University of Duisburg-Essen
Faculty of Business Administration and Economics
Chair of Health Economics

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Males, Catch Up!

Replication and summarization of **brunello** findings

Term Paper

Submitted to the Faculty of
Economics
at the
University of Duisburg-Essen

from:

Nils Paffen

Reviewer:	Norman Bannenberg (M.Sc)
Deadline:	30.09.2020

Name:	Nils Paffen
Matriculation Number:	3071594
E-Mail:	nils.paffen@stud.uni-due.de
Study Path:	M.Sc. Economics
Semester:	3rd
Graduation (est.):	Winter Term 2021

Contents

List of Figures

List of Tables

List of Abbreviations

1 Introduction

This term paper will focus on the summarization and the replicated results of **brunello**. The main idea of the paper is to show if extending years of compulsory schooling has an effect on the distribution of wages. Further findings might be that compulsory school reforms significantly affect educational attainment, which holds in this replication study for almost none qunatile, since the given data does not help to show the possible effect of the instrument years of compulsory schooling . This term paper, is limited to the data of the SHARE data set, compared to a mixed dataset consisting of a "data[set] drawn from the 8th wave of the European Community Household Panel (ECHP) for the year 2001, the first wave of the Survey on Household Health, Ageing and Retirement in Europe, or SHARE, for the year 2004, and the waves 1993 to 2002 of the International Social Survey Program (ISSP)"**brunello**. Due to the latter difference, replicated results differ heavily from those in the original paper. More about this in the dicussion section. This paper will start with a summary of the empirical model and the strategy the authors used, followed by a short overview of the SHARE dataset. Afterwards the replicated results limited to the latter dataset were presented and a short comparison to the findings of **brunello** is included. A discussion of the replicated results is followed by a conclusion section which closes this term paper.

2 Empirical Model

The author of the original paper starts with an introduction to schooling and the direct correlation on wages. They claim that the individuals or their parents choose years of schooling to maximise

$$u\left(w_i, s_i\right) = \ln\left(w_i\right) - c\left(s_i\right) \tag{1}$$

where w_i is interpreted as a function of s_i ($w_i = g(s_i)$), s_i itself is defined as year of schooling, $c(s_i)$ is the cost of schooling, while the index i indicates the individual. The optimal s_i is then given by the s_i which satisfies the equation of the marignal costs and the (expected) marginal benefits of schooling. The marginal costs rise in schooling, decrease in cognitive ability a_i and are a function of extenal controls X_i and z

$$mc\left(s_i\right) = r\left(X_i, z\right) + \theta s_i - \kappa.a_i \tag{2}$$

Further the author follow a Miniceria earnings function

$$\ln(w_i) = \beta s_i + s_i(\lambda a_i + \phi u_i) + \gamma_w X_i + a_i + u_i \tag{3}$$

where constant term is incldued in X, in the model of the replication study a constant is added seperatly to the model, the variable $a \sim G_a(0, \sigma_a^2)$ (cognitive ability) is knwon to the individuals at the time of their choice. Following the authors interpretation $u \sim G_u(0, \sigma_u^2)$ can be described as as fortune in the labour market and formally as an error term orthogonal to ability. Besides, u_i hold an zero mean demand shock which is corelated with relative productivity of jobs and skills. Equation (3) shows that schooling influences the location, due to $\beta * s_i$, on one hand and on the other the scale of the earnings distribution, through the interaction with ability and labour market fortune. Ability influences the individual earnings through a_i in a direkt way and via its product with schooling. As mentioned in the beginning, optimal schooling s_i^* needs to satisfy

$$mc(s_i) = mb(si) \tag{4}$$

where $mb(s_i)$ is defined as

$$mb(s_i) = \beta + \lambda a_i \tag{5}$$

therefore s_i^* can be written as

$$s_i^* = \frac{\beta - r(X_i, z_i)}{\theta} + \frac{\lambda + \kappa}{\theta} * a_i \tag{6}$$

When $\lambda > 0$ ability and schooling are complements and both can be defined as substitues if $\lambda < 0$ holds. Further the authors assume that $1 + \phi s_i > 0$, which secures the endogenoues variable log earnings are a (increasing or decreasing) monotonic function of the labour market fortune variable u_i. The authors propose an exactly identified triangular model as in Chesher's approach (7) and (8) but mention that the orthogonality condition for the consistency of the OLS estimation of (3) fails if one cant appropriately control for ability. In this case the triangular model can be explained as the following. Assume that the earnings of an individual are correlated with their educational level. Let this educational level in this model be defined as years

of schooling (s). Therefore we need an instrument which correlates with the educational level measurement but not with the earnings.

To solve the latter and thereby generate an consistent estimatator, **brunello** propose a variable z that is corellated with schooling but orthogonal to individual ability conditional on schooling and orthognal to the endogenous variable of (8). The instrumental variable in this model is years of compulsory schooling *ycomp*. Taken this into account the (exactly indentified triangular) model can be expressed as :

$$\ln(w) = \beta s + s(\lambda a + \phi u) + \gamma_w X + a + u \tag{7}$$

$$s = \gamma_s X + \pi z + \xi a \tag{8}$$

with $\xi = (\lambda + \kappa)/\theta$ Let $\tau_a = G_a(a_{\tau_a})$ and $\tau_u = G_u(u_{\tau_u})$, where a_{τ_a} and u_{τ_u} are the $\tau-$ quantiles of the distributions of a and u, respectively. Additionally define $Q_w(\tau_u \mid s, X, z)$ and $Q_s(\tau_a \mid X, z)$ as the conditional quantile functions corresponding to log wages and years of education. To achieve the recursive conditioning model one needs to compute the control variates first. Step one is to estimate the conditional quantile functions of schooling s and afterwards subtract the estimated values of the specific qunatile from years of schooling. Considering (8) again and the fact that the model is exactly identified one only remains with the value of ability at the specific quantile tau. Formally :

$$a(\tau_a) = s - \bar{Q}_s(\tau_a \mid X, z). \tag{9}$$

Afterwards one adjust the conditional quantile functions of $ln(w)$ with the control variate of (9) so that the residuals, orthogonal to ability, of the estimated conditional qunatile regression of $ln(w)$ yields to $u(\tau_u)$ of the following regression equation :

$$\tilde{Q}_w[\tau_u|X, s, a(\tau_a)] = \beta s + s(\lambda a(\tau_a) + \phi u(\tau_u)) + \gamma_w X + G_a^{-1}(\tau_a) + G_u^{-1}(\tau_u) \tag{10}$$

Now one can construct the parameter $\Pi(\tau_a, \tau_u)$ which is a matrix with the following structure :

$$\Pi(\tau_a, \tau_u) = \beta + \lambda G_a^{-1}(\tau_a) + \phi G_u^{-1}(\tau_u) \tag{11}$$

Due to recursive conditioning $Q_s(\tau_a \mid X, z)$ on $Q_w[\tau_u \mid X, z)$ one yields to the following model :

$$Q_w[\tau_u \mid Q_s(\tau_a \mid X, z), X, z] = Q_s(\tau_a \mid X, z)\,\Pi(\tau_a, \tau_u) + \gamma_w X + G_a^{-1}(\tau_a) + G_u^{-1}(\tau_u) \tag{12}$$

$$Q_s(\tau_a \mid X, z) = \gamma_s X + \pi z + \xi G_a^{-1}(\tau_a) \tag{13}$$

A a two stage fit of the the latter models then gives us the coeficient of $Q_s(\tau_a \mid X, z)$ After we plug this into $\Pi(\tau_a, \tau_u)$ for the coefficient β. We will repeat this step for each quantile tau (0.1, 0.3, 0.5, 0.9) but always altering only the quantiles of either τu of $Q_w[\tau_u \mid Q_s(\tau_a \mid X, z), X, z]$ or τa of the parameter $Q_s(\tau_a \mid X, z)\,\Pi(\tau_a, \tau_u)$. Therebey the study observes in the first case the effect of how a specific quantile τ_a of $Q_s(\tau_a \mid X, z)\,\Pi(\tau_a, \tau_u)$ interacts with the entire distribution of the log hourly earnings , while the latter measure the effect of the different quantiles of the ability distribution on the fixed τ_u of the endogenous variable $ln(w)$. Integrating the key parameter $\Pi(\tau_a, \tau_u)$ with respect to τ_a results in mean quantile treatment effects. The latter gives an overview of how an individual with average abilites is rewarded for educatinal attainment in the different qunatiles of the labour market luck distribution.

3 Empirical Strategy

The crucial change in the papers strategy is, compared to other papers using the same instrument, is that the results are not limited to the conditional but also support the unconditional effect (marginalized effect) of the quantile regression which can be interpreted as OLS results. The latter will be shown by the mean quantile treatment effect, as described in the last section. Following **brunello** serveal assumption for correct identification have to be made, namely :

- Due to monotonicity, with respect to u in (7) and a in (8), individuals in a higher qunatile of the labour market fortune recieve higher wages, will individuals with higher ability tend to stay longer in educational training;
- Once the decision of schooling is made an individual can't foresee their future draw from the the distribution of labour market fortune distribution, but can form expactations about their draw;
- the instrument ycomp (years of compulsory schooling) has an remote impact on the distribution of education or the attainment to the latter. The treatment

is assigned quasi-randomly due to their date of birth, whithout any parental influence;

- the variation in the timing of the implementation might vary between muncipalities in a country but this has no effect on the general education level
- there is no other channel besides the individual's education level, how the educational reform influences the log wages, so they are excluded form the wage equation of the observables (7)

Pooling the data from all countries helps to support the instrument ycomp, since more observations help to measure the effect more robustly. Doing so one can exploit the fact that due to the different timings of compulsory school reforms we exclude the possibilty of a specific cohort helps the instrument to become more valuable. This summarization does not include the table 1 of the paper which is only informative for the fact that compulsory schooling reforms were introduced in different years at each country. To create the post and pre-treatment sample-size the reform dates of Table 1 from **brunello** were used, with the excpetion of Germany. Since muncipalities were not observed in the SHARE dataset the mean of the reform date at each muncipalities was used to identify post and pre-treatment individuals. The latter choosing was done with the distance between birth cohort b and cohort $\bar{b}_k$, while the latter is identifyed as the first cohort potentially affected by the change in mandatory school leaving age in country k.

A first feeling for the effect of *ycomp* respectivly to the SHARE data can be seen in Figure 1. The graph shows the longitudanal data of individuals five years before until five years after the compulsory school reform happend in the respective countrys.

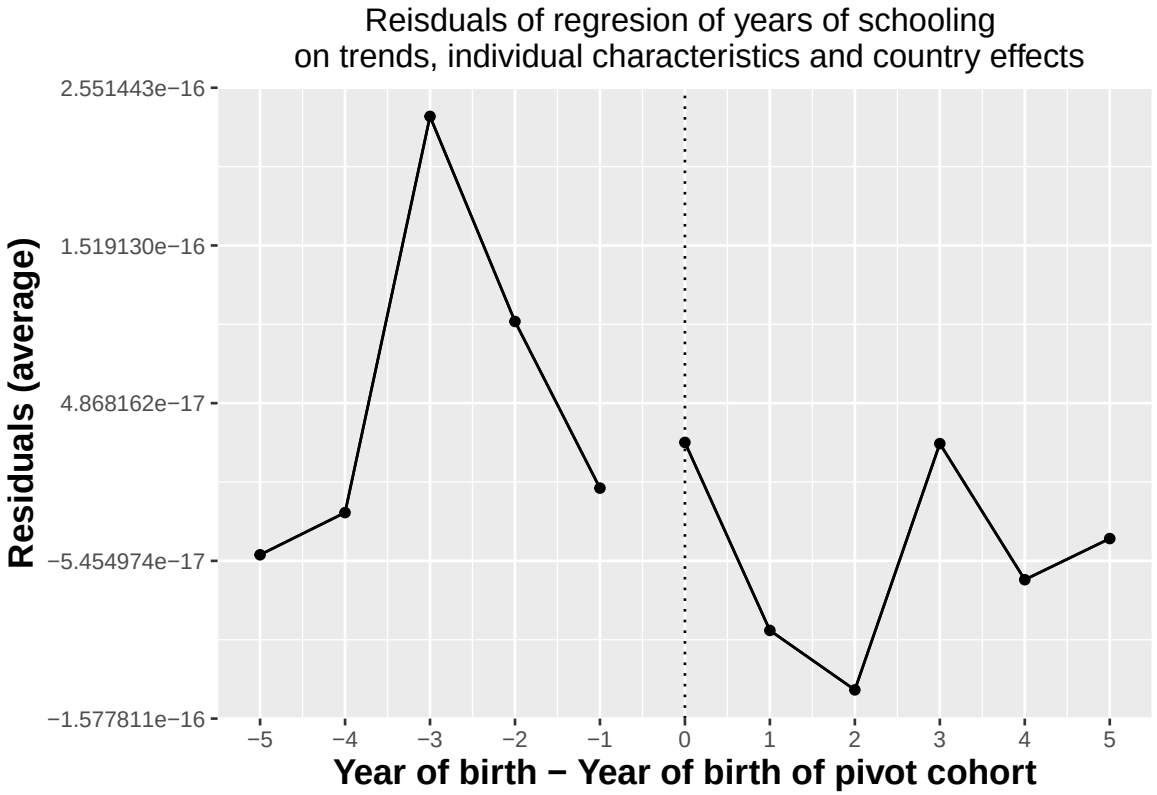

Figure 1: *The Effect of School Reforms on Educational Attainment*

Note. The OLS gender-specififc regressions included a constant, country dummies, q, q^2 and their interactions with country dummies and the GDP per head at the age when the pupil would have finished compulsory schooling.

The residuals show that even this simple OLS regression produces results with residuals which are more or less zero. This indicates that the OLS model achives to nearly perfectly explain the years of schooling. There is even an downward jump when the reform hits the pupils. We will see in latter results that in the first stage regression the instrumnent ycomp is always insignificant and mostly negative.

4 Data

As mentioned in the beginning this recursive study is done using only the data of the SHARE dataset. More than a decade has passed since the paper of **brunello** was published. During this time the SHARE data set of the first wave has been revised several times. In this term paper the dataset of version 7.0[1] is used. There is an more actual version, namely 7.1, but the stata data of this version are saved as .dat files and the .sav files of the 7.0 version were easier to implement. As mentioned in the

[1] https://releases.sharedataportal.eu/releases

empirical strategy **brunello** uses several controls to claim that their results are exactly identified. Attempts to gain access to these controls worked only partly. Therefore the GDP and their first lags, as reported in the Empirical strategy section, were retrieved from the OECD[2]. All other controls were either not avaiable or did not match the requiered time-length. The dataset is restricted to individuals aged between 26 to 65, following the argument of **brunello** that educational attainment does ot change after age 25. The final sample contains 7165 observations from 0 countries instead of 12 of the original paper.

```
\newcommand{\range}[1]{\eqparbox{ra}{#1}}
\newcolumntype{R}{>{\collectcell\range}c<{\endcollectcell}}



\begin{table}[!htbp]
\setlength\tabcolsep{3pt}
\sisetup{table-format=1.3, table-number-alignment=center}
\centering
\caption{Means of the key variable\strut} \label{}
\small
\begin{tabular}{@{} l SS[table-format=2.3]SRS[table-format=2.3]SS[table-format=4]S
\toprule
Country & {log w} & {s} & {ycomp} & \multicolumn{1}{c}{\makecell{Change in yrs.\\ o
 & {Age} & {\makecell{\%\\ Males}} & {Nobs} & {\makecell{\%\\ Complier}} \\
\midrule
Austria & 2.022 & 12.632 & 8.679 & 8 to 9 & 55.368 & 0.526 & 209 & 0.679 \\
Belgium & 2.288 & 12.162 & 8.005 & 8 to 12 & 53.234 & 0.543 & 877 & 0.001 \\
Denmark & 2.973 & 13.523 & 7.115 & 7 to 9 & 54.592 & 0.489 & 681 & 0.057 \\
France & 2.401 & 11.303 & 8.64 & 8 to 10 & 53.328 & 0.474 & 878 & 0.32 \\
Germany & 2.58 & 14.705 & 8.7 & 8 to 9 & 55.309 & 0.51 & 857 & 0.7 \\
Greece & 2.038 & 12.22 & 6.021 & 6 to 9 & 54.182 & 0.617 & 708 & 0.007 \\
Italy & 2.196 & 10.066 & 7.112 & 5 to 9 & 55.606 & 0.601 & 411 & 0.528 \\
Netherlands & 2.748 & 13.289 & 9.02 & 9 to 10 & 54.69 & 0.535 & 910 & 0.02 \\
Spain & 2.597 & 12.766 & 6.04 & 6 to 8 & 56.396 & 0.46 & 1251 & 0.02 \\
Sweden & 2.039 & 9.487 & 8.454 & 8 to 9 & 55.355 & 0.567 & 383 & 0.454 \\
\bottomrule
\end{tabular}
\end{table}
```

[2] https://stats.oecd.org/viewhtml.aspx?datasetcode=PRICES_CPI&lang=en#

Table 1 shows the log hourly real earnings, years of schooling, years of compulsory schooling, average age and percentage of males. Education attainment is highest in Germany (14.705) and lowest in Sweden (9.487). Average age is highest in Spain (56.396) and lowest in Belgium (53.234). Compared to the results of the paper by **brunello** the table is extended by the column *(*% Complier) which indicates how many individuals of the sample took the treatment and the column "Change in years of comp. school" of Table 1 from **brunello**. The latter columns indicate that the SHARE dataset contains mostly people which where not affected by the reforms. This might be an first explantion why the residuals of Figure 1 do not show the results **brunello** found in their study. The reason might be lack of observed treated individuals. For Belgium in the SHARE dataset this is roughly around 0.1 % of all individuals. Keeping in mind, that Belgium was one of the last countrys to extend the years of compulsory schooing, this might be also an answear to the the question, why the average age of Belgian individuals in this replication study differ that much from the average age of the original paper. To capture trend-like changes in the log earnings the study chooses, as described in the original paper, a second order polynomial in $q = t+7$, where t describes the distance between the individual and the first cohort affected by the reform, and the effect of the interactions with country specific dummies. This study follows the market-entry approach, which matches each individual with the first lags of their country specific GDP at that time when they would have applied to the job market for the first time without the reform. So a Spanish citizen born in 1960, where the cirtical age before the reform was 12 would be matched with the GDP values around 1972 to control for the posibility that the changes in educational attainment after the reform can be credited to the reform itself and not to some other economic time and/or country-specific factors.

5 Empirical Evidence

Before this paper starts with the discussion of the results, one problem that arised during almost all regression results that will follow in this chapter is singular matrices. A square matrix can be described as singular, that is, its *determinant* is zero, in other words, one or more of its rows(columns) can be exactly expressed as a linear combination of all or some or some other its rows (columns). In the multivariate data case, like the one in this paper, this can habben if there is linear interdependances among the variables. Since **brunello** adds many dummy variables it is possible to run into this problem when your dataset is smaller, as in the reproduction studies case. There are several ways to fix this problem such as covariate reduction techniques such as LASSO, which only keeps variables which are signifcant for the regression. The latter is explicitly usefull if you have more variables than observations. But this does not reflect this

papers case. Another way to at least solve the problem of exact linear combination is jittering. This means nothing else than a small *noise* is added to all values. In this paper the noise level is a random value choosen from the interval $[-0.1; 0.1]$ for each value of the dataset. If this value is *small* the interferences of the results is negligible.

As in the paper by **brunello** the first relationship presented in this paper will be the quantile effect of education, as expressed by years of schooling, on the log earnings, under the condtion that education is treated as exogenous.

```
\begin{table}[!htbp]
\captionsetup{labelsep=newline, justification=centering}
  \begin{threeparttable}
       \caption{\textit{Quantile Effects When Education is Treated as Exogenous} \\
    \scriptsize (Sample size : 7,165) By gender (3,735 males and 3,430 females)}
     \begin{tabular}{*{6}{l}}
        \toprule
         & \( \tau=0.10\) & \( \tau= 0.30\) & \( \tau= 0.50\) & \( \tau= 0.70\) & \
        \midrule
        \addlinespace
        \textit{Males}   & $\underset{(0.0025)}{0.0093^{***}}$ & $\underset{(0.0052
\\
        \textit{Females}& $\underset{(0.0025)}{0.0065^{***}}$ & $\underset{(0.0052)
        \bottomrule
     \end{tabular}
    \begin{tablenotes}[flushleft]
      \small
      \item \textit{Note.} Each regression included a constant, country dummies, \t
    \end{tablenotes}
  \end{threeparttable}
\end{table}
```

Table 2 highlights returns to a one year increase in education from the 10th to the 90th quantile. All results are statistically significant at the highest level. In difference to **brunello** the returns for males mostly higher than those for females. The latter beats the other gender only in the 30th and 50th quantile. Following this results the 90-10 log wage differential would indicate that one additional year of education lead to an increas of 2.02 percantage points for males and an increase of 1.67 for females. This results draw an picture of a world where the males need to catch up to their gender

counterpart. Besided that, the results look robust and as expected, since in this model the instrument do not play any role in the causal inference of log earnings.

As argued by **brunello** we can't treat education as exogenous since we expect a correlation between the log earnings and the latter. Therefore we will use the described instrument ycomp to explain schooling and use these values as a substitute for education in our log earnings regression model.

```
\begin{table}[!htbp]
\captionsetup{justification=centering}
  \begin{threeparttable}
       \caption{\textit{First Stage Effect of ycomp on s} (Sample size : 7,165)}
        \begin{tabular}{*{6}{l}}
        \toprule
        \textit{Males}  & \( \tau_a=0.10 \) & \( \tau_a= 0.30 \) & \( \tau_a= 0.50
        \midrule     \\
        Coeff. (s.e.)   & $\underset{(0.0872)}{-0.0049}$ & $\phantom{-}\underset{(0
        F-test (p-value) & $\phantom{-}\underset{(0.304)}{1.057}$ & $\phantom{-}\un
        \midrule
        \textit{Females}  & \( \tau_a=0.10 \) & \( \tau_a= 0.30 \) & \( \tau_a= 0.5
        \midrule \\
        Coeff. (s.e.)    & $\phantom{-}\underset{(0.1373)}{0.0094}$ & $\underset{(0
        F-test (p-value) & $\phantom{-}\underset{(0.3545)}{0.8576}$ & $\phantom{-}\
        \bottomrule
     \end{tabular}
    \begin{tablenotes}[flushleft]
      \small
      \item \textit{Note.} See Table 3. $\tau_a$ denotes the quantile of the distri
    \end{tablenotes}
  \end{threeparttable}
\end{table}
```

In the section about the empirical strategy we already got a feeling for for the effect of the instrument ycomp. Figure 1 shows almost zero residuals which indicates that it is likely that another coefficient don't add any significant value to explain the years of schooling. The results in Table 3 shows the expected result. Given the SHARE dataset the instrumental variable *ycomp* is insignificant at all quantiles for both genders. Besides, this step it is important to analyze if our instrument is a valid instrument. Using the Stock and Staiger rule of thumb, an selected instrument apears to be a

weak instrument if the F-test for it's inclusion is lower than 10. The results, again for both genders, clearly show that the instrument is weak and that all F-tests appear to be insignifcant.[3] Under condition that those result would be representative, males at the 10th quantile of the distribution would face a slightly decrease in educational attainment around -0.5% per extra year of compulsory schooling while females at the same quantile would face an increase of roughly 1% in educational attainment per extra year of compulsory schooling.

In the next step

```
\begin{table}[!htbp]
\captionsetup{labelsep=newline, justification=centering}
  \begin{threeparttable}
       \caption{\textit{Quantile Effects When Education is Treated as Exogenous} \\
    \scriptsize (Sample size : 7,165) By gender (3,735 males and 3,430 females)}
     \begin{tabular}{*{6}{l}}
        \toprule
        \textit{Males}  & \( \tau_u=0.10\) & \( \tau_u= 0.30\) & \( \tau_u= 0.50\)
        \midrule     \\
        $\tau_a = 0.1$    & $\phantom{-}\underset{(1.6418)}{18.5705^{***}}$ & $\pha
\\
         $\tau_a = 0.3$   & $\phantom{-}\underset{(0.1817)}{0.7117}$ & $\phantom{-}
\\
          $\tau_a = 0.5$    & $\phantom{-}\underset{(0.218)}{1.7513}$ & $\phantom{-
           $\tau_a = 0.7$   & $\phantom{-}\underset{(0.4483)}{1.8353}$ & $\phantom{
           \\

           $\tau_a = 0.9$    & $\phantom{-}\underset{(0.4067)}{7.9706}$ & $\phantom
\\
Mean effect+ &6.167885 & 2.649243 & 2.522340 &  3.702072 &  2.544303
\\ \toprule
        \textit{Females} &  \( \tau_u=0.10\) & \( \tau_u= 0.30\) & \( \tau_u= 0.50\
        \midrule
        \addlinespace
        $\tau_a = 0.1$ & $\underset{(0.3065)}{-0.0087}$ & $\phantom{-}\underset{(0.
         $\tau_a = 0.3$   & $\underset{(1.3782)}{-0.414}$ & $\phantom{-}\underset{(
```

[3]Typically those results should lead the model designer to a reconsideration of either the model, the strategy, the data or all three of them. Since this replication study just reproduce given the limitations of the dataset SHARE, this paper continous to replicate the results and ignores the obvious warning signs the latter results show.

```
          $\tau_a = 0.5$   & $\underset{(0.2067)}{-0.0047}$ & $\phantom{-}\underset
           $\tau_a = 0.7$   &  $\phantom{-}\underset{(0.268)}{0.0131}$ & $\underset
           $\tau_a = 0.9$    & $\phantom{-}\underset{(0.2297)}{0.0258}$ & $\underse
           Mean effect+ & -0.07770027 &  5.05951296 & 4.65119676 &  3.55031764 & 9.
        \bottomrule
     \end{tabular}
    \begin{tablenotes}[flushleft]
      \small
      \item \textit{Note.} See Table
    \end{tablenotes}
  \end{threeparttable}
\end{table}
```

the full matrix of quantile treatment effects is estimated by applying the control variate approach as described by **MaKoenk** in the sense of **brunello**. The results can be seen in Table 4. As expected through the results given by the first stage in Table 3, most of the coefficients are insignificant and or unlikely high. Examples are the 10th and 90th quantile for males of τ_a when τ_u and the 30th qunatile for females.

In compariso **brunello** finds that the estimated increase in earnings for males at the lowest qunatile are 7.48% compared to roughly 1800% in the result of the replication study and stay at a very high percantage level for mostly around 700% to 800% percent. The page for the females looks quit the opposite. It seems that, again, males have to catch up extremly since many of the results from the bottom to the highest quantile of ability show some negative coefficients. But there are some mentionable positive results for the latter gender, too. The findings would expect the estimated returns of education of females which find themself in the 30th qunatile of the ability distribution and the highest quantile of the labour market luck distribution of roughly 5000%. If the latter would be the case we would expect much more women from the lower ability distribution to earn a fortune after leaving school with an extra year of education while hitting the jackpot in the "labour market luck lottery" aka presenting an indivdiual in the 90th percentile of the labor market fortune distribution.

In a more realistic world presented by the results of **brunello** females are thoses to catch up at the fight for wages. The authors found that espically women from the lowest percentile benefit from an extra year of compulsory schooling be returns for the surplus education of roughly 10%.

6 Discussion of the findings

Either the quantile regression method fails completly on this kind of dataset or the dataset is not good to show the effect of the choosen treatment *ycomp*. **brunello** add in their section about robustness checks that it is likely that an measurment error of the key models ((7) and (8) in this paper and (6) and (7) in **brunello**) occur but not for the ECHP data, "because years of education are computet there by using the information on the age when full time education was stopped". Since the replication study only uses the data given by the SHARE study, it is possible that the erroenous looking results of this replication might be biased due to this problem. **brunello** further add that another bias in their results might be due to lack of controls for parental background, which they try to control for by using the unenmployment rate, which isn't avaiable for this replication, and the GDP per capita. They argue that other explanation like easier access to credit funds or higher education of the parents could give the treated individuals an unobserved exogenous boost which might be falsly captured by ability. The key problem of this replication study seems to be the SHARE dataset. Rembering the results of Table 1, one can easily see that many of the individuals are not accounted as compliers. For example, Spain, which is the country where most individuals come from, compared to all other countries, has a rate of only 2% of individuals affected by the instrument. The Netherlands show the same qoute and beeing the second highest country in terms of share of observed individuals. Greece, Belgium and Denmark show extremly low compliance rates aswell. Given these numbers it should be no surprise that the instrument is never significant due to lack of observed treated individuals. An plausible explanation why **brunello** added these data to their study, might be as mostly control individuals for the treatment effect.

7 Conclusions

The results of the replication study shed light on how returns of compulsory schooling to both genders affects their earnings. Mostly for men in a positive manner while their gender counterpart should leave school as early as possbile since in the world of the replicated results women even loose if they stay in school much longer as needed. The latter do not hold for those women of the 30th quantile of ability who mostly benefit extremly if they are able to not the the lowest end of the labour market fortune distribution. Further attempts to replicate the results and want to achieve plausible results are highly recommended to not limit their study to only the SHARE dataset.