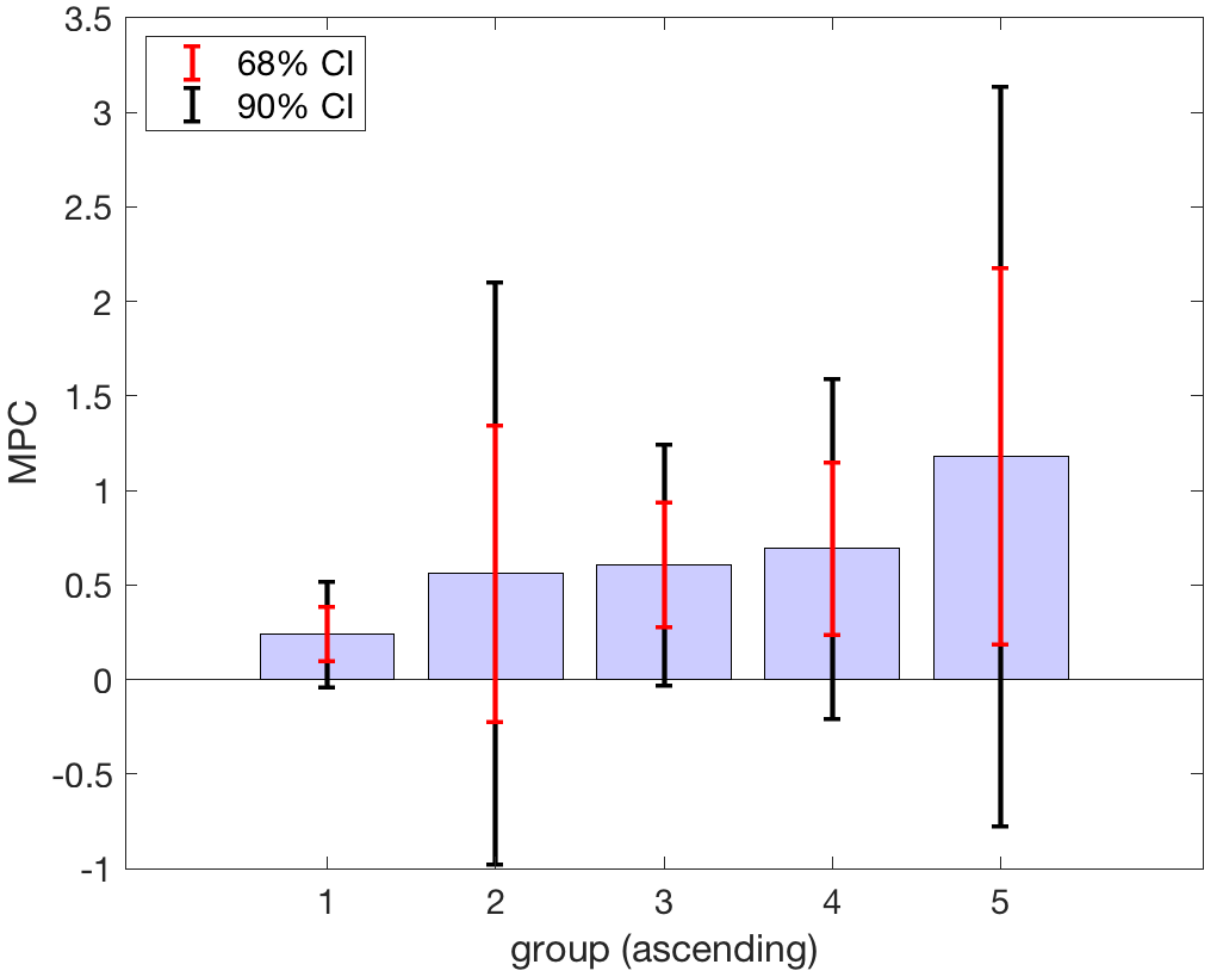
back

The MPC distribution: TSLS



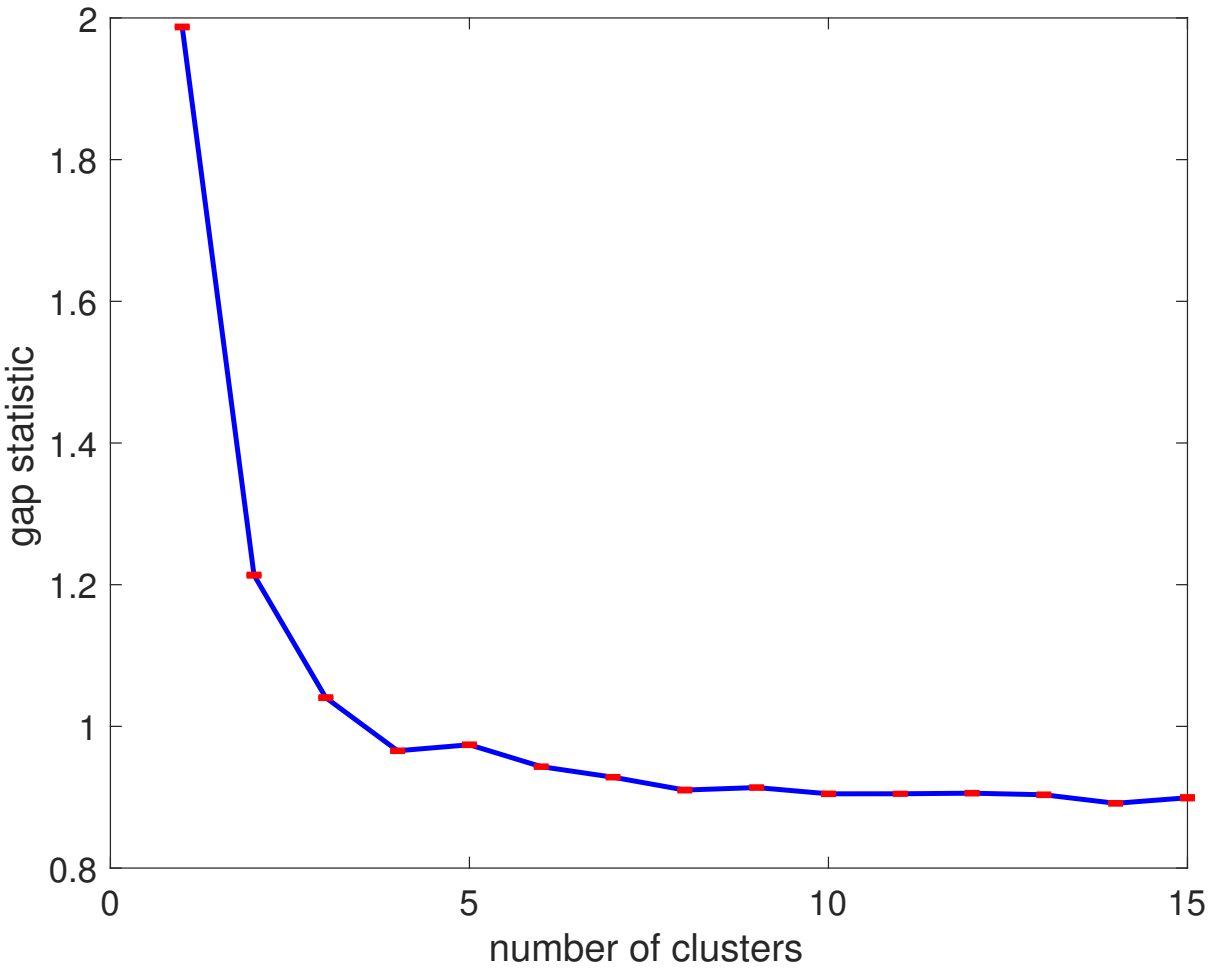
confidence bands gapstat drop never-rebaters

Inference on heterogeneous MPC: 2sls



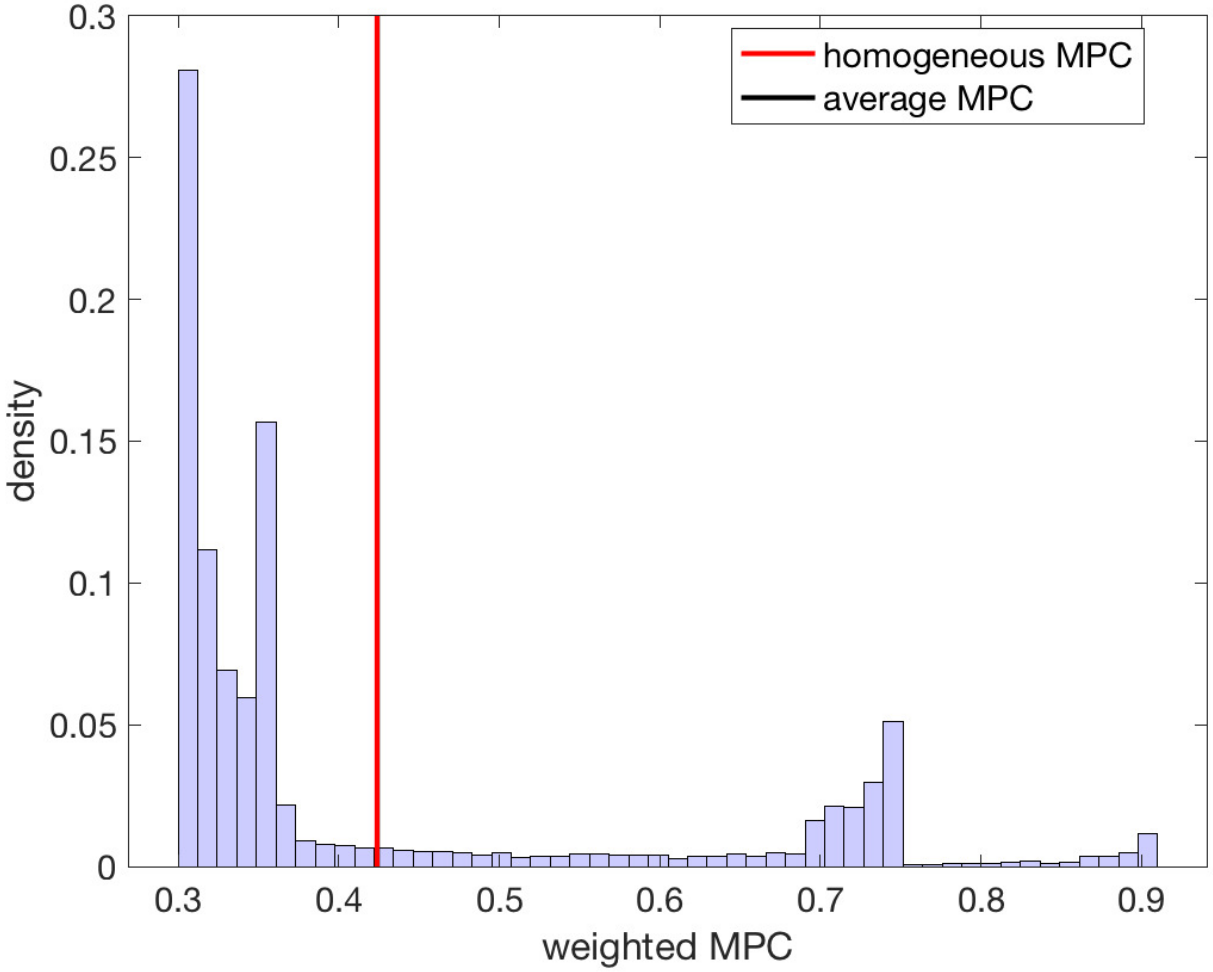
back

Gap statistic: TSLS



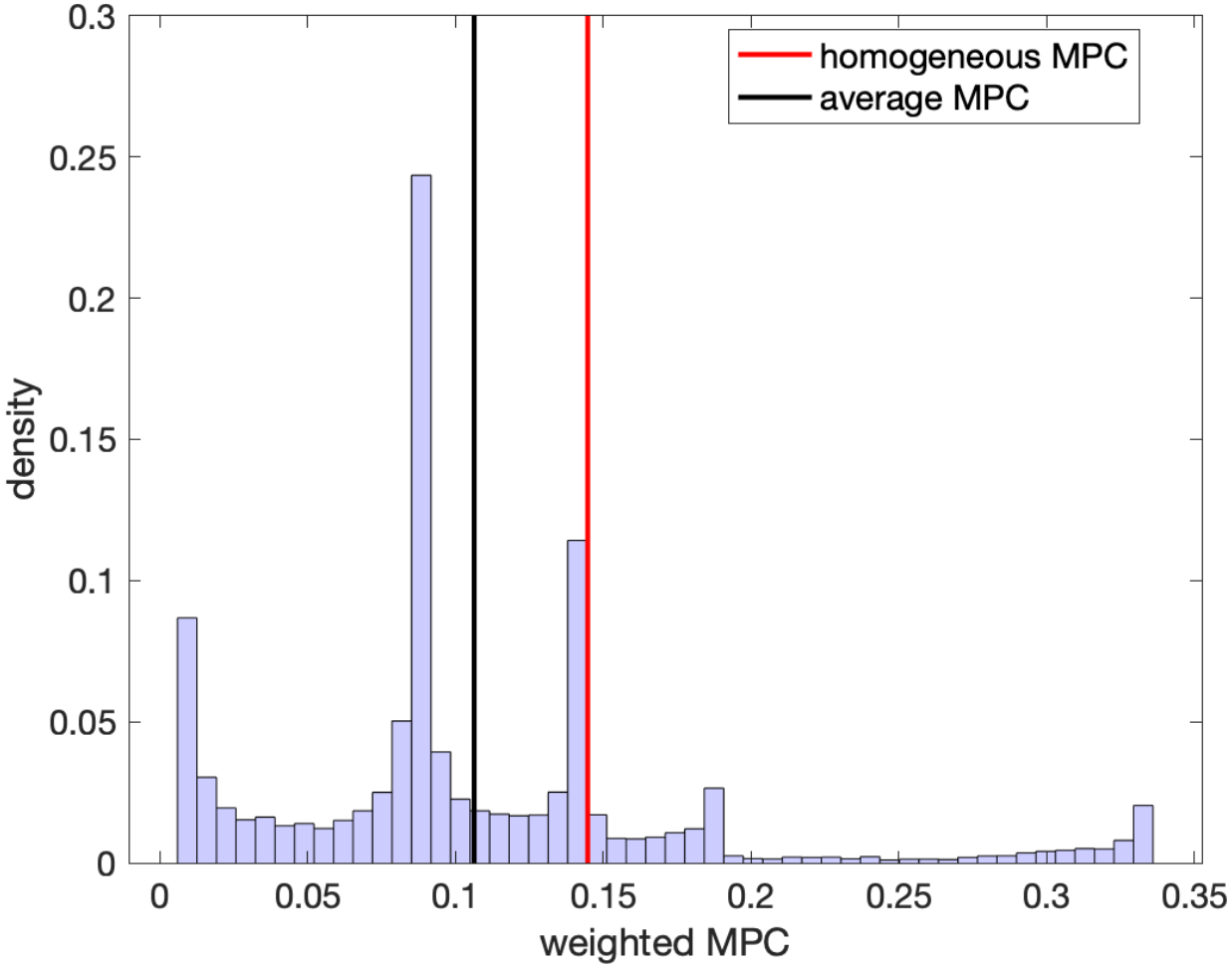
back

Drop households that never get the rebate



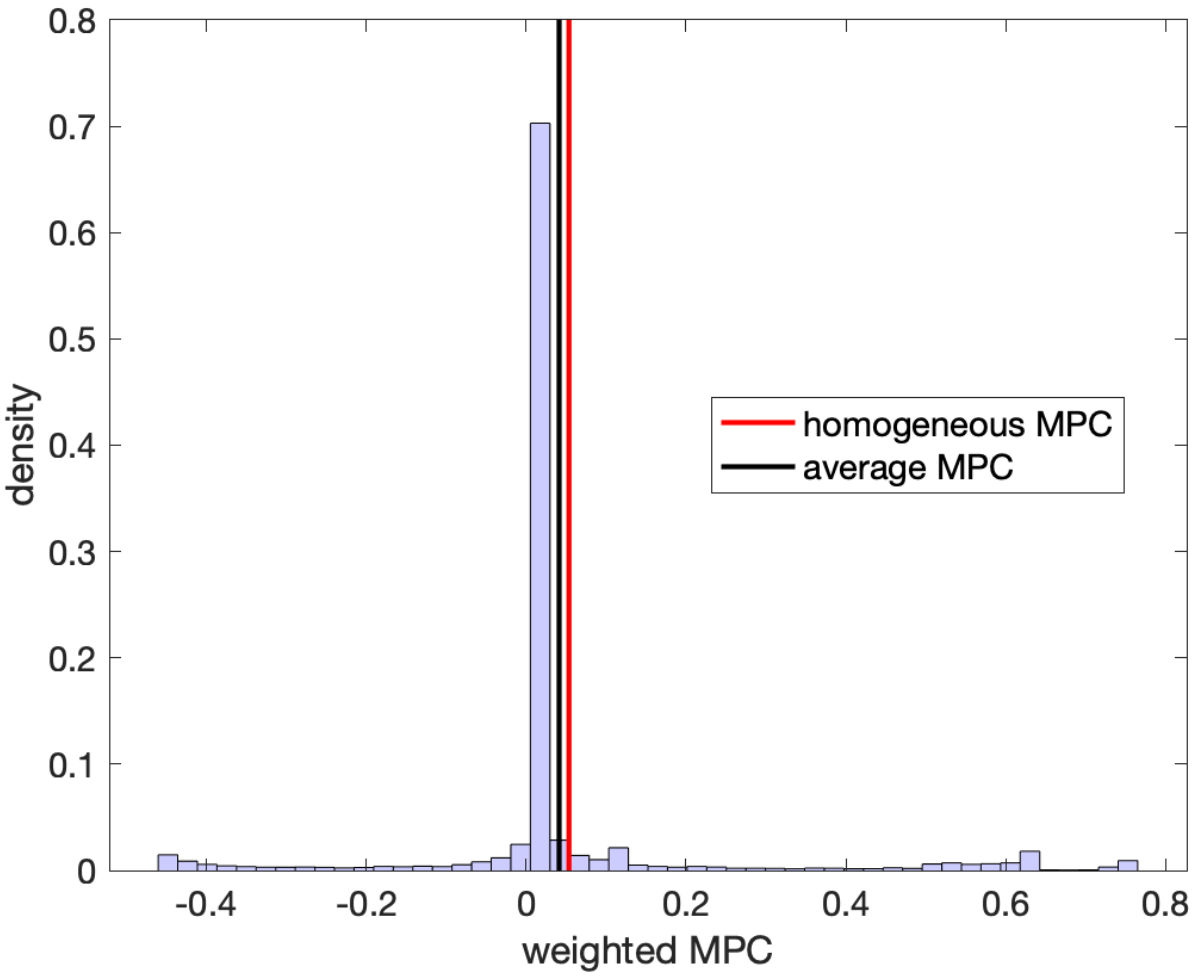
back

The MPC distribution: nondurables



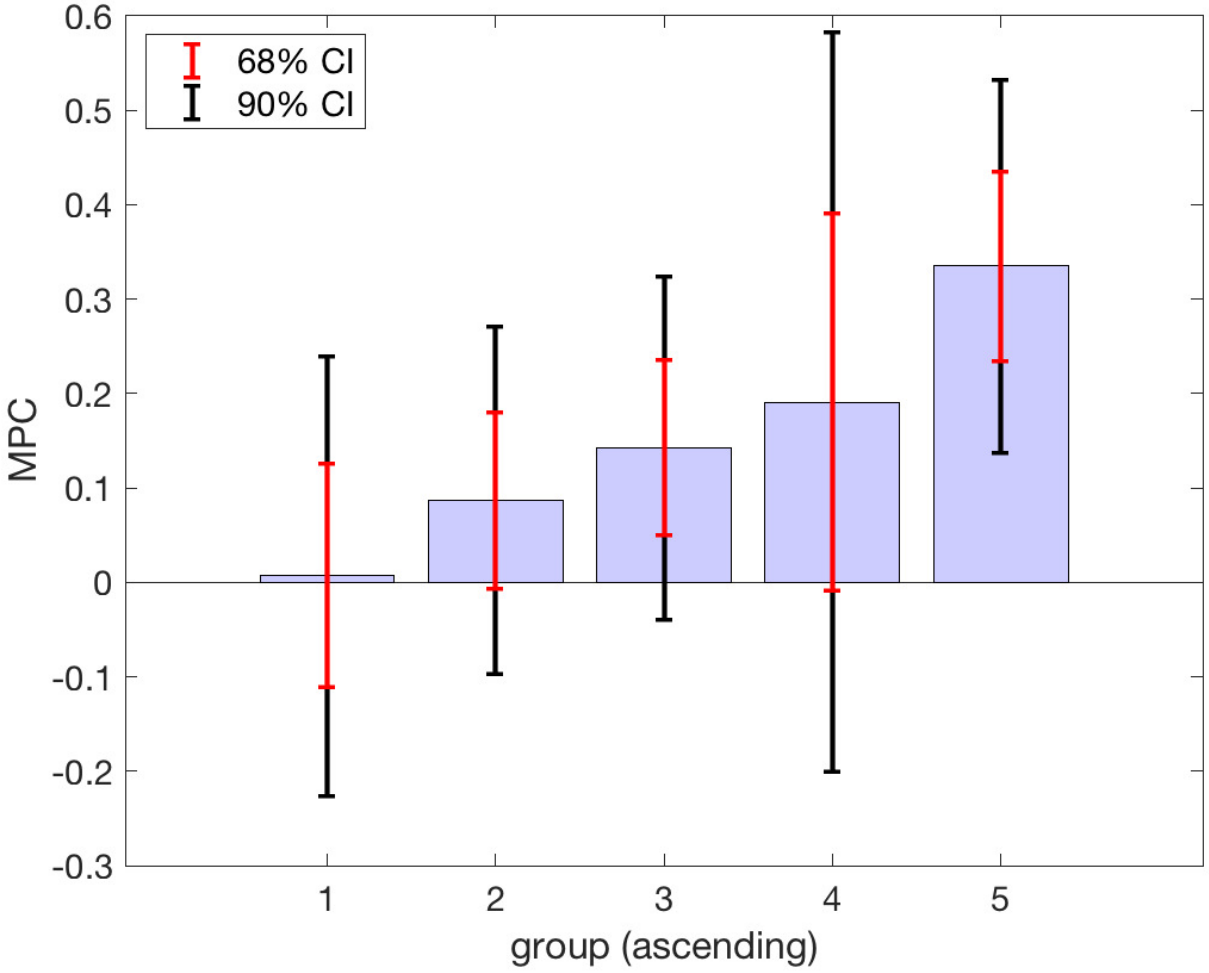
confidence bands distribution

The distribution of MPC: durables



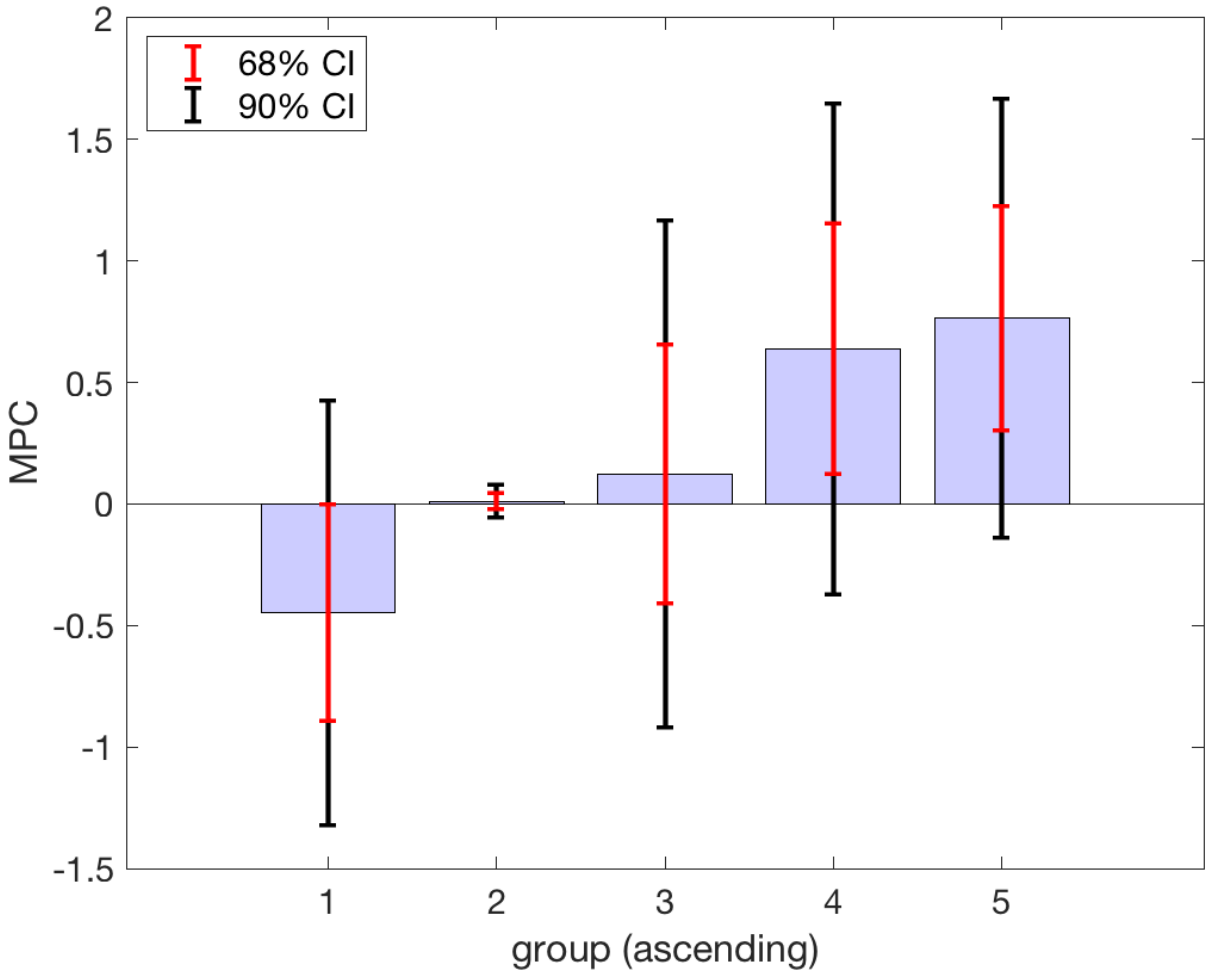
confidence bands distribution

Inference on heterogeneous MPC: non-durables



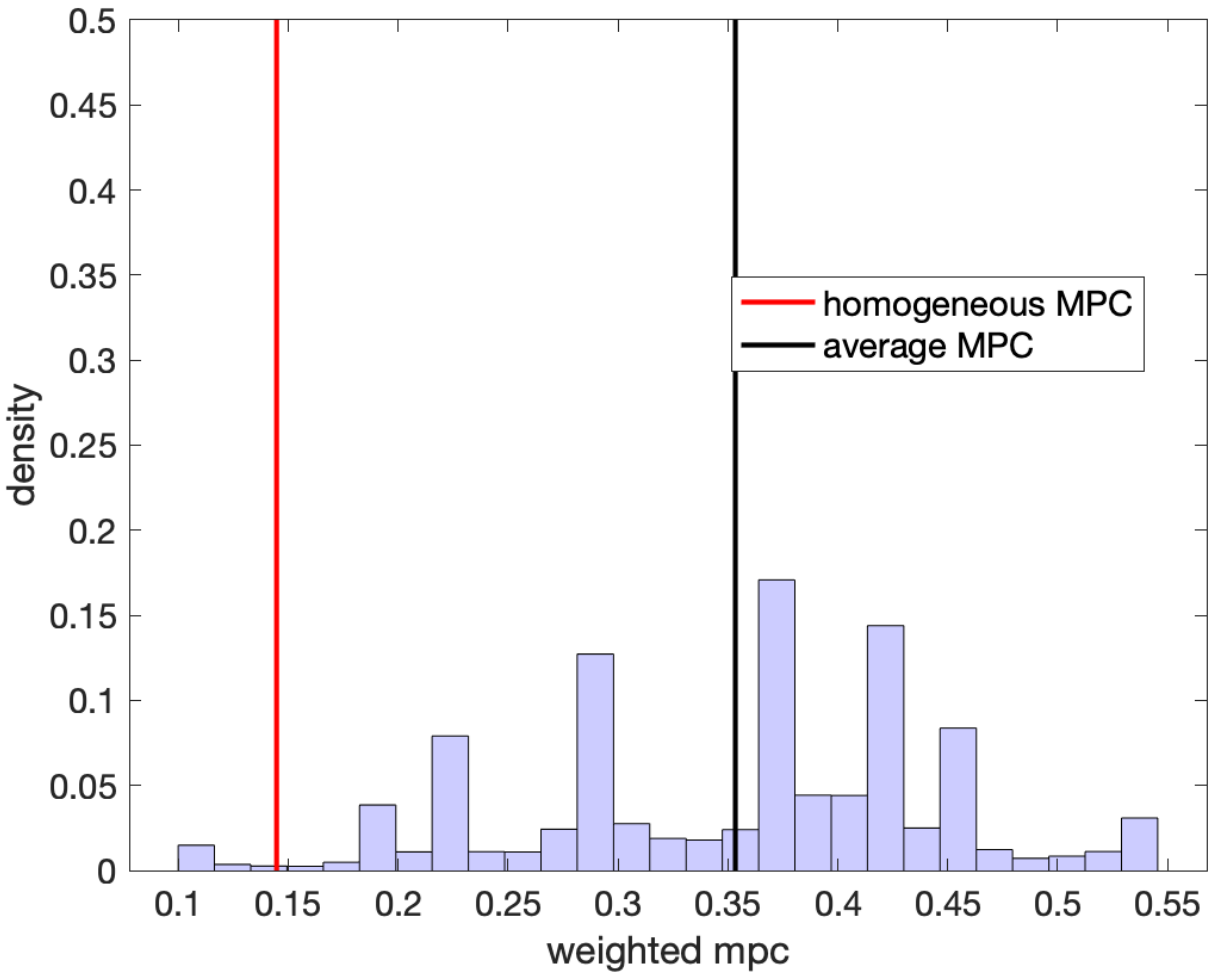
back

Inference on heterogeneous MPC: durables



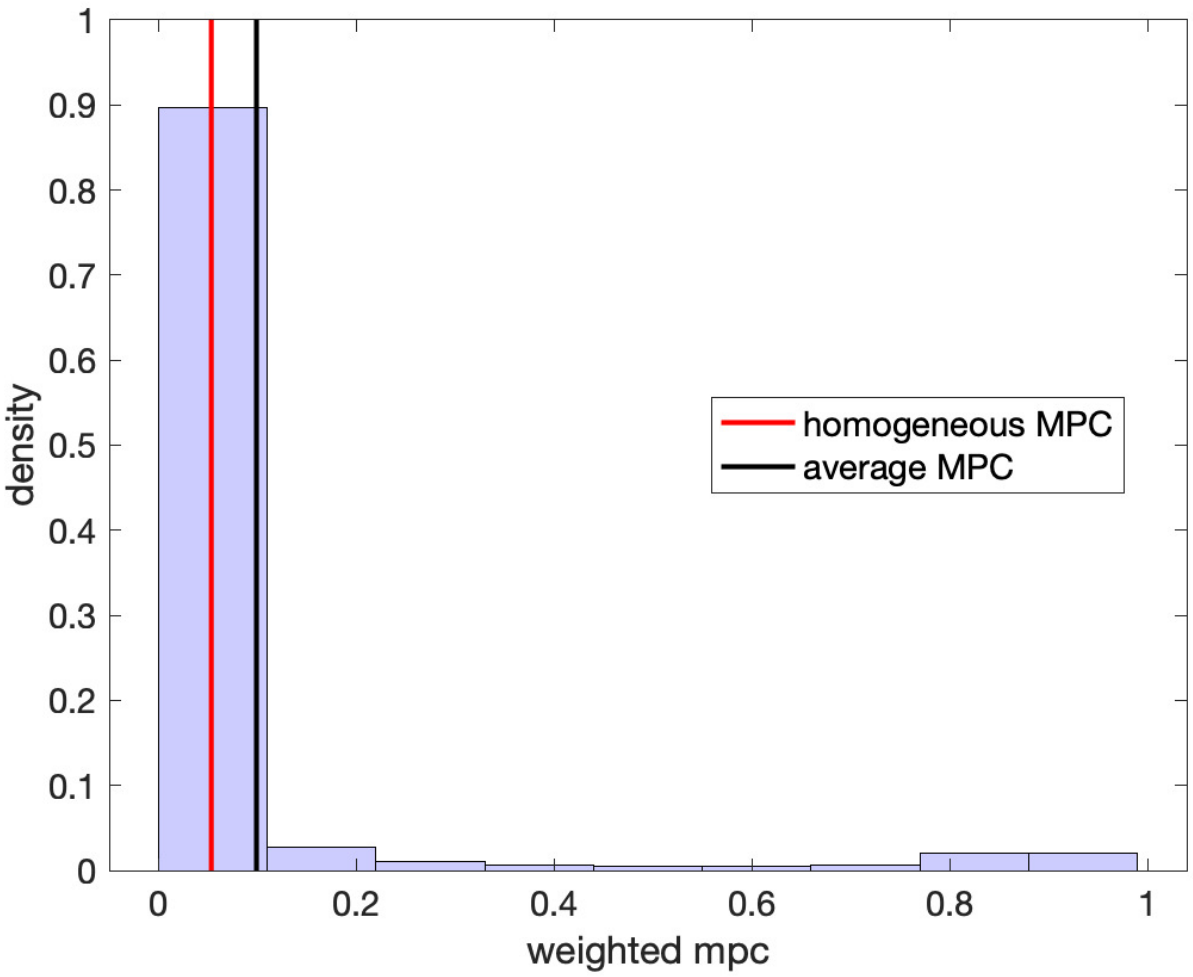
back

The distribution of MPC: nondurables at G*



back

The distribution of MPC: durables at G*



back

Additional results

Additional household characteristics unconditionally correlate with the MPC, but are not robust to inclusion of controls:

- Liquid wealth is positive correlated with MPC...
  - ...but liquid wealth to income ratio is not.
    - saving in an educational savings plan is associated with higher MPC (Campbell and Hercowitz (2018)).
  - Households receiving a larger rebate display greater MPCs (also in TSLS).
  - Married households have a larger MPC.

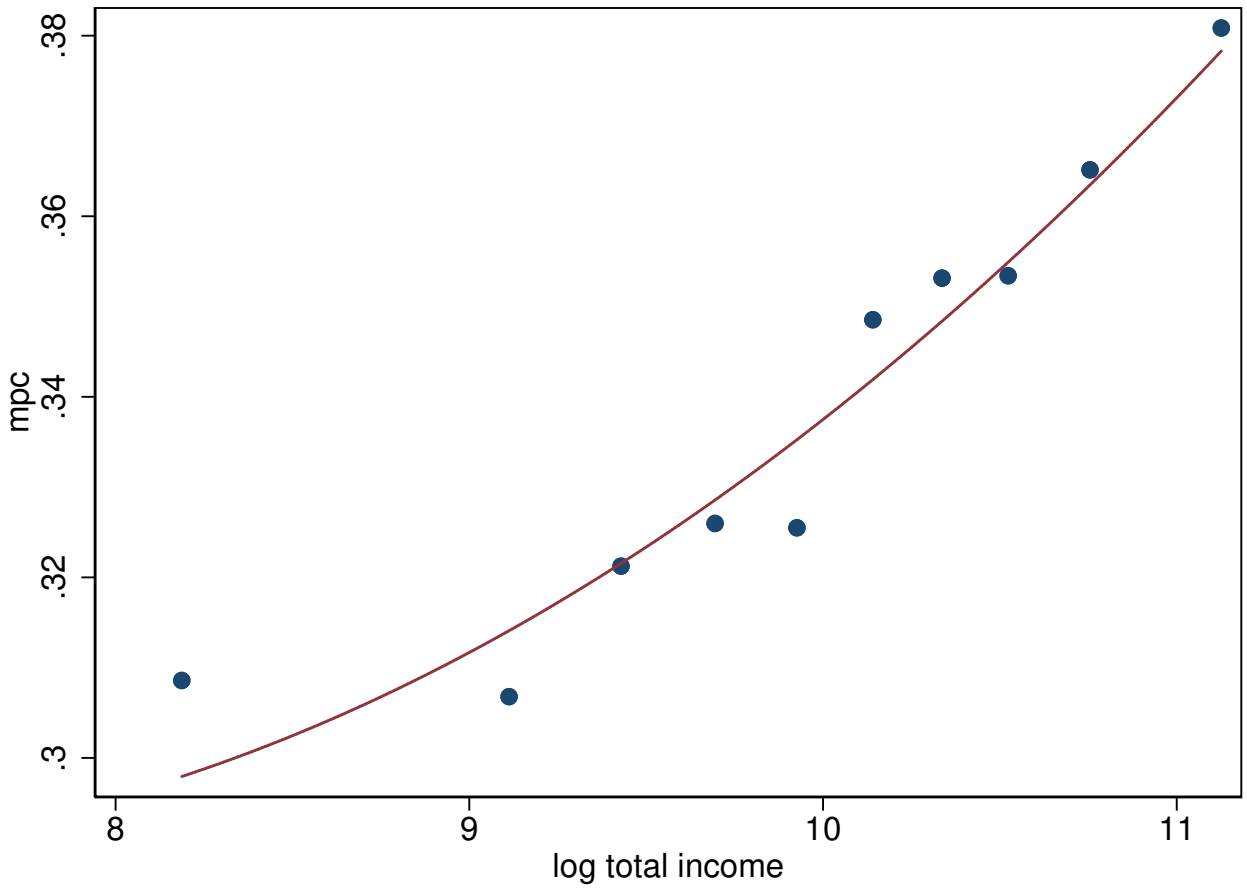
back

Robustness

- Results are confirmed with TSLS MPCs and modal MPC distribution
  - Most of the correlations with observables are confirmed with non-durable MPCs, except for mortgage.
  - Results are confirmed if we disregard households associated with a MPC that is not statistically different from 0.

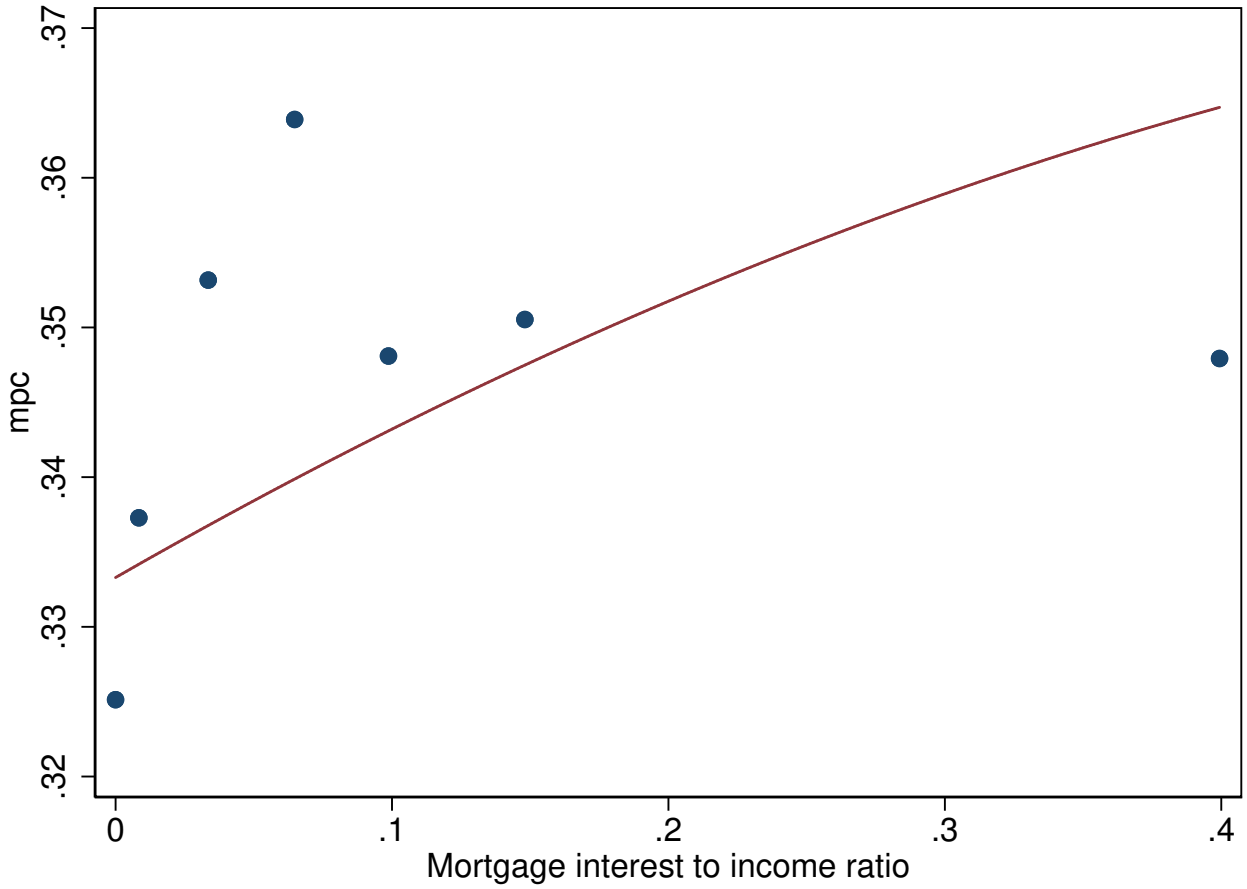
Total income and the MPC

- 1% in total income increases MPC by 0.02.
- Aligns with Shapiro and Slemrod (2010), Misra and Surico (2014).



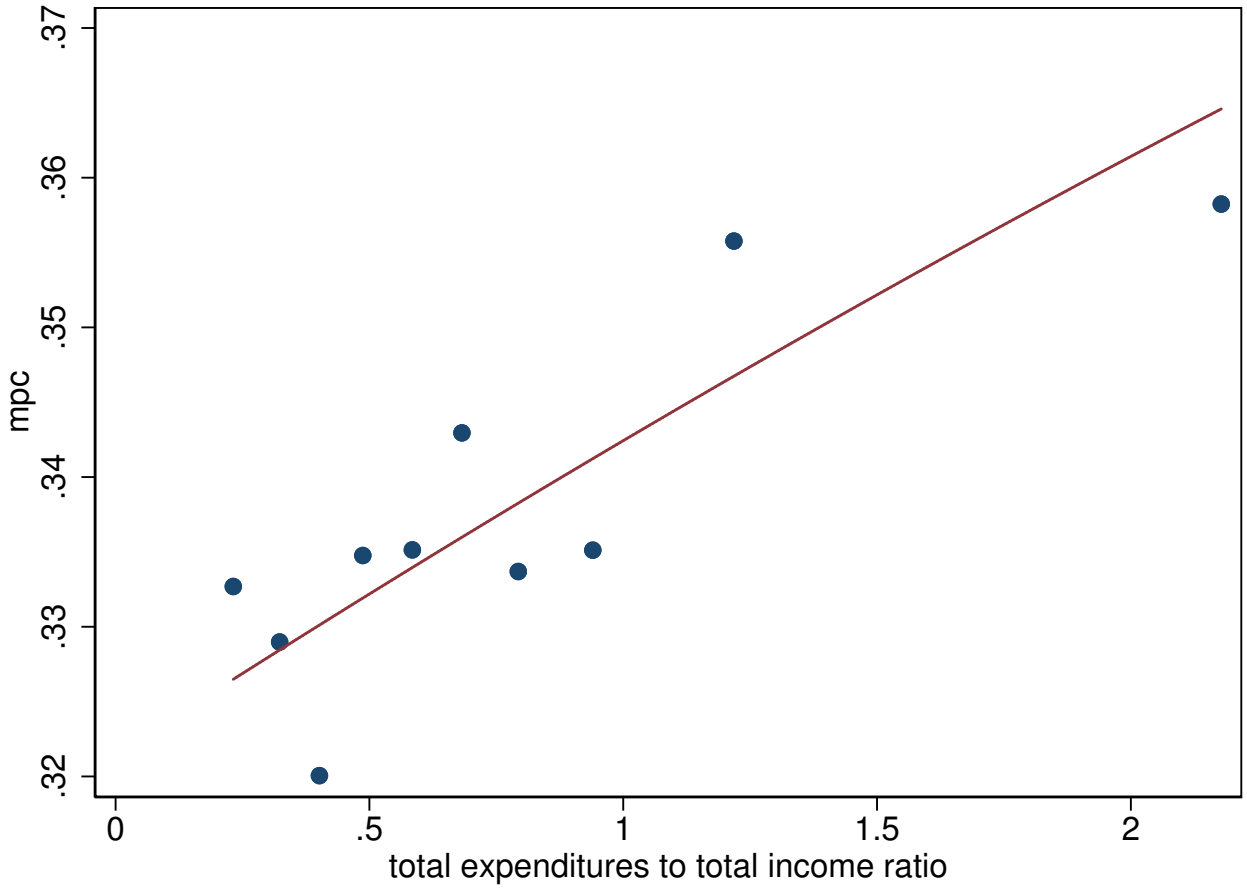
Mortgages and the MPC

- Households with a mortgage have a 0.03 higher MPC.
  - Parker et al. (2017), Misra and Surico (2014)
- 1 percentage point higher mortgage to income ratio increases MPC by 0.09.



APC and MPC

- APC is computed with lagged average expenditures to avoid mechanical correlation.
- Households spending 1 percentage point more of their income have a 0.04 higher MPC.
  - There appears to be a "spender" type.



APC and MPC

- Positive correlation is present at different ages and income levels.
- Life-cycle consumption savings model can generate the positive correlation:
  - Constrained households have APC $= 1$ and high MPC.
  - As they become less constrained, they build up assets (APC < 1) and have lower MPC.
  - Unconstrained households might have higher APC (no precautionary savings) and low MPC...
  - ...but effective discount factor can increase MPC

References

Bezdek, J. C., R. Ehrlich, and W. Full (1984): "FCM: The fuzzy c-means clustering algorithm," *Computers Geosciences*, 10, 191 – 203.

Bonhomme, S., T. Lamadon, and E. Manresa (2017): "Discretizing unobserved heterogeneity," IFS Working Papers W17/03, Institute for Fiscal Studies.

Bonhomme, S. and E. Manresa (2015): "GROUPED PATTERNS OF HETEROGENEITY IN PANEL DATA," *Econometrica*, 83, 1147–1184.

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.

Lin, C.-C. and S. Ng (2012): "Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown," *Journal of Econometric Methods*, 1, 1–14.

Pollard, D. (1981): "Strong Consistency of K-Means Clustering," *Ann. Statist.*, 9, 135–140.

——— (1982): "A Central Limit Theorem for k-Means Clustering," *Ann. Probab.*, 10, 919–926.

Tibshirani, R., G. Walther, and T. Hastie (2001): "Estimating