Detecting Heterogeneity in the MPC: A Machine Learning Approach

Last Update: 03.09.2021

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

The Marginal Propensity to Consume (MPC) is at the centre of the macroeconomic model introduced by John Maynard Keynes in his General Economic Theory. Eversince its introduction, the role and size of the MPC has been subject to debate. While Keynes declared the MPC to be meaningfully different from zero, the permanent income hypothesis developed by Milton Friedman and a corner-stone of modern macroeconomics declares it to be irrelevant to current consumption decisions and, thus, irrelevant to economic policy making. However, both are wrong and right at the same time. More recently, the focus of research concerned with undertanding the MPC to guide policies - such as stimulus payments - has shifted to painting a more diverse picture of households' willingness to spend out of a transitory income shock. New, sophisticated models formalize the heterogeneity of agents in the macroeconomy, including their MPC. Additionally, empirical work has shifted from trying to prove an MPC of zero - or the opposite - to understanding the difference across households and allowing for heterogeneity in the MPC. **add two examples of channels - liquidity constraint and ...**

Using the 2008 tax stimulus as an exogenous income shock, my contribution to the empirical literature is twofold: First, I use a new and highly flexible estimation approach, that allows me to identify a wider range of heterogenous effects. The so-called Double Machine Learning Approach allows for a semi-parametric setup in which the functional form of any confounding factor does not have to be specified. Second, the literature using the data from the 2008 stimulus (or the general literature? **check!**) so far has investigated its effect (poor) methods that lack the advantage of the DML approach while at the same time not allowing to account for the panel data setting of the data. These approaches implicitly impose a strict exogeneity condition, while the Panel DML model is capable of accounting for possible effects of past characteristics on the change in current income. Therefore, I am able to identify the causal effect of the tax rebate on consumption change more clearly (or: actually identify it, but maybe to harsh).

I rely on data collected by the Consumer Expenditure Survey (CEX), which included a special part in the 2008 and 2009 surveys dedicated to the tax stimulus. This effort was promoted by Johnson, Parker and Souleles (2013; henceforth JPS) who quantify the effect of the tax stimulus on consumption changes. The data they use is publicly available and also used by Misra and Surico (2014; henceforth MS). Hence, to improve comparability with two of the more recent and prominent contributions, I use the data provided publicly by JPS as well. While both document some heterogeneity in the MPC, there are several drawbacks in their respective analysis. Meanwhile, the DML estimation allows me to identify household level point estimates and standard errors, allowing me quantify whether the estimated MPC is significantly different from zero for each individual to uncover which households actually experience a temporary increase in consumption due to a temporary income shock (**rephrase**).

rewrite and put this somewhere else Understanding which underlying factors drive heterogeneity in the MPC is crucial for policy makers. While short-term untargeted tax-stimuli such as the one in 2008 are reasonable in times of economic crisis when time is short, targeted stimuli can improve the payoff of each dollar invested into an economic stimulus.

add a brief summary/overview of what I find

The rest of the paper is structured as follows: Section 2 summarizes the theoretical and empirical literature on MPC heterogeneity putting a focus on the issues concering JPS and MS analysis. Section 3 discusses the data source and challenges connected with it. The empirical methodoogy I use is described in Section 4, while Section 5 presents the results. Section 6 concludes.

2 Literature

The literature investigating the size of the MPC and potential heterogeneity can be summarized in three different strands.

The first one uses quasi-experimental settings to identify income shocks and the resulting reaction of consumption. Settings considered are US tax stimulus progrmas during the times of economic crisis in 2001 and 2008 or lottery wins by individuals. Johnson, Parker and Souleles (2006) estimate the size of the MPC out of the 2001 tax stimulus and in a more recent contribution also take a look at the 2008 tax rebate program (Johnson, Parker and Souleles, 2013). The latter is closely related to our procedure and is hence discussed in more detail further below. They find XXX. Meanwhile, Fagereng et al. (2020) estimate the heterogeneous MPC out of lottery winners in Norway. Golosov et al. (2021) do the same using bla data. The second strand of literature uses self-reported MPC from household surveys. However, studies based on self-reported data are prone for measurement error - specifically the so-called self-report bias which leads respondents to misreport their data. In the case of the Marginal Propensity to Consume we expect this to be even larger than in the survey data exploited in quasi-experimental settings since respondents do not only have to document their raw spending behaviour (e.g. indicating how much money was spent in total) but assess their MPC on their own. Such calculations are likely to increase the risk of measurement error, especially the more abstract the concept becomes. There is also a more theoretical side to the discussion focusing on calibrating heterogenous agent new keynesian models (HANK) to uncover general equilibrium effects of single agents' MPCs on the aggregate MPC out of income shocks. Finally, there are two contributions that are by default most closely related to our setting since we make use of

the same data. Namely, these are Johnson, Parker and Souleles (2013) and Misra and Surico (2014). They estimate a simple fixed-effects regression in which they interact their income schock variable with pre-defined dummies. Those dummies are based on continuous variables and created by choosing discrete cut-off points. However, this prohibits the detection of heterogeneous patterns that are not captured by the variables considered or are not inside the defined thresholds. Using the Parker et al. data, Misra and Surico (2014) replicate their approach but use quantile regression to analyse the heterogeneity in the MPC distribution. While quantile regression can be of service to detect heterogeneity in coefficients, it does not allow for the correct interpretation. The treatment effects they uncover are the effect of the income shock on the difference in consumption before and after for a respective quantile. However, this quantile does not need to include the same individuals. Hence, the quantile regression only uncovers shifts in the overall distribution but is silent on how specific individuals changed their consumption pattern - and hence the actual MPC.
Lastly, as Kaplan and Violante **(or who exaclty was it?)** point out, empirical analysis that use stimulus payments as a temporary income shock to identify the MPC might actually estimate another coefficient, which they coin the coefficient of rebate. They argue that the conditions of a stimulus payment as well as the overall economic conditions that lead to such a payment are too specific (**rephrase**) to

In the JPS and MS specifications, they also consider simple linear estimators (OLS and Quantile Regression) that imply the assumption of strict exogeneity. Since we are looking at quarterly data and JPS/MS only consider age and change in the size of family as confounders, one could argue that there is little to no variation in these variables between quarters. In that case,

3 Methodology

To identify the causal effect of the income shock on households' change in consumption, we employ the Double Machine Learning (DML) approach developed by Chernozhukov et al. (2017). More precisely, to account for the dynamic structure of our data we use the variant from Semenova et al. (????) who present a DML estimator that allows for less constraining conditions.
In a setting like ours where one is interested to estimate heterogenous treatment effects, the DML estimator has a major advantage over classical econometric approaches which have been adopted in the literature so far. Namely, it does not restrict the effect of confounders on the outcome to a specific functional form but uses Machine Learning methods to freely estimate this relationship. Simultaneously, the orthogonalization step discussed below takes care of identifying the true effect stemming only from the treatment. Lastly, from a theoretical perspective this estimator yields very efficient properties when it comes to its asymptotic analysis, especially a rate of convergence that is faster than other nonparametric estimators, in part even achieving root-n consistency. 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. (????). Instead, this section introduces the general idea behind the DML estimator and how the Panel DML setup differs. These differences

3.1 Idea behind DML

We start with considering a Partially Linear Model of treatment and outcome. However, be aware that the DML estimator is not constrained to this form but rather it helps us to understand the idea and mechanics behind it in a clear way. Section X.x 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.

The DML now follows a two step procedure to clearly identify $\theta(X)$. First, we define

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

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

where (4) follows from (2). The rewrite (1)

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

The first stage now consists of choosing an appropriate Machine Learning method to find estimates of the conditional expectation functions $f(\ddot{;})$ and $h(X_{it}, W_{it})$ (**which is nicer way of presentation?**). 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 $\hat{f}(X_{it}, W_{it})$ and $\hat{f}(X_{it}, W_{it})$, 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 orthogonalization guarantees that the residuals only contian variation that is not induced by any of the confounders in treatment and outcome, respectively. Therefore, the second stage then only consists of a simple linear regression of $\tilde{Y}_{it}$ on $\tilde{D}_{it}$ that yields $\hat{\theta}(X)$. **How do we get the individual level point estimates?**

3.1.1 First Stage Cross-Fitting

Should I additionally desribe the cross-fitting approach or just mention it as a side thing? -¿YES

3.2 Panel DML

So far we have considered the original DML estimator that considers a cross-sectional setting, however, we will look at a dynamic panel of households. However, under the assumption of strict exogeneity, it is also reasonable to use this estimator in our setting. The reasoning in favor and against this assumption have been laid out in Section X. Since we aim to compare how well this assumption holds up contrary to when relaxing it, we have to introduce the Panel DML approach by Semenova et al. (????), which only assumes conditional sequential exogeneity. More sepcifically, 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. Under some more technical assumptions, Semenova et al. show that in a setting with low-dimensional treatment, the estimator is similar to the original DML. The only difference in the estimation procedure is the cross-fitting algorithm in the first stage. The folds for the cross-fitting procedure are formed based on the time index instead of simply randomly splitting up the sample.
In practice, we will include lagged values of treatment and outcome as confounders in W_{it} to model the information set Φ_t. This procedure is also proposed by Semenova et al.

3.3 Nonparametric DML

The above described estimators both aim to estimate a partially linear model of treatment and outcome. While this allows for flexible estimation of the confounders, it still assumes a linear relationship between the treatment and outcome, resulting in the estimate $\hat{\theta}(X)$ being linear in X. While this is similar to the current literature, we also propose a specification in which we make no assumption on the functional form whatsoever. This nonparametric approach has the same first stage as the PLM based one, but estimates the second stage using the Causal Forest estimator proposed by Athey 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 a treatment effect study framework and developing theory that allows retrieving standard errors of the estimated coefficients. (**obviously re-write but no more detail in general I guess**)
Using the CF as a second stage enables us to write the model as

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

There are no specific assumption in what what form the treatment effect will depend on the confounders X_{it}. 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 have not been considered yet.

Estimation and Results

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

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

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.

References

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.

SEMENOVA, V., M. GOLDMAN, M. AI, V. CHERNOZHUKOV, M. TADDY, AND M. AI (????): "Orthogonal Machine Learning for Demand Estimation: High Dimensional Causal Inference in Dynamic Panels," 66.