Detecting Heterogeneity in the MPC: A Machine Learning Approach

Last Update: 01.10.2021

Master's Thesis (Draft)

Department of Economics

University of Mannheim

submitted to:
Prof. Krzyzstof Pytka, PhD

submitted by:
Lucas Cruz Fernandez

Student ID: *********
Studies: Master of Science Economics (M.Sc.)

Adress: ******
Phone: ******
E-Mail: ******

Mannheim, ADD DATE

1 Introduction

How do households respond to income shocks and how do their responses differ given their personal characteristics and economic circumstances? These questions are not only at the centre of a wide academic debate in economics but also of major importance for policy makers. While the former revolves around verifying or neglecting the main mechanisms of the Permanent Income Hypothesis (PIH), the latter are interested in improving government transfers to more efficiently use public funds. These two sides have sparked many investigations using a wide array of approaches to quantify hosueholds' responses to income shocks. This Marginal Propensity to Consume (MPC) - at the centre of macroeconomics since Keynes introduced it at the heart of his General Economic Theory - quantifies how much households will spend on consumption of each dollar they receive from an income shock. While research has long focused on quantifying whether the MPC out of income shocks is zero and thus in line with the PIH, the literature has seen a shift of focus over the last decade. On average most studies support the notion of a zero MPC, however, more recent evidence suggests that for specific groups the response is significantly different.

Empirical research related to the MPC and its heterogeneity has used several settings to identify income shocks. One of the most prominent is to use natural experiments in which households receive exogenous income shocks. Following Parker et al. (2013) and Misra and Surico (2014), we exploit the 2008 tax rebate in the USA to estimate households' MPC using data collected by the Consumer Expenditure Survey (CEX). Similar to these two prior studies, we are able to use the rich information on consumption the CEX provides to not only identify heterogeneities in the overall MPC but also to analyse which categories of consumption goods households spend their rebate on. However, our econometric approach sets us apart as it is more sophisticated and more precise in detecting heterogeneities compared to any contribution we are aware of.

Namely, the two main channels are life-cycle dynamics and liquidity. The former is driven by a consumer's age and the associated fluctuation in income. As data consistently shows (Attanasio and Weber, 2010), consumption expenditures follow a hump shape over the life-cycle rather than being roughly constant as we would expect under the PIH. This anomaly is often referred to as the retirement-consumption puzzle and connected to an increased amount of free time which allows households to reduce the cost of their consumption. Therefore, we would expect a reduction in the measured MPC in age.

In the case of liquidity, its role is linked to the nature of the income shock and borrowing constraints. If a positive income shock is anticipated, households that are already close to or at their borrowing constraint cannot borrow new funds to smooth consumption in anticipation of a higher future income. Thus, once the shock realises, we will observe an

increase in consumption. On the other hand, saving is always possible for any household and hence we will not see a reaction once the shock realises in case of a negative anticipated shock. Thus, more liquid households react less to a positive anticipated shock in comparison with liquidity constrained ones. In contrast, in case of an unanticipated shock, we expect the opposite. Think of an agent that is temporarily out of work and has no liquid wealth at their disposal. In case of a negative shock, the agent is forced to adjust their consumption behaviour downward. Meanwhile, a positive shock will always be saved and stretched over future periods, no matter the level of households' liquidity. E.g. Bunn et al. (2018) document this asymmetry depending on the sign of the shock.
It is important to highlight that in our setting, households experience an anticipated positive income shock. The Economic Stimulus Act was signed into law by President Bush in February 2008 and payments of tax rebates started in April of the same year. Therefore, following the previous arguments, theory would expect older households to react less, but liquidity not being a major driver. However, the tax rebate was disbursed to US taxpayers during a time of national and global economic downturn. Thus, many households receiving the stimulus might have been in economic turmoil when receiving the payment and actually spend it to cover regular expenses that they otherwise would not have been able to cover (e.g. rent or utilities). However, Parker et al. (2013) emphasise that some rebates were reported to be received outside of the disbursal window, which suggests that the income shocks might not have been anticipated and only noticed after their arrival.
One major issue in the existing literature is the way how heterogeneity is measured. Most studies rely on either splitting their sample into smaller sub samples and estimating the MPC within each sample or use dummy variables that are defined by the authors based on some continuous variable. These approaches suffer from the severe issue that any heterogeneity that does not fall into this pre-defined pattern is not captured and will muddy the results of these investigations. In the worst case these procedures miss to pick up existing heterogeneity or missing patterns within these pre-defined subgroups. On the contrary, our Double Machine Learning (DML) approach allows us to estimate the conditional MPC out of the tax rebate of each individual household. Prior studies have to rely on looking at the correlation between their estimates and characteristics such as liquidity, but our setting enables us to calculate more sophisticated measures that capture the influence of each variable on the MPC.
The fine-grained consumption data of the CEX allows us to identify what kind of goods households consumed and what they spent their stimulus money on. As Kaplan and Violante (2014) note, the tax stimulus is anticiapted and is subject to these special circumstances. Therefore, one might also speak of our estimated coefficients as a 'Propensity to consume the rebate' or 'rebate coeffcient', which is not necessarily equivalent to households' overall MPC. We compare our estimates with the range found in the literature

using different income shock sources to get a grasp of whether this difference might play a role and for what households it does. However, while a government stimulus program might not be perfectly appropriate to verify theoretical models concerned with the MPC, providing evidence on their effect on individuals is of major importance for future policy making. While in some cases when economic relief is urgent broadly defined, non-targeted stimuli might be a good option to pursue, targeted transfers can play a major role in many policy settings. **(rephrase)**

this needs to be implemented somewhere By exact definition, the MPC is the reaction to an unanticipated, transitory income shock. However, in out setting, the shock cannot be fully seen as unanticipated as the tax stimulus payment was signed into law and therefore known to the public several months before the payments were conducted. However, in such cases to identify only the contemporaneous reaction of households the empirical literature (e.g. Parker et al. (2013)) .

section on results here We show that indeed both these channels play a role in the heterogeneity of households' response to the 2008 tax stimulus. Similar to the existing literature we find... However, additionally we are able to show that the heterogeneity is not only linear/indeed linear...
The rest of the paper is structured as follows: Section 2 summarizes the theoretical and empirical literature on MPC heterogeneity. Section 4 discusses the data source and challenges connected with it. The empirical methodology we use is described in Section 3, while Section 5 presents the identification and estimation results. Section 6 concludes.

2 Literature Review

The literature investigating the size of the MPC and potential heterogeneity can broadly be categorized into three different strands. The first one uses quasi-experimental settings to exploit variation in income to estimate households' MPC. The second uses surveys that explicitly question participants about their MPC - be it out of actual or hyopthetical income shocks. Lastly, a vast literature focuses on building sophisticated macroeconomic models that are calibrated to match real world data and subsequently estimate the MPC agents experience in these models. Our work falls into the first of these categories.
In this section we briefly summarize the findings in all three and additionally discuss two studies - Parker et al. (2013) and Misra and Surico (2014) - in more detail as they investigate MPC heterogeneity using the same data as we do.
Quasi-experimental settings appear all the time in the real world, e.g. in case of a specific policy being implemented or another exogenous shock happening. Researchers interested

in MPC heterogeneity focus on shocks that alter the income of a household. For example, Fagereng et al. (forthcoming) use panel data from Norwegian administrative data on winners of a state lottery in which most citizens participate. Receiving a payment from the lottery can be seen as an unanticipated income shock because the chances of winning are so low. They find that households winning the lottery spend almost half of their win within one year and 90% after 5 years. Moreover, liquidity and age are the only variables correlated with the MPC, providing evidence for the existence of the liquidity and age channels. In similar vein, Golosov et al. (2021) construct a dataset of lottery winners in the USA to estimate their MPC and labor market response. They make use of tax forms provided by the lottery winners and general income tax statements. Their main goal is to estimate the labor market responses to windfall gains in unearned income but their strategy allows them to identify the MPC as well. Using a Difference-in-Difference estimator, their estimated MPC is around 60ct out of each dollar earned on average, while labor earnings are reduced by 50ct. To investigate heterogeneity in these responses, the authors split their sample based on the quartile along the liquidity distribution. Further supporting the liqudity channel, they find that households in the highest quartile spend only 49ct while the lowest quartile spends almost 80ct of each dollar won in the lottery. However, these two lottery-based approaches suffer from the drawback that they do not measure consumption directly. Instead they have to either construct consumption out of households balance sheet data (Fagereng et al. (forthcoming)) or model consumption as a function of their observed variables (Golosov et al. (2021)).

Gelman et al. (2018) use the government shutdown in the U.S. as a transitory liquidity shock. Hence, contrary to other literature they only estimate how liquidity changes the consumption behavior and not the MPC directly. Still, their setup allows them to disentangle the pure effect a liquidity shock has on consumers spending as government workers receive a payback of their wage once the government shutdown is over. Hence, there are no changes in expected income. Meanwhile, studies using income shocks cannot quantify what effect stems from the liquidity channel and what stems from changes in expected income. Their findings highlight that low liquid households react more to a negative liquidity shock as they have no assets to fall back on. Low liquid government workers started postponing their credit card payments, while simultaneously increasing the amount spend using them. **probably only add very short inside of this as not so much related to raw MPC and little/bad heterogeneity investigation**
The second strand of literature uses survey data from field surveys that question households about potential or actually relaized income shocks and how their reaction looks like. Bunn et al. (2018) use (**use twice here**) data collected by the Bank of England to assess the asymmetry that we expect in households' reaction depending on the sign of the income shock. As mentioned before, the liquidity channel suggests that an unantic-

ipated shock calls for a stronger reaction if its negative. Indeed, the authors are able to provide ample evidence for such a reaction with their estimated MPC out of a negative shock being being between 5 to 12 times as high as the reaction to a positive shock. The balance sheet data their source provides also enables them to show that borrowers show a more pronounced asymmetry and reaction, which is also in line with the theoretical mechanisms of the liquidity channel. This also holds for households that face some kind of liquidity constraint. Additionally, Bunn et al. are capable of replicating their estimated MPCs in a model with households at the borrowing constraint. These findings are further underlined by Christelis et al. (2019) who use Dutch data for a similar study. Summarizing these studies strongly suggest the existence of the mechanisms related to households at their borrowing constraint and precautionatry saving motives (**the latter must be elaborated on in intro**).

2.1 2008 Tax Stimulus

Lastly, we want to elaborate in more detail two already mentioned quasi-experimental studies: Parker et al. (2013) and Misra and Surico (2014). These two studies have been quite influential when it comes to studying MPC heterogeneity and use the same dataset as we do. Therefore, we want to lay out their approache sin detecting heterogeneity in greater detail here and highlight the major advantages our estimation approach offers. Compared to the general literature, we are the first to focus on the form of the heterogeneities reported in the literature so far (**check whether this isn't too bold of a claim**).

Thanks to Parker et al. (2013), the 2008 and 2009 BLS added questions about the tax stimulus to the CEX survey. Parker et al. (2013) use this data to first estimate a simple homogenous model controlling for age and the change in the family size using Ordinary Least Squares (OLS). Meanwhile, they set off to detect heterogeneity using interaction terms. More precisely, similar to Golosov et al. (2021) they create dummies that signal to which quartile a household belongs, where the quartiles are defined based on the distribution of various variables. The interaction terms then capture systematic differences between households across these quartiles. However, as we have already pointed out, this procedure can be quite problematic. It increases the potential to miss substantial heterogeneity across households. For example, consider the case where the largest heterogeneity is between the top 10% and the top 20% of households along some distribution. The approach using sample splitting or dummy variables is not capable of detecting these heterogeneities as those households belong to the same category. Additionally to this problem, using an OLS estimator, Parker et al. (2013) completely ignore any panel dynamics that might take place. Meanwhile, individual level fixed effects are identified as a

strong potential channel that drives heterogeneity in MPC. The results by Parker et al. (2013) are therfore potentially biased (**two times potential**).
To adress the heterogeneity issue, Misra and Surico (2014) follow a new approach: they employ a quantile regression estimator to estimate the heterogenous response of households to receiving the rebate. The quantile regression estimator is similar to the OLS but estimating conditional quantiles instead of the conditional mean. The authors therefore claim to estimate the conditional MPC out of the rebate for any quantile of consumption change. However, there are severe dissimilarities with what they are actually estimating and what they interpret. QR estimates how much a variable affects the outcome at a given quantile, e.g. the 10% quantile, of the conditional distribution of the outcome.

after mentioning all studies, here notes on why sample splitting or dummies are way worse than my approach While their procedure of sample cutting is common in the literature, it is an inconvenient procedure to detect heterogeneity as it only allows the authors to analyse differences between their pre-defined groups. Hence, in case heterogeneity is strongest between other groups, their findings underestimate or even completely miss patterns in the data.

2.2 Channel description

This will be pushed/worked into another section: decide whether at beginning of literature review or in introduction; for now here only to formulate something in a more consistent manner than what's in the intro right now
The theoretical literature has identified several channels which drive MPC heterogeneity. The two most prominent ones are life-cycle dynamics and liquidity. The former is driven by a consumer's age and the associated fluctuation in income. As data consistently shows (sources), consumption follows a hump shape over the life-cycle. In the case of liquiditiy, its role is linked to the nature of the income shock and completeness of the credit market. If a positive income shock is anticipated, households that are already close to or at their borrowing constraint cannot borrow new funds to smooth consumption in anticipation of a higher future income. Thus, once the shock realizes, we will observe an increase in consumption - although if we follow the PIH, this increase is rather small as the additional income is spread out over all future periods. In case of a negative anticipated shock, saving is always possible for any household and hence we will not see a reaction once the shock realizes. E.g. Bunn et al. (2018) document this asymmetry depending on the sign of the shock (**they document this for unanticipated shocks, shift back**). Thus, more liquid households react less to a positive anticipated shock in comparison with liquiditiy constrained ones. In contrast, in case of an unanticipated shock, we expect the opposite. Think of an agent that is temporarily out of work and has no liquid wealth at

their disposal. In case of a negative shock, the agent is forced to adjust their consumption behavior downward. Meanwhile, a positive shock will always be saved and stretched over future periods, no matter the level of households' liquidity. However, these theoretical predictions are made within a permanent income framework in which households try to smooth consumption over time.

Literature Notes

Parker et al. (2013)

- three sources of variation exist:
  1. timingand type of payment
  2. amount
  3. type of payment
- result: on avg. hosuseholds spent 12-30% of rebate on nondurable consumption goods → significant
- life-cycle/PIH model (LCPIH) is rejected by these findings
- LCPIH: no response to anticipated shocks at timing of arrival → borrowing/liquidity constraints may be main driver
- prior research: larger payments may skew consumption/usage of rebate towards durables (Souleles (1999) finds significant increase in ND and D goods in response to larger payments (federal tax refund in springtime))
- problem data: assets are not measured in detail and frequently
- find no significant response (interpret it anyway)
- keep in mind that tax rebate was disbursed during time of major economic downturn/turmoil
- look at relationship to owning house compared to renters → homeowners spend more than renters
- closest literature: Agarwal, Liu and Souleles (2007); Broda and Parker (2008); Bertrand and Morse (2009)
- Tax rebate
  - at least 300 (couples 600), at most 600 (1200) if 300+tax liability are above 600

  - at least 3000$ qualifying income
  - phased out with income starting at 75000$: 5% reduction of amount gross income exceeds 75k (couples 150k)
  - hh received a notice in advance of payment → anticipated shock!

- CEX
  - questions were added from june 2008 to march 2009
  - use 2007 and 2008 waves of CEX (2008 data includes first quarter of 2009)

- identification
  - use

3 Methodology

3.1 Notes on Methodology for writing

- see: `https://econml.azurewebsites.net/_autosummary/econml.dml.DML.html#econml.dml.DML`
- how is the CATE achieved in second stage?
  - second stage in Linear DML is OLS regression
  - CATE is achieved through interaction of some mapping $\phi(X)$ of the confounders X with the 'base' treatment effect Θ.
  - effectively running an OLS regression with interaction terms $\tilde{D} \otimes \phi(X)$
  - hence, assume that CATE is linear in X unless using polynomial mapping $\phi(X)$
  - nonparametric DML such as Causal Forest DML circumvents this assumption and detects non-linear heterogeneity in case there is any (run this as kind of robustness check/better specification)
- causal forests/generalized random forests:
  - random forests only allow for prediction, no inference possible
  - Athey and Wager (2016) and Athey, Tibshirani and Wager (2019) develop causal forests/grf
  - these allow to estimate any desired local moment equation - e.g. those of a treatment effect analysis
  - more specifically, in the DML case the moment equation estimated is

$$E\big[\,(Y - E[Y|X,W] - <\theta(X), T - E[T|X,W]> - \beta(x))\,(T;1)|X = x\big] = 0$$

  - I should push this explanation into the Appendx though and just explain the GRF in more general terms (it detects heterogeneity that is not specified before)

To identify the causal effect of the income shock on households' change in consumption, we use the Double Machine Learning (DML) estimator developed by Chernozhukov et al. (2017). This rather new kind of estimator allows to efficiently estimate semi- or non-parametric models of treatment effects. The DML estimator has the major advantage that it does not restrict the effect of confounders on the outcome to a specific functional form. Instead it uses Machine Learning methods to freely estimate this relationship.

Through the orthogonalization step discussed below it takes care of any confounding effects and cleanly identifies the pure effect of treatment on the outcome. Meanwhile, its implementation procedure deals with common biases arising in more naive estimation procedures that employ Machine Learning methods, opening the door to making use of these spohisticated methods in causal inference studies. Even in settings in which the raw predictions of the ML alogrithms used are not of high quality, the estimator yields desirable results and properties (**rephrase this last sentence again**).
From a more theoretical perspective the DML estimator also yields very efficient properties when it comes to its asymptotic analysis, especially a rate of convergence that is faster than other nonparametric estimators. Under certain assumption, Chernozhukov et al. (2017) are able to prove root-n consistency of the estimator. However, we will not further elaborate on these lattertechnical details but rather focus on how the estimator works in general. For a more technical discussion the reader is referred to Chernozhukov et al. (2017) and Semenova et al. (2021). Instead, this section introduces the general idea behind the DML estimator as well as the different variants we will use in Section 5.

3.2 Idea behind DML

We start with considering a Partially Linear Model of treatment and outcome. Note that the DML estimator is capable of estimating various models, but we will also use a PLM specification later on. Additionally, it helps to intuitively understand the main idea of the DML estimator and its mechanics. Section 3.5 will briefly present what a fully non-parametric approach looks like.
The Partially Linear Model (PLM) is given by

$$Y_{it} = \theta(X_{it})D_{it} + g(X_{it}, W_{it}) + \epsilon_{it} \tag{1}$$

$$D_{it} = h(X_{it}, W_{it}) + u_{it}, \tag{2}$$

where Y_{it} is the outcome, D_{it} is the treatment and X_{it} and W_{it} are observable variables. We distinct between simple confounders W_{it} which affect the outcome and also potentially the treatment and X_{it} which additionally are considered to impact the average treatment effect of D_{it} on Y_{it}. The choice of these variables is left to the researcher. We are interested in $\theta(X)$, the conditional average treatment effect (CATE), which in Rubin's **missing citation here** potential outcomes framework is defined as

$$\theta(X) = E[Y_1 - Y_0|X = x].$$

where Y_d is the outcome when treatment is $D = d$. In our setting, treatment is not binary but constant, hence $\theta(X)$ represents the marginal CATE

$$\theta(X) = E[\frac{\delta Y(d)}{\delta d}|X = x].$$

The marginal CATE measures how much a marginal increase in the continuous treatment changes the outcome, conditional on the individual having a set of characteristics $X = x$. The DML now follows a two step procedure to identify $\theta(X)$. We define

$$E[Y_{it}|X_{it}, W_{it}] \equiv f(X_{it}, W_{it}) \tag{3}$$

$$E[D_{it}|X_{it}, W_{it}] \equiv h(X_{it}, W_{it}) \tag{4}$$

where (4) follows from (2). We can use these two conditional expectations to show that the PLM can be boiled down to

$$Y_{it} - f(X_{it}, W_{it}) = \theta(X_{it})(D_{it} - h(X_{it}, W_{it})) + \epsilon_{it}.$$

This orthogonalization of treatment and outcome guarantees that the regression coefficient $\theta(X)$ only captures variation that the treatment invokes in the outcome and is free of any confounders acting through treatment D. Moreover, as Chernozhukov et al. (2017) show, simply plugging in estimates of f and h into the PLM results in a regularization bias when estimating $\theta(X)$ that does not vanish asymptotically, effectively prohibiting consistency of the estimator. It is circumvented by the orthogonalization of treatment and outcome, while there is no difference in orthogonalization and controlling for confounders. **(Frisch-Waugh-Lovell Theorem kind of, also potentially missing details on orthogonalization)**

The first stage of the estimation process consists of choosing an appropriate Machine Learning method and finding estimates of the conditional expectation functions f and h. A welcome property of the DML estimation is its agnostic to the first stage estimator. Thus, it allows choosing the appropriate prediction method for the given setting.

Once we obtain the first stage predictions $\hat{f}$ and $\hat{h}$, we use them to orthogonalize treatment and outcome to retrieve the residuals

$$\tilde{Y}_{it} = Y_{it} - \hat{f}(X_{it}, W_{it})$$
$$\tilde{D}_{it} = D_{it} - \hat{h}(X_{it}, W_{it}).$$

The second stage then only consists of a linear regression of $\tilde{Y}_{it}$ on $\tilde{D}_{it}$ that yields $\hat{\theta}(X)$. More precisely, the partially linear model we consider here implicitly assumes a parametric form of the CATE

$$\theta(X) = \phi(X) \times \Theta,$$

where Θ is the base treatment effect and $\phi(X)$ is a mapping of confounders X. In practice, the estimator boils down to a linear regression which includes interaction terms $\tilde{D} \otimes \phi(X)$. The mapping $\phi(X)$ can take any parametric form we might think of. In Section 3.5 we discuss how this assumption can be avoided, allowing to detect any heterogeneity without pre-defining its functional form.

3.3 Cross-Fitting

The DML estimator achieves its desirable consistency results by avoiding two common biases arising in settings that employ Machine Learning estimators: overfitting and regularization bias. We have already briefly discussed that the regularization bias is avoided through orthogonalization. However, the overfitting problem needs to be adressed on its own.

Overfitting is the result of an estimator adjusting too much to the given data such that when it is exposed to new data it performs very badly resulting in a high variance of this estimator. Again, Chernozhukov et al. (2017) show how this variance term does not vanish in the asymptotic analysis leading to an inconsistent estimator. However, this issue can be avoided using a technique they coin *cross-fitting*. Instead of using all observations to find the estimates of f and h and then estimate $\theta(X)$ using the whole sample, consider the case in which we split the sample into two. The first sample is used to estimate the first stage estimators. Those are used to predict the values in the second sample, which are then subsequently used for orthogonalization and the second stage estimation. In case we are interested in the unconditional average treatment effect (ATE), this procedure is repeated with the role of the samples reversed and the resulting estimators are averaged. However, in the CATE case we are interested in individual-level point estimates. Therefore, while the role of both samples are switched, we do not average any results but keep the individual level estimates of all observations. The cross-fitting procedure for splitting up the sample into any K folds is described in Algorithm 1 which summarizes the baseline DML estimator overall.[1]

[1]Note that Chernozhukov et al. (2017) argue that K=4 or K=5 performs reasonably well, even for smaller samples.

This has to look better and be more 'algorithmic'.

Algorithm 1 Double Machine Learning Estimator

1: Split up sample into K folds.
2: To estimate $\widehat{h}$ and $\widehat{f}$ for for the k^{th} fold use observations $j \notin k$.
3: To get residuals for observations in k, calculate $\widehat{h}(X_i)$ and $\widehat{f}(X_i, W_i)$ for $i \in k$ and use to retrieve residuals.
4: Once residuals of each fold retrieved, estimate $\theta(X_i)$.

3.4 Panel DML

So far we have considered the original DML estimator that relies on the assumption of strict exogeneity. While even in the panel setting at hand this assumption might be reasonable, Semenova et al. (2021) propose another DML estimation method that relaxes it. More precisely, for their estimator to be consistent we only have to assume conditional sequential exogeneity, which enables us to control for panel dynamics in a more precise manner. We lay out why the assumption of strict exogeneity might be reasonable in Section 2 and will compare these arguments to assuming conditional sequential exogeneity in Section 5.
More formally, in the Panel DML setting we assume

$$E[\epsilon_{it}|X_{it}, W_{it}, \Phi_t] = 0$$
$$E[u_{it}|X_{it}, W_{it}, \Phi_t] = 0,$$

where Φ_t is the information set in period t. Semenova et al. (2021) show in a setting with low-dimensional treatment, this assumption still results in the same second-stage estimator as the original DML. The only difference in the estimation procedure is the cross-fitting algorithm in the first stage. We form its folds based on the time index instead of simply randomly splitting up the sample. Moreover, we can include lagged values of treatment and outcome to account for the information set Φ_t. As discussed in Section 2 including lagged treatment actually improves the proper identification of the MPC.

3.5 Nonparametric DML

In the PLM setting, the functional relationship how the CATE is influenced by confounders X via the mapping $\phi(X)$. However, the DML estimator also enables us to use a nonparametric approach that can detect any interaction between treatment and confounders to uncover heteroegeneity. It has the same first stage, but estimates the second stage using the Causal Forest estimator proposed by Athey, Tibshirani and Wager (????).

The Causal Forest is a generalization of the Random Forest prediction method developed by Breimann (2001), which has found application in a wide array of predictive tasks. However, the original algorithm - as most Machine Learning methods focusing on prediction - does not allow for any causal inference. The Causal Forest solves this problem by generalizing the objective function to fit the potential outcomes framework and developing theory that allows retrieving standard errors of the estimated coefficients. Appendix A elaborates in more detail how the Causal Forest algorithm works and how it identifies the treatment effect. Using the CF as a second stage enables us to estimate the model

$$\begin{aligned} Y_{it} &= g(D_{it}, X_{it}, W_{it}) + \epsilon_{it} \\ D_{it} &= m(D_{it}, X_{it}, W_{it}) + u_{it}. \end{aligned}$$

As part of our analysis we will compare the results to check whether the relationship is indeed linear or whether we discover non-linear heterogeinities that the PLM approach does not account for and have not been considered in literature yet. However, note that when using a nonparametric second stage the convergence rate of the estimator declines. While still achieving faster rates than most other nonparametric estimators, this implies that the Causal Forest based approach is more demanding when it comes to the number of observations.

4 Data

We use data collected by the Consumer Expenditure Survey (CEX) that is administered by the Bureau of Labor Statistics. Additionally to the main questionnaire, the CEX added questions about the stimulus payments to their surveys conducted between June 2008 and March 2009 (**is this correct time in dataset?**). The main advantage of the CEX is that it creates a unique representative sample that contains finely grained information on the type of goods households consume. While its main purpose is to serve as the benchmark to determine the goods basket used to measure inflation in the USA, it enables a detailed analysis on what households spent their rebate on. In the following, we briefly outline the stimulus program and describe the CEX data.

4.1 The 2008 Tax Stimulus Program

Due to the global financial crisis and the subsequent recession, the United States government passed the Economic Stimulus Act (ESA) in February 2008. With projected costs of more than 150 billion USD it was the largest relief program passed in the USA's history. Next to the stimulus payments, which made up roughly two thirds of the program, the ESA also enacted other steps meant to provide economic relief such as enabling government owned entities (Fannie Mae and Freddie Mac) to buy up more mortgages. However, we focus on the effects of the stimulus payments.

Each household that filed for taxes in 2007 was eligible to receive a minimum amount of 300 USD (600 USD for couples filing jointly) and a maximum of 600 (1200) USD. The exact amount of each individual was determined by their net tax liability in 2007 up to the maximum amount. Additionally, households received 300 USD per dependent child under the age of 17. Meanwhile, the payment was also connected to households gross income, with the rebate being reduced by 5% of the amount the income exceeded 75,000 (150,000) USD. (**the last two sentences are very close to wikipedia source**)

4.2 Consumer Expenditure Survey

The CEX is a representative survey of households in the USA. Once a household is selected to participate, they are interviewed a total of five times. The first interview is a baseline interview in which some general household characteristics such as the financial circumstances are determined and their stock of nondurable goods are documented. The next four interviews are administered every 3 months in which households are asked to document their expenditures over these past three months. After that, the household is rotated out of the CEX and replaced with a new one. Hence, each month of data documented in the CEX contains a different set of households as new ones are added and others do not show up anymore as they are rotated out of the survey.

5 Estimation and Results

$$\Delta C_{it+1} = \theta(X_{it})R_{it+1} + g(X_{it}, W_{it}) + \epsilon_{it} \tag{5}$$

$$R_{it} = h(X_{it}, W_{it}) + u_{it} \tag{6}$$

where ΔC_{it+1} is change in consumption, R_{it+1} is the amount of rebate received by the household and $g(X_{it}, W_{it})$ and $h(g(X_{it}, W_{it})$ are non-parametric functions of confounders. X_{it} and W_{it} are distinct by the assumption that only X_{it} influences the marginal effect of the rebate, $\theta(X)$, while W_{it} denotes the set of confounders that play no role in the effect.

6 Conclusion

(from intro 2 draft but doesn't fit that much anymore) Our contribution is twofold: for one, we estimate the conditional MPC out of the tax stimulus in the most precise and rigorous manner thus far. Second, we use an estimator that exploits the power of machine learning methods for causal inference and contribute to the wider understanding and promoition of this method among applied researchers. Machine Learning predictors are powerful tools when it comes to handling large data and/or complex relationships between variables without any specification of those.

References

Bunn, P., J. Le Roux, K. Reinold, and P. Surico (2018): "The consumption response to positive and negative income shocks," *Journal of Monetary Economics*, 96, 1–15.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2017): "Double/Debiased Machine Learning for Treatment and Causal Parameters," *arXiv:1608.00060 [econ, stat]*, arXiv: 1608.00060.

Fagereng, A., M. B. Holm, and G. J. Natvik (forthcoming): "MPC Heterogeneity and Household Balance Sheets," *American Economic Journal: Macroeconomics.*

Golosov, M., M. Graber, M. Mogstad, and D. Novgorodsky (2021): "How Americans Respond to Idiosyncratic and Exogenous Changes in Household Wealth and Unearned Income," Working Paper 29000, National Bureau of Economic Research, series: Working Paper Series.

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

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, 2530–2553.

Semenova, V., M. Goldman, M. Ai, V. Chernozhukov, M. Taddy, and M. Ai (2021): "Orthogonal Machine Learning for Demand Estimation: High Dimensional Causal Inference in Dynamic Panels," 66.