



Econometric Evaluation

Reserach Paper:
Secondary Analysis and Evaluation of the Tristan Auvray 's and Joel Rabinovich's
"The financialization-offshoring nexus and the capital accumulation of US non-
financial firm" (2019)

Economic Policies in the age of Globalisation

Heterodox Econometrics

Instructor:

Hugo Harari-Kermadec

Author:

Lukas Bernhard Thomas Maria Marschallek

Submission Date:

23.01.2020

Table of Contents

<u>I. Introduction.....</u>	<u>3</u>
<u>II. Results of Auvray and Rabinovich.....</u>	<u>3</u>
<u>III. Databases.....</u>	<u>5</u>
<u>IV. The Variable Investment.....</u>	<u>6</u>
<u>V. Replication of Figures 1,2 & 5.....</u>	<u>7</u>
<u>V.1. Replication of Figure 1.....</u>	<u>7</u>
<u>V.2. Replication of Figure 2.....</u>	<u>8</u>
<u>V.3. Replication of Figure 5.....</u>	<u>8</u>
<u>VI. Replication of Table 6.....</u>	<u>11</u>
<u>Table 3 / Replication of Table 6.....</u>	<u>15</u>
<u>Robustness of the results.....</u>	<u>16</u>
<u>VII. Conclusion.....</u>	<u>17</u>
<u>VIII. Annex.....</u>	<u>18</u>
<u>IX. Bibliography.....</u>	<u>22</u>
<u>X. R-Code.....</u>	<u>23</u>
<u>Figures 1,2,5 and Residual Plot Figure 5.....</u>	<u>23</u>
<u>Table 6 and 7.....</u>	<u>31</u>

I. Introduction

Ever since the decline of Fordist mode of accumulation towards the end of the 20th century, the process of financialization set in. The meaning of the term and what it entails is highly debated, but a vague and widely accepted definition of financialization has been brought forward by the economist Gerald A. Epstein: “the increasing role of financial motives, financial markets, financial actors and financial institutions in the operation of the domestic and international economies” (Epstein 2005, 3). It is characterized by a proportional overgrowth of the financial sector in relation to the real sectors of the economy.

This heightened importance in turn is resembled in the changed behavior and structure, of financial institutions as well as of non-financial corporations (NFCs). As identified by scholars, the two main channels that characterize the financialization of NFCs are a “higher proportion of financial assets compared to non-financial ones” and a “higher amount of resources diverted to financial markets” (Auvray and Rabinovich 2019, 1). While those two channels and a consequential decline in aggregate investment have become subject of intensive scrutiny, the potential linkage between financialization and offshoring represents a less empirically researched aspect. The authors Tristan Auvray and Joel Rabinovich picked up this point in their 2019 published paper “The financialization-offshoring nexus and the capital accumulation of US non-financial firm”. They examine this linkage by looking at industry-level information on offshoring and company data regarding capital accumulation for companies based in the United States of America (US).

The goal of this present report is to replicate, reanalyze and evaluate the robustness of their findings from an econometric viewpoint and to further discuss potential findings.

II. Results of Auvray and Rabinovich

Guided by a phenomenon also known as ‘investment-profit-puzzle’, a counter-intuitive trend for heterodox economists that is characterized by a decline in aggregate investment coupled with high profits, Auvray and Rabinovich set out in order to empirically validate the hypothesis brought forward by Milberg (2008) and Milberg and Winkler (2010; 2013). Postulating that the phenomenon of the ‘investment-profit-puzzle’ is only made possible by gains from the offshoring of productive capacities in order to sustain financialization, the aforementioned authors brought forward an intuitive explanation.

On a bigger level the authors conclude that their data and the way they used it¹ indicates that there is indeed a linkage between financialization and offshoring, but several factors exists which carry implications for the strength of this linkage (Auvray and Rabinovich, 2019, 1).

¹ The used macroeconomic data will be discussed extensively in Chapters 3 & 4

Starting out with the verification of the financialization thesis, the authors see increased distribution of profits at the costs of investment as apparent in the whole US-economy. When only focusing on listed firms this effect is much more pronounced (ibid., 2). A similar trend is identified regarding financial payouts in line with the shareholder value orientation, as firms shifted towards strategies of downsizing and distributing (increased transfer of operating surplus towards shareholders through share buybacks and dividends) instead of retaining and reinvesting (ibid., 4-8).

After having demonstrated the phenomenon of financialization, the authors turn to offshoring. The authors look at the increase of foreign direct investment (FDI) in relation to US investment, the increase in stock of FDI measured as percentage of the GDP, and the increase in stock of for the rest of the world (ROW) and Mexico as part of the US outward FDI stock in order to conclude that the phenomenon also exists (ibid., 8-10). Auvray and Rabinovich furthermore deduce that the institutional set-up carries implications for the linkage of financialization and offshoring: If companies are capable of offshoring of non-core non-energy activities, the investment is reduced and non-core offshoring can thus explain the prevalence of firms with high financial payouts and low investment (ibid., 10-16). This is done by the construction of two offshoring variables, one looking at core and the other looking at non-core non-energy activities, with the help of input-output data for different sectors.

Applying this distinction, the authors find that high payouts combined with low investment, most likely occurs in companies that are a part of industries that are heavily involved in global value chains (GVCs) (ibid., 3). In working with the different sectors, the authors explicitly mention that utilities, wholesale and retail sectors are special cases. Utilities, as dividend and investment in this industry may be strongly affected by regulation, indeed affects the data and generates a negative and significant correlation between stock repurchases and investment for one of their subsamples (ibid., 27). The wholesale and retail trade industries provide another difficulty as they import final products rather than intermediary ones (ibid., 13). Overall though, the authors conclude that the inclusion of those industries does not affect the general results (ibid., 16). Drawing from the aforementioned aspects, the authors reveal additionally that investment of firms in low offshoring sectors is not significantly correlated to their financial payouts.

By means of empirical robustness checks the authors conclude that there is a significant linkage between financialization and offshoring (ibid., 27-32). Based on firm and industry-level data the authors ultimately verify their proposed framework, which captures the channels by which a firms' offshoring affects employment and investment (Annex Figure 1).

After having provided a short overview of the paper by Auvray and Rabinovich, the analysis will start out by evaluating the used databases. Following the usage of the variable of the Investment

will be researched, before their results will be further analyzed by an intended replication and interpretation of the results.

III. Databases

For their empirical validation Auvray and Rabinovich rely on three databases, two of which are going to be assessed: Standard and Poors' (S&P) Compustat Annual Industrial Database (Compustat) and the World Input-Output Database (WOID).

The Compustat database, which is developed and marketed by S&P, covers approximately 98% of the world's market capitalization with historical key restated and un-restated data. S&P standardizes the data used in the Compustat report, according to the data subsets – Global Data or North American Data – and the then applicable country specific accounting principles. As Auvray and Rabinovich focus on US companies they use the North American Data subset. Hence their data is standardized according to the “financial statement presentation and specific data item definitions” in order to provide “consistent comparable data with which to analyze companies and industries.” (S&P Compustat 2019, 2). The American company S&P, as one of the ‘Big Three’ credit-rating agencies, alongside Moody's and Fitch, is specialized in the provision of financial services, more precisely the provision of financial research and analysis and thus has an incentive to sell their product to customers. This implies that their products are being marketed as seen in this statement: “We empower people to make informed, confident decisions” (About Us | S&P Global Ratings).

The second database, the WOID, is used by the authors to look at the industry-level information on offshoring. Initially funded by the European Union in order to assess the effects of globalization on socio-economic and environmental trends, it is a freely accessible database. Through the linkage of statistics on international trade with national supply and use tables, an international input-output table has been created that covers all major economies and at least 80% of the world's GDP from 1995 until 2014. Limiting this table to a specific time period already poses a limitation for the research of the authors: Starting point of financialization can be traced back to the 1980s and the crisis of 1973, which heralded the end of the ‘Golden Age of Capitalism’ (Fasianos, Guevara, and Pierros 2018). Thus, the linkage cannot be verified by analyzing only two decades of a relatively young phenomenon. In order to make the given data comparable the authors of the WOID had to resort to standardization, as national supply and use tables do not only differ across countries, but also across time (Dietzenbacher et al. 2013).

Both, for this paper relevant databases, use different standards: While the WIOD follows the third or fourth revision - data set depended - of the International Standard Industrial Classification of All Economic Activities (ISIC), Compustat follows the Standard Industrial Classification (SIC). In

order to circumvent this problem, the authors used a concordance table. Congruent measurements and standardized variables consequently enabled a combination of the two databases.

While standardization of data is essential in order to aggregate and compare data for different companies, it does not necessarily lead to a more precise depiction of reality, since measurements, just like the data sets and their creation – meaning the chosen and not-chosen variables -, are influenced by ideology. Alain Desrosières described it as follows: “The accountant’s work carries within it a language of reality that defies the statistician’s language; the accountant constructs, from the ground up, a universe to serve as a reference for action.” (Desrosières 2001, 351–52). These constructed languages therefore provide a viewpoint from which the reality can be approximated but never fully reached. As Desrosières further explains: “The terms ‘index’ and ‘indicator’ suggest that the reported measures are like the visible symptoms of a hidden reality that is impossible to reach directly.” (ibid., 353).²

It should be taken into consideration that, throughout the analysis of the paper by Auvray and Rabinovich, all assumptions are based on, and limited by, the data that is provided. This trend becomes even more drastic when statistics or econometrics are used in order to ‘depict reality’.

Thus ultimately one has to be careful with the interpretation of one’s findings and ultimately critically evaluate the data used. The fact that one has data supporting his ideas does not mean that his data is necessarily correct and appropriately used. One always only captures a frame of the reality which is limited by the used terms and the data set: Auvray and Rabinovich for example were only able to sample 2,049 companies that represent 68% of the total US market capitalization (Auvray and Rabinovich 2019, 20), hence disregarding 32% of the ‘reality’ due to the lack of standardized variables.

IV. The Variable Investment

Comparing the definition and usage of the term investment through different stages such as subsistence, classical, and modern economies, it is possible to observe that its definition and usage has varied or as Laurent Thévenot would call it, it has been subjected to coding. Thévenot concluded that the usage of coding and hence the usage of standardized variables comes at the cost of data, a process he termed ‘investment in form’ (Thévenot 1984). He points out that each definition entails other features and therefore creates a different ‘reality’ (ibid., 8-10).

Investment within the frame of the paper by Auvray and Rabinovich depends on the respective level the authors look at: on the industry level the investment is defined as gross fixed capital formation (GFCF), drawn from the financial accounts of the US, while on the company level

² Desrosières describes four possible attitudes towards ‘reality’ that each have different tools, “reality tests” as Desrosières calls them, in order to verify and articulate one’s reality. See Desrosières (2001)

investment is drawn from the Compustat database as capital expenditures (Capx). As each of those definitions of investment has a different asset boundary, meaning the precise financial measurable quantification of investment, the authors committed to such an ‘investment in form’. The Capx, as drawn from Compustat database with the annual data item number 128, represents the cash outflow or funds used for additions to the company’s property, plant, and equipment. The GFCF on the other hand, based on the System of National Accounts (SNA), generally looks at a much wider asset boundaries and goes beyond the simple purchase of fixed assets. The GFCF asset boundaries exemplary include computer software or a recognition of research and development, both variables that are lacking in the Compustat definition. While both concepts are generally very similar, both provide different perspectives and thus lead to a different ‘investment-in-form’.

It furthermore has to be mentioned that in opting for those definitions, the authors limit their view to the ‘modern’ definition of investment which is disregarding several factors such as investment into human capital or the acquisition of land. Thus, when talking about a general macroeconomic trend it might prove difficulty to rely on financial variables that only capture a framed reality. If Capx and the GFCF would include all measures that affect the output other variables have to be taken into account that increase the value of investment. Exemplary, investment in non-scientific research, such as brand equity or organizational capacities, or expenditures on human capital would have to be treated as investment. While it is not the frame of the work to estimate such a value, it is not hard to see that including those increased values for investment would bring about a change in several figures of their paper.

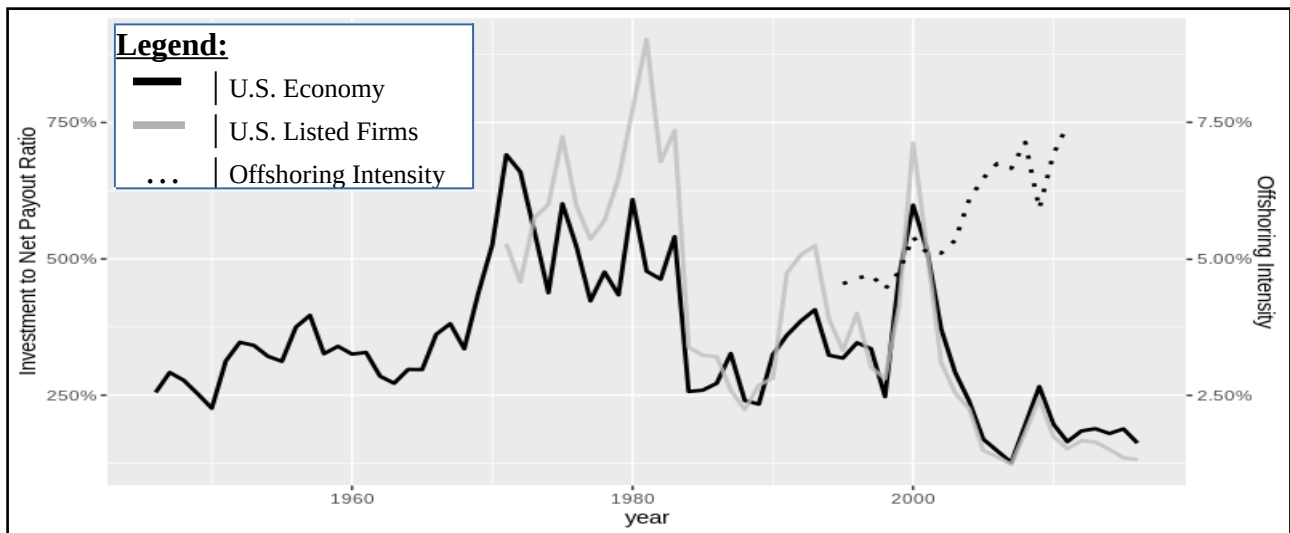
V. Replication of Figures 1,2 & 5

V.1. Replication of Figure 1

The general idea of Figure 1 is to show the negative relationship of gross fixed investment and net financial payouts for the US for listed firms as well as for the economy as a whole. This trend is much more pronounced for the listed firms. Auvray and Rabinovich state that for the calculation of the offshoring intensity they used “the updated WIOD [...] organized following the ISIC third revision” (Auvray and Rabinovich 2019, 20), which leads to believe that they use the WIOD data 2013 release. This data set only contains values from 1995 until 2011 though, as opposed to the 2016 release which is created in accordance with the fourth ISIC revision and containing data from 2000 until 2014. The authors graph offshoring intensity from 1995 up until 2014, but the used dataset thus does not become apparent. The results of the 2013 WIOD database were much more in line with their graph but did not exceed the year 2011. Consequently, it was chosen to use this database in the present paper.

Figure 1:

Investment as a ratio of net financial payouts for the US economy and US-listed firms, 1946 - 2016



Source: Table Z1, WIOD and Compustat, Author's calculations

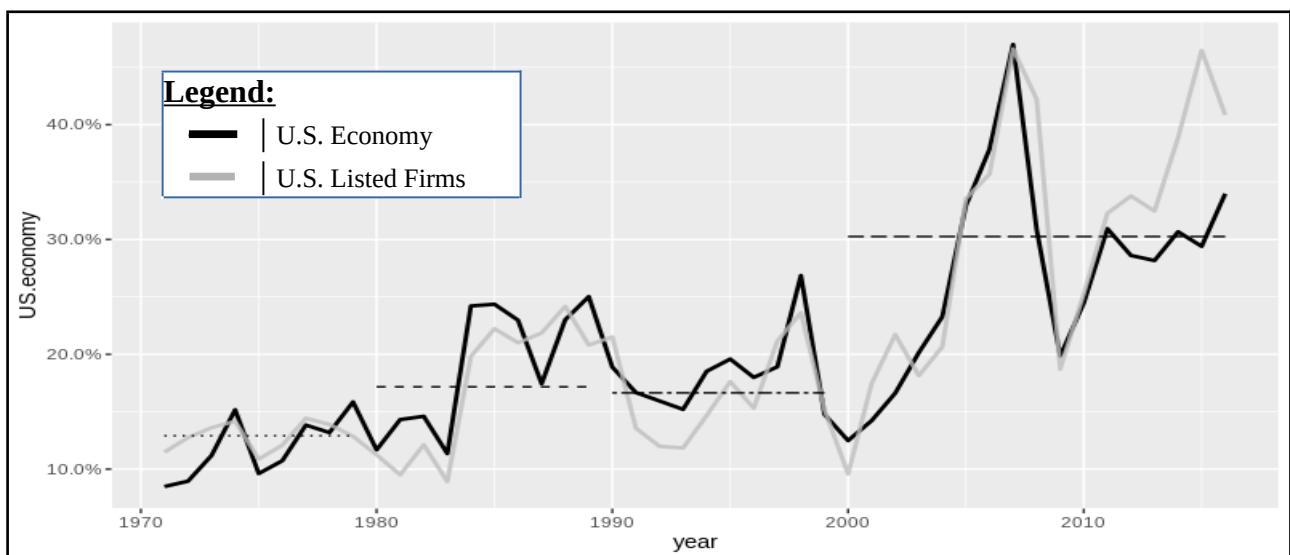
Replication of: (Auvray and Rabinovich 2019, 2, Figure 1)

V.2. Replication of Figure 2

The goal of figure 2 is to display the increasing financial payouts as a percentage of operating surplus for the US economy and US-listed firms.

Figure 2:

Net financial payouts as a percentage of operating surplus for the US economy and US-listed firms, 1971 - 2016



Source: Table Z1 and Compustat, Author's calculations

Replication of: (Auvray and Rabinovich 2019, 4, Figure 2)

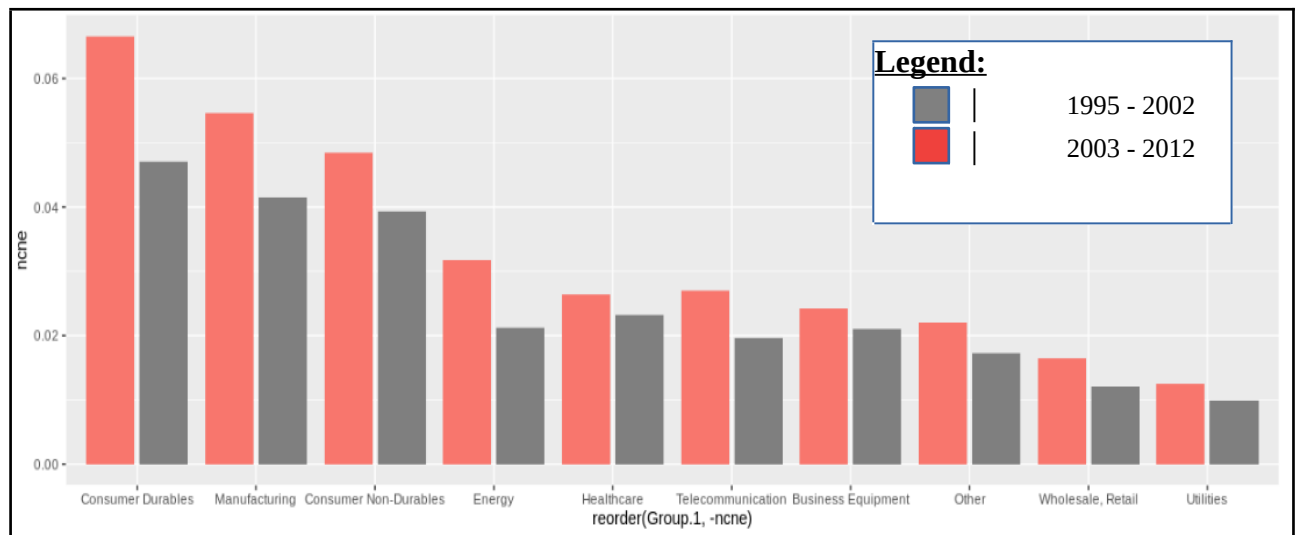
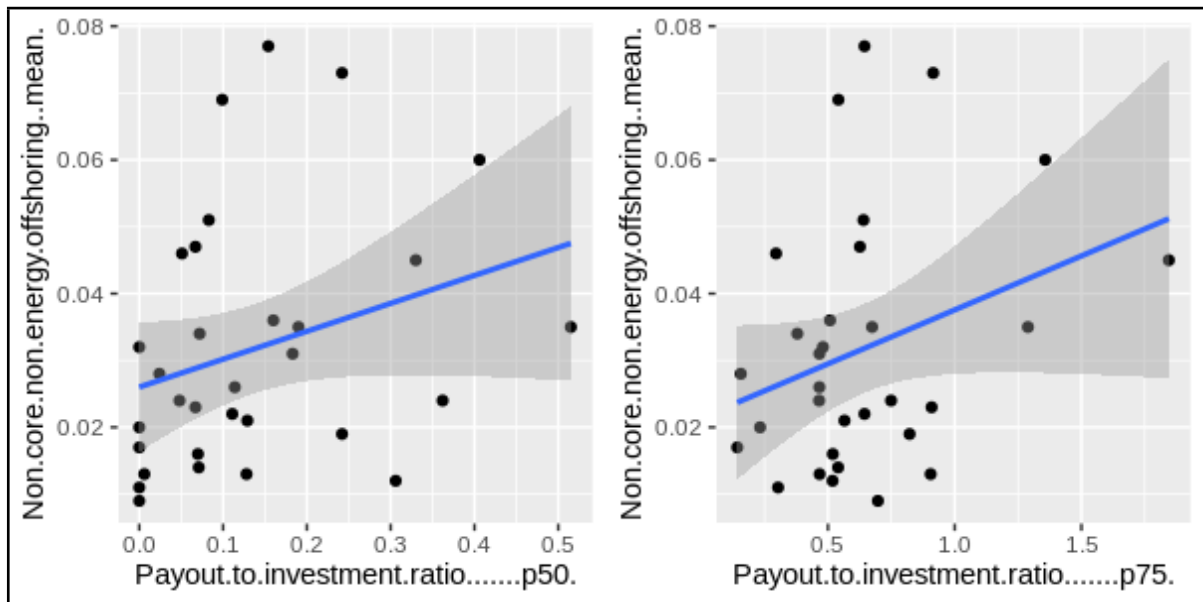
V.3. Replication of Figure 5

Figure 5 is at the core of the authors paper as it tries to illustrate that the existence of financialized NFCs is linked to offshoring. The idea behind the scatter plots is to portray a linear and positive

correlation between the payout-to-investment ratio and offshoring as it reinforces the linkage between financialization and offshoring. This is further increased by the fact that offshoring in relation to the 75th percentile of the payout-to-investment ratio creates an even steeper positive relationship. Hence the payout-to-investment ratio increases the more companies offshore their non-core non-energy activities.

Figure 3:

Non-core offshoring and payout-to-investment ratio, 1995-2011



Source: Compustat, Author's calculations

Replication of: (Auvray and Rabinovich 2019, 13, Figure 5)

Looking at the graphic replication of both linear models, for the median or the 75th percentile of the payout-to-investment ratio, one can see that a lot of the observations are neither laying on the estimated model and neither in the confidence interval, which is verified when running a simple linear regression.³

³ The paper will only replicate and discuss the results of the regression for the median ratio

By running a simple linear regression, the correlation between two variables can be evaluated. In this specific case this would mean that one tries to estimate with the regression the correlation between the payout-to-investment-ratio and the non-core non-energy offshoring. While the assumptions of the linear model are never perfectly met, it must be checked if they are even reasonable assumptions to work with.

Figure 4:

Results of the linear regression

```
Call:
lm(formula = Non.core.non.energy.offshoring..mean. ~ Payout.to.investment.ratio.....p50.,
    data = fig5)
Residuals:
    Min       1Q   Median       3Q      Max
-0.026789 -0.013099 -0.004772  0.005598  0.044558
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.026012   0.004735   5.494 6.43e-06 ***
Payout.to.investment.ratio.....p50.  0.041756   0.025085   1.665  0.107
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01821 on 29 degrees of freedom
Multiple R-squared:  0.08721, Adjusted R-squared:  0.05574
F-statistic: 2.771 on 1 and 29 DF, p-value: 0.1068
```

Source: Author’s calculations

Looking at the p-values the statistical significance for the estimates can be determined. For the estimation of the relationship between the mean non-core energy offshoring and the median of the payout-to-investment ration one can note a very high p-value (Figure4). Large p-values, above the determined level of significance, indicate that there is a very small probability of the null hypothesis being inconsistent with the data. As the null hypothesis for the simple linear regression states that the slope is equal to zero, this means that the statistical significance between payout-ratio and offshoring should be rejected unless the chosen significance level is above 10%. Looking at the residual standard error, telling one how well or poorly the model does at predicting y-values in comparison to the used data, one can conclude that it is poor in relation to the data set. As the adjusted R² value is very low at 0.087, it can be noted that the model explains only around 9% of the observed variability of the response data around its mean.

Looking at the residual plots of the function, one can also observe a few irregularities that point towards flaws in the model (Annex Figure 2)⁴. These are the indication of a structure of mostly negative residuals, especially when looking at large and small values, and a possible issue with the normally distribution of the residuals in the Q-Q plot especially for observations #1,2,3.

⁴ See the [annex](#) for the graphic display of the residual plot.

Overall it has to be said, that the authors' chosen representation, following this present evaluation using a simple linear model, does not present a statistical meaningful link for offshoring and financialization. Furthermore, it has to be taken into consideration that the explicit goal of the authors is to explain the link of offshoring on investment, yet they chose to set offshoring as the dependent variable and the payout-to-investment ration as the independent variable.

VI. Replication of Table 6

As a first step towards the replication of Table 6, it was intended to clean the Compustat database in line with the authors. Yet, following the authors' guidelines no entries are removed and it is thus assumed that a cleaned database was received (see annexed R-code). This cleaned database consists of 2326.471 firm-level observations annually, compared to 2049 companies on average annually as analyzed by Auvray and Rabinovich. This difference in analyzed companies might explain the different results when looking at the relevant variables and their Winsorized means (Table 1).

Table 1:
Descriptive statistics:

<u>Variable Name</u>	<u>Mean – paper</u>	<u>Mean – calculated</u>
Capital expenditures	0.242	0.221
Profits	-0.052	-0.239
Sales	9.112	9.772
Long-Term debt	2.110	2.001
Interest Expenditure	0.253	0.283
Interest and Investment Income	0.041	0.038
Dividends	0.036	0.034
Stock issue	0.484	0.565
Stock repurchase	0.109	0.104
Net debt issue	0.255	0.148
Internal finance	1.302	1.447
Tobin's Q	1.927	1.958
Narrow or core offshoring*	0.024	0.025
Non-core non-energy offshoring*	0.028	0.028

Source: Auvray and Rabinovich (2019, 21, Table 4) and Author's calculations

The authors apply Winsorization, the process of a transformation of extreme values to reduce the effect of outliers, for the 0.5th and 99.5th percentile of their dataset. Winsorizing with proper justification, e.g. Winsorization of measurement mistakes or impossible values presents a good way of dealing with those accounted errors as the data would in their original condition falsify

results and it thus accounts for outliers in the paper. If data is removed simply on the basis of outlier values, one might end up transforming information, that either is relevant for the whole model or as such does not have relevant impact as in a large enough dataset the impact would be fairly small. Regrettably the authors fail to justify the application of the mechanism, but at as the authors apply the process of Winsorization only for very small percentiles, the result is a change in the data in comparison to the original population which ultimately gives us a less skewed distribution. Yet, in order to reanalyze the data as comparable as possible, and regardless of the missing justification, the data used in the replication is also Winsorized. Therefore, similar values in the both calculations, theirs and the one in the paper, are observable (Table 1).

After having established a similar starting point in the two databases one has to note that a lot of the data provided to us in the dataset contained NA-values, meaning non available or missing values (Table 2), whose origin is unknown. In order to work with the data and in order to not lose too many observations, certain values are transformed to 0. Chosen were only those economic variables where a transformation makes sense from economic viewpoint or where it was necessary due to the amount of possible data-loss. As such no transformation of long-term debt occurred, despite 2294 observed NAs, as it is hardly imaginable that listed firms from the Compustat database are completely without debt.

It remains unclear, whether Auvray and Rabinovich performed a similar transformation or whether the issues are intrinsic to the dataset, but if the authors proceeded and transformed the data, simply because no data was available, it seems a bit arbitrary.

Differing calculation methods for different variables change their significance, as mentioned in earlier chapters and as it is visible for Tobin's Q. Here Tobin's Q was calculated with either the market-cap at the end of the year or with the average market-cap and then proceeded twice with the calculated values, once with the end of the year (Table 6) and once with the average market-cap (Annex Table 2). In Annex Table 2 significant differences to Table 3 were highlighted. A drastic change in the significance of Tobin's Q as a regressor can be seen. This change in the significance of Tobin's Q, which loses significance almost throughout the whole data-set, could be explained by the volatility of shared prices that usually occurs at the end of the year, either pushing the shared prices strongly up- or downwards, as it has been noted and became subject of asset-pricing theory.⁵ Accounting only for a specific date and not looking at the development over the whole year, it is much more likely that the data is affected by such extraordinary events that have a detrimental impact on the data as it is observable.

⁵ As it is not frame of the paper though, I will refrain from explaining the issue at full.

Table 2:⁶

Descriptive statistics 2:

<u>Variable Name</u>	<u>Relevant Variables</u>	<u>NA's</u>	<u>NA's to 0?</u>
Capital expenditures	IQ_CAPEX	0	No
Profits	IQ_OPER_INC	3	No
Sales	IQ_REV	74	No
Long-Term debt	LT_DEBT	2294	No
Interest Expenditure	IQ_INTEREST_EXP	359	No
Interest and Investment Income	IQ_INTEREST_INCOME	15034	Yes
Dividends	IQ_TOTAL_DIV_PAID_CF	24164	Yes
Stock issue	IQ_COMMON_ISSUED	7528	Yes
	IQ_PREF_ISSUED	38471	Yes
Stock repurchase	IQ_COMMON_REP	24388	Yes
	IQ_PREF_REP	38538	Yes
Net debt issue	IQ_TOTAL_DEBT_ISSUED	10231	Yes
	IQ_TOTAL_DEBT_REPAID	2150	Yes
Internal finance	IQ_CASH_ST_INVEST	209	No
Tobin's Q (End of Year)	IQ_MARKETCAP_average	32	No
	IQ_TOTAL_LIAB	0	No
	IQ_TOTAL_ASSETS	0	No
Tobin's Q (Avg. Market_Cap)	IQ_MARKETCAP_end_of_year	359	No
	IQ_TOTAL_LIAB	0	No
	IQ_TOTAL_ASSETS	0	No

Source: Authors' calculations

Furthermore, one can also see changing p-values for variables such as long-term debt and net-debt-issue, which depending on the chosen significance level either makes the variable significant or not. This can be explained with the change in the observed and researched variables. As the variable IQ_MARKETCAP_end_of_year has 359 NAs, the variable IQ_MARKETCAP_average has only 32 NAs and hence 327 NAs less. As missing values were not transformed to 0, as a share price of 0 would represent an unphysiological value, observations that by themselves have an impact on the results as well as on the results of what the authors try to explain, were eliminated. Thus, one can clearly observe that a small change in the dataset carries implications for the results.

⁶Note: All Variables are divided by IQ_NPNE with 0 NA's

In trying to assess the impact on the slowdown of investment the authors measure the relatedness of financialization and offshoring, taking into consideration always lagged variables as post-Keynesian theory dictates that firms take investment decisions under fundamental uncertainty on previous experiences. This paper will follow their proposed framework in order to calculate the impact of financialization on investment, which is further for the calculation of offshoring expanded, replicating their offshoring model (Auvray and Rabinovich 2019, 19).⁷

Figure 5:

Statistical specification of the financialization model:

$$\ln(I/K)_{it} = \alpha_0 + \alpha_1 \ln(I/K)_{(i,t-1)} + \alpha_2 \ln(\pi/K)_{(i,t-1)} + \alpha_3 \ln(S/K)_{(i,t-1)} + \alpha_4 \ln(Q)_{(i,t-1)} + \alpha_5 \ln(LONGDEBT/K)_{(i,t)} + \alpha_6 \ln(INTEXP/K)_{(i,t-1)} + \alpha_7 \ln(INTINC/K)_{(i,t-1)} + \alpha_8 \ln(DIV/K)_{(i,t-1)} + \alpha_9 \ln(STKISSUE/K)_{(i,t-1)} + \alpha_{10} \ln(STKREP/K)_{(i,t-1)} + \alpha_{11} \ln(NETDEBTISSUE/K)_{(i,t-1)} + \alpha_{12} \ln(INTERNF/K)_{(i,t-1)} + \gamma_{it} + \sum_{t=1996}^{t=2011} \beta_t + \varepsilon_{(it)}$$

Source: Auvray and Rabinovich (2019, 18, equation 1)

Including the lagged value of investment as an explanatory variable introduces two different sources of autocorrelation, namely “autocorrelation due to the lagged dependent variable among regressors and individual persistent effects “ (Auvray and Rabinovich 2019, 20). The authors opted for the computed generalized method of moments (GMM) proposed by Arellano-Bond (1991).

Trying to replicate their results it was decided to use the GMM model.

Regarding the general results (Table 3), it was possible to replicate some of their results.⁸ While finding very similar results in terms of impact and significance for capital-expenditures and internal funds, the present results started deviating from the results of Auvray and Rabinovich in the significance of sales and profits. While in present case profits kept across the whole dataset at at least a significance of 10%, therefore also being more in line with the theory (Auvray and Rabinovich 2019, 17), a negative and significant influence of sales on investment was found. This represents a contradiction with the findings of Auvray and Rabinovich as well as with the theory the paper was placed in. Further a deviation from their results in terms of long-term debt was observed, as their results display a negative and not significant relationship, while here a positive and partially significant relationship could be found.

⁷ While the authors include a coefficient for the age, here it will be discarded

⁸ This paper follows the standard output significance codes of R. The significance translates to theirs according to Annex Table 1

Table 3: with Tobin's Q calculated with IQ_MARKETCAP_end_of_year

Dependent: $\ln(I/K)_{i,t}$	Financialisation model			Financialisation and offshoring model			High-non-core non-energy offshoring sectors			Low-non-core non-energy offshoring sectors		
	All	Large	Small	All	Large	Small	All	Large	Small	All	Large	Small
Capital Expenditures	0.234*** (0.021)	0.264*** (0.033)	0.207*** (0.031)	0.237*** (0.021)	0.272*** (0.034)	0.210*** (0.031)	0.209*** (0.030)	0.216*** (0.043)	0.157*** (0.039)	0.249*** (0.027)	0.315*** (0.044)	0.254*** (0.044)
Profits	0.019*** (0.005)	0.028*** (0.007)	0.017** (0.005)	0.019*** (0.005)	0.027*** (0.007)	0.017** (0.005)	0.026*** (0.005)	0.031*** (0.008)	0.027*** (0.007)	0.0144* (0.007)	0.032* (0.015)	0.013. (0.007)
Sales	0.0241** (0.009)	-0.020* (0.010)	0.0463*** (0.011)	0.023** (0.009)	-0.024* (0.010)	0.045*** (0.011)	0.020* (0.010)	-0.022. (0.012)	0.032* (0.013)	0.022. (0.012)	-0.035* (0.014)	0.049*** (0.0137)
Long-term debt	0.007. (0.004)	0.011 (0.008)	0.010. (0.006)	0.007 (0.004)	0.007 (0.008)	0.010. (0.006)	0.007 (0.005)	-0.002 (0.006)	0.009 (0.008)	0.011. (0.006)	0.008 (0.013)	0.006 (0.008)
Tobin's Q	0.049*** (0.005)	0.054*** (0.006)	0.047*** (0.008)	0.048*** (0.005)	0.054*** (0.006)	0.047*** (0.008)	0.050*** (0.007)	0.055*** (0.009)	0.057*** (0.011)	0.045*** (0.008)	0.047*** (0.009)	0.034** (0.011)
Interest expenditure	0.0002 (0.019)	-0.056. (0.032)	-0.006 (0.020)	0.001 (0.019)	-0.045 (0.035)	-0.006 (0.020)	0.026 (0.025)	0.025 (0.029)	0.020 (0.024)	-0.026 (0.024)	-0.061 (0.048)	-0.005 (0.024)
Interest and investment income	0.019 (0.047)	0.014 (0.053)	0.027 (0.064)	0.013 (0.047)	0.011 (0.055)	0.025 (0.064)	-0.042 (0.048)	-0.072 (0.048)	0.007 (0.065)	0.012 (0.061)	0.010 (0.080)	-0.032 (0.088)
Dividends	-0.054 (0.037)	-0.045 (0.036)	-0.026 (0.059)	-0.056 (0.037)	-0.049 (0.036)	-0.028 (0.059)	0.009 (0.044)	-0.030 (0.037)	0.154* (0.061)	-0.134** (0.045)	-0.066 (0.051)	-0.163* (0.074)
Stock issue	0.007 (0.004)	0.015* (0.006)	-0.002 (0.005)	0.007 (0.004)	0.015* (0.006)	-0.001 (0.005)	0.006 (0.007)	0.0344*** (0.010)	-0.013 (0.009)	0.006 (0.005)	0.001 (0.007)	0.003 (0.007)
Stock repurchase	0.002 (0.006)	-0.008 (0.006)	0.015 (0.012)	0.002 (0.006)	-0.008 (0.006)	0.014 (0.012)	-0.001 (0.008)	-0.008 (0.007)	0.002 (0.013)	-0.001 (0.005)	-0.012 (0.009)	0.016 (0.155)
Net debt issue	-0.004. (0.002)	0.001 (0.003)	-0.005 (0.003)	-0.004. (0.002)	0.001 (0.003)	-0.005 (0.003)	-0.004 (0.003)	-0.003 (0.003)	-0.006 (0.005)	-0.0001 (0.003)	0.006 (0.004)	-0.001 (0.004)
Internal finance	0.040*** (0.006)	0.030*** (0.006)	0.048*** (0.009)	0.040*** (0.006)	0.030*** (0.006)	0.047*** (0.009)	0.033*** (0.007)	0.026*** (0.007)	0.045*** (0.010)	0.040*** (0.003)	0.034*** (0.009)	0.045*** (0.012)
NCO	-	-	-	0.139 (0.172)	0.440* (0.182)	-0.136 (0.302)	0.086 (0.195)	0.570** (0.204)	-0.322 (0.356)	0.182 (0.369)	-0.247 (0.349)	0.424 (0.569)
NCNE	-	-	-	0.147 (0.238)	0.077 (0.236)	0.509 (0.478)	-0.152 (0.299)	-0.299 (0.276)	0.090 (0.593)	1.52*** (0.455)	0.969* (0.440)	2.56* (0.997)
Observations	28877	15568	12266	28827	15528	12259	14698	7815	6362	14138	7714	5892
Firms	5583	2551	3947	5583	2551	3947	2864	1253	2078	2719	1298	1869
ar1p	0	0	0	0	0	0	0	0	0	0	0	0
ar2p	0.525	0.077	0.154	0.447	0.111	0.133	0.762	0.091	0.767	0.390	0.298	0.079
Hansen-Sargan	0.008	0.003	0.488	0.015	0.002	0.515	0.098	0.015	0.388	0.774	0.148	0.938

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: Author's calculations

Differences also arise comparing the results of the inclusion of offshoring in the estimation. The authors found a “negative and significant elasticity of 0.039” (ibid., p.26) for all non-financial firms and a 0.026 elasticity for narrow or core offshoring. Furthermore, they conclude that “results are in line with the framework [...] and support the idea that firms in industry with a high level of offshoring non-core activities can have a lower demand for capital since part of their production is probably outsourced.” (ibid., p.26). Their main hypothesis is furthermore supported by results that, according to them, the effect played by offshoring on the rise of financialization.

While it is possible to observe an increase in significance of the dividend channel, when looking at the differentiated sectors - into high- and low offshoring sectors – one is also able to observe a positive elasticity of dividends for small high non-core non-energy sectors, which contradicts with their results, their story of financialization and on top of that with the proposed theoretical framework. The biggest contradiction can be found though when looking at the estimation results for both kinds of offshoring. The here found results do not display a significant negative elasticity in regard to non-core non-energy offshoring, but they much rather paint the opposite picture, at least in the case of low non-core non-energy offshoring sectors. In results of the paper one can find a positive highly significant relationship. Regarding the narrow or core offshoring two found cases were significant at much higher elasticities. As their results with the given dataset cannot be reproduced, their results cannot be confirmed here.

Robustness of the results

Having discussed the obtained results, it is also essential to examine the robustness of the results and the output of the GMM, just as the authors did. When computing the results of the GMM-model results for autocorrelation of the first order difference equation residuals and the second order are also received. As expected, when including the lagged value of investment as an explanatory variable, a persistent source of autocorrelation in the first order has been added to the paper (Auvray and Rabinovich 2019, 20). This autocorrelation is confirmed by the ar1 results (Table 3). The further diagnostics test for the paper also reveal one cannot reject the absence of the second-order serial correlation in disturbances, depending on the chosen significance level but when assuming a 5% level, as the lowest ar2 value at 0.077 exceeds the 5% significance level (Table 3). Seeing thus, that the expected first order serial correlation exists as expected and the second order serial correlation in disturbances can be rejected, based on a 5% significance level, a certain robustness in regard to autocorrelation for the model has thus been proven.

Another result that is automatically received when running the GMM model is the result for the Hansen Sargan test, which is used for overidentified equations in order to test the validity of the instruments used. Under the null hypothesis of the Hansen Sargan test all instruments are jointly uncorrelated with the term and one ends up rejecting it. This implies that at least one instrument is

not valid. As persistently low p-values results are received (Table 3) a possible problem of overidentification can be identified, but in the frame of this paper this issue will not be further pursued.

Another thing which could be addressed in is the question of non-stationarity of data within the frame of the paper. Non-stationarity means that there is deterministic dynamic, something that varies for deterministic (not random) reasons. This means, one has to make sure that it is not capital expenditure that is explained through variables, that just as capital expenditure increases over time. In the case of the paper, as the authors have already transformed their data through the creation of the log and furthermore divided it by the capital stock, it is simply assumed that all variables are stationary, even though it is not definitely sure that this is sufficient.

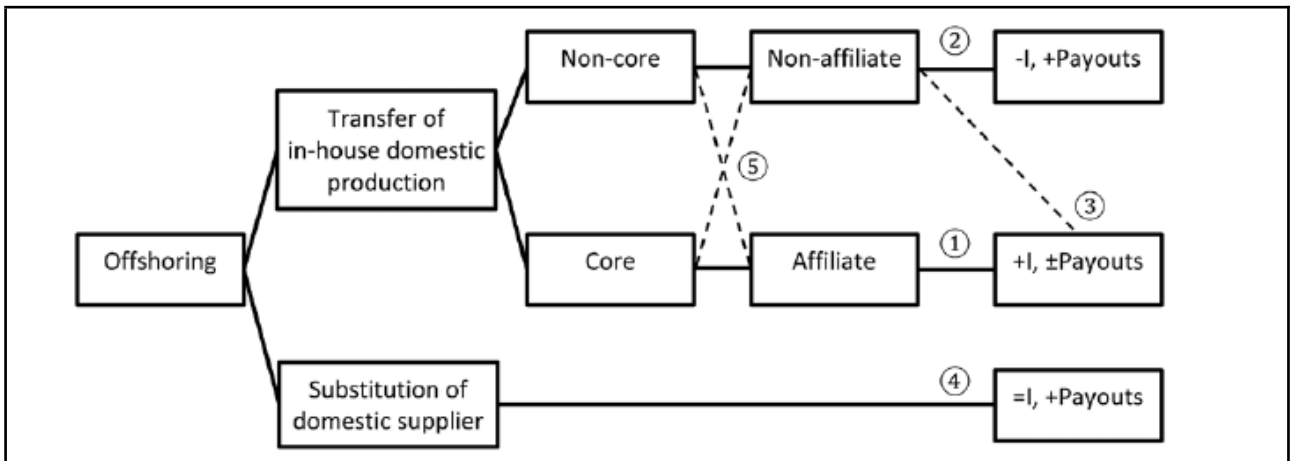
VII. Conclusion

As the goal and onset in this report was to check and evaluate the robustness of their findings from an econometric viewpoint and thus determine if a linkage between financialization and offshoring exists, it has to be mentioned that it was not able in replicating this linkage, but much rather also found opposing results. In doing so, it found some contradicting and verifying results in regards to the theory which Auvray and Rabinovich propose in their paper. If the differences in the results are due to differently applied methods, which were not mentioned in their paper, or due to different dataset cannot be answered at this point. Differences in results could have also arisen due to different transformations or slight differences in the calculation of variables, which can have a detrimental impact as it was pointed out with the calculation of the Tobin's Q. Lastly it also always have to be kept in mind what was pointed out in Chapter 3 and 4. Thus creating an omnifarious result that depicts the objective reality is virtually impossible, as Desrosières has noted: "By multiplying the points of view from different positions, we can always dream of 'encircling reality'. But reality will slip away, for new systems- and languages to make them real- are born every day" (Desrosières 2001, 353).

VIII. Annex

Annex Figure 1:

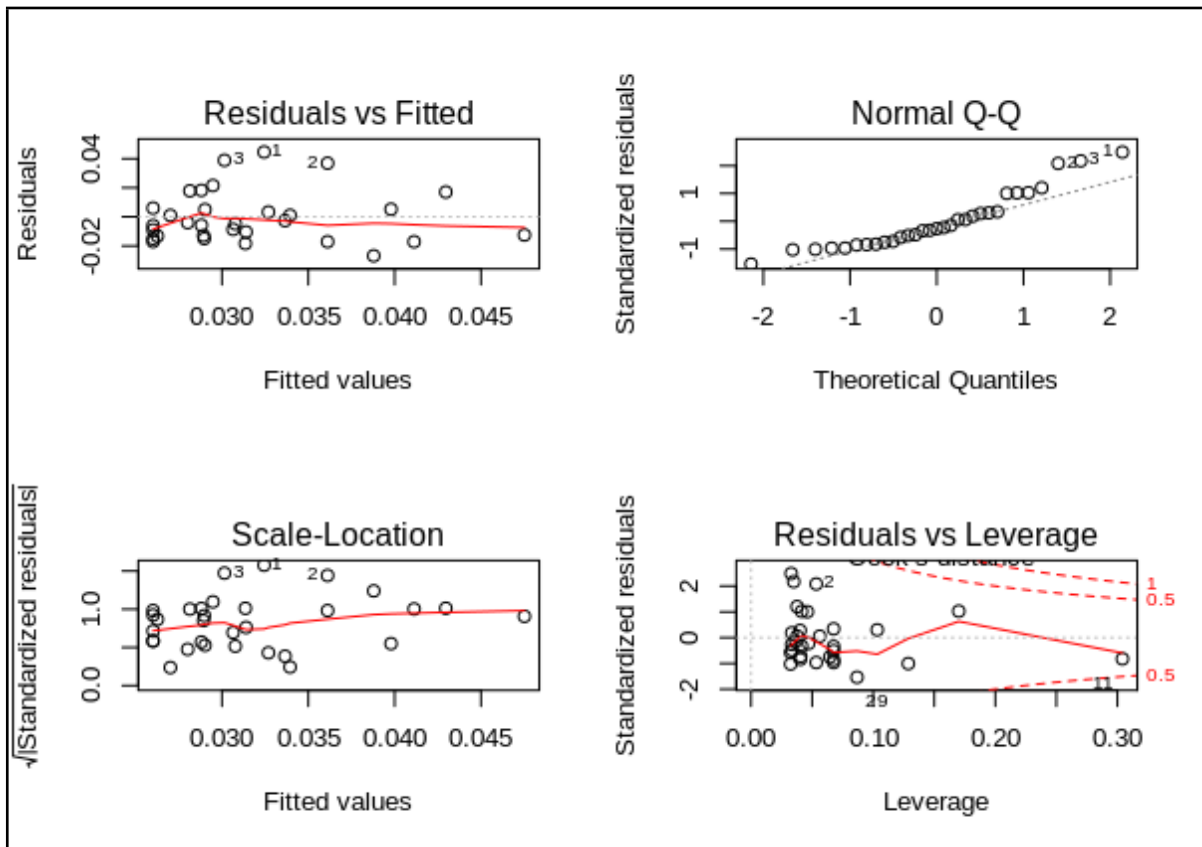
Offshoring, investment and payout from the firm's perspective



Source: Auvray and Rabinovich (2019, 11)

Annex Figure 2:

Residual plots Figure 3



Source: Author's calculations

Annex Table 1
Significance codes

Auvray & Rabinovich	Replication
* ~ Significance at 10%	. ~ Significance at 10 %
** ~ Significance at 5%	* ~ Significance at 5%
*** ~ Significance at 1%	** ~ Significance at 1%
	*** ~ Significance at 0%

Source: Author's Calculations

Annex Table 2 with Tobin's Q calculated with IQ_MARKETCAP_average

Dependent variable: $\ln(I/K)_{i,t}$	Financialisation model			Financialisation and offshoring model			High-non-core non-energy offshoring sectors			Low-non-core non-energy offshoring sectors		
	All	Large	Small	All	Large	Small	All	Large	Small	All	Large	Small
Capital Expenditures	0.246*** (0.022)	0.285*** (0.035)	0.210*** (0.032)	0.250*** (0.022)	0.290*** (0.035)	0.2133*** (0.031)	0.225*** (0.032)	0.237*** (0.045)	0.154*** (0.041)	0.256*** (0.028)	0.320*** (0.046)	0.253*** (0.044)
Profits	0.020*** (0.005)	0.029*** (0.007)	0.018*** (0.005)	0.019*** (0.005)	0.029*** (0.007)	0.018*** (0.005)	0.027*** (0.006)	0.032*** (0.009)	0.029*** (0.007)	0.016* (0.007)	0.036* (0.015)	0.0134. (0.007)
Sales	0.028** (0.009)	-0.020. (0.010)	0.050*** (0.12)	0.027** (0.009)	-0.023* (0.011)	0.049*** (0.012)	0.024* (0.011)	-0.022. (0.013)	0.036** (0.014)	0.027* (0.012)	-0.034* (0.014)	0.052*** (0.014)
Long-term debt	0.009* (0.004)	0.010 (0.008)	0.012* (0.006)	0.009* (0.004)	0.006 (0.008)	0.012* (0.006)	0.007 (0.005)	-0.002 (0.006)	0.010 (0.008)	0.013* (0.006)	0.008 (0.013)	0.008 (0.008)
Tobin's Q	-0.0002 (0.008)	0.014. (0.008)	0.003 (0.011)	-0.001 (0.008)	0.014 (0.008)	0.003 (0.011)	0.003 (0.011)	0.014 (0.012)	0.016 (0.016)	0.009 (0.010)	0.021. (0.011)	0.005 (0.014)
Interest expenditure	-0.003 (0.020)	-0.061. (0.323)	-0.007 (0.020)	-0.002 (0.020)	-0.050 (0.036)	-0.007 (0.020)	0.027 (0.026)	0.024 (0.030)	0.024 (0.024)	-0.029 (0.024)	-0.062 (0.049)	-0.009 (0.025)
Interest and investment income	0.013 (0.048)	0.011 (0.054)	0.020 (0.063)	0.008 (0.048)	0.012 (0.056)	0.019 (0.064)	-0.041 (0.049)	-0.076 (0.048)	0.005 (0.066)	0.013 (0.063)	0.007 (0.084)	-0.041 (0.089)
Dividends	-0.050 (0.037)	-0.045 (0.036)	-0.018 (0.058)	-0.051 (0.037)	-0.047 (0.036)	-0.020 (0.058)	0.011 (0.045)	-0.027 (0.035)	0.161** (0.061)	-0.130** (0.046)	-0.069 (0.052)	-0.151* (0.074)
Stock issue	0.007. (0.004)	0.017** (0.036)	-0.002 (0.005)	0.007. (0.004)	0.018** (0.007)	-0.002 (0.005)	0.006 (0.007)	0.038 (0.010)	-0.011 (0.009)	0.007 (0.005)	0.004 (0.007)	0.003 (0.007)
Stock repurchase	0.003 (0.006)	-0.008 (0.006)	0.015 (0.012)	0.003 (0.006)	-0.008 (0.006)	0.015 (0.012)	0.00005 (0.008)	-0.007 (0.007)	0.003 (0.014)	-0.001 (0.008)	-0.013 (0.009)	0.015 (0.016)
Net debt issue	-0.005* (0.002)	0.0001 (0.003)	-0.006. (0.003)	-0.005* (0.002)	0.0004 (0.003)	-0.006. (0.003)	-0.006. (0.003)	-0.004 (0.003)	-0.007 (0.005)	-0.001 (0.003)	0.005 (0.004)	-0.002 (0.004)
Internal finance	0.041*** (0.006)	0.032*** (0.006)	0.047*** (0.009)	0.041*** (0.006)	0.032*** (0.006)	0.047*** (0.009)	0.034*** (0.007)	0.028*** (0.007)	0.046*** (0.011)	0.039*** (0.008)	0.035*** (0.009)	0.044*** (0.012)
Narrow or core offshoring	-	-	-	0.170 (0.176)	0.391* (0.189)	-0.093 (0.306)	0.121 (0.200)	0.565** (0.212)	-0.373 (0.357)	0.241 (0.373)	-0.323 (0.353)	0.493 (0.571)

Non-core non-energy offshoring	-	-	-	0.141 (0.240)	0.033 (0.238)	0.532 (0.474)	-0.214 (0.300)	-0.314 (0.281)	-0.173 (0.595)	1.488** (0.456)	0.974* (0.449)	2.678** (0.993)
Number of observations	29024	15664	12325	28974	15624	12318	14747	7844	6383	14242	7780	5929
Number of firms	5583	2551	3947	5583	2551	3947	2864	1253	2078	2719	1298	1869

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: Author's calculations

IX. Bibliography

- “About Us | S&P Global Ratings.” n.d. Accessed November 6, 2019.
<https://www.spglobal.com/ratings/en/about/>.
- Arellano, Manuel, and Stephen Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *The Review of Economic Studies* 58 (2): 277–97. <https://doi.org/10.2307/2297968>.
- Auvray, Tristan, and Joel Rabinovich. 2019. “The Financialisation–Offshoring Nexus and the Capital Accumulation of US Non-Financial Firms.” *Cambridge Journal of Economics* 43 (5): 1183–1218. <https://doi.org/10.1093/cje/bey058>.
- Desrosières, A. 2001. “How Real Are Statistics? Four Possible Attitudes.” *Social Research* 68: 339–55.
- Dietzenbacher, Erik, Bart Los, Robert Stehrer, Marcel Timmer, and Gaaitzen de Vries. 2013. “THE CONSTRUCTION OF WORLD INPUT–OUTPUT TABLES IN THE WIOD PROJECT.” *Economic Systems Research* 25 (1): 71–98. <https://doi.org/10.1080/09535314.2012.761180>.
- Epstein, Gerald A., ed. 2005. *Financialization and the World Economy*. Paperback edition 2006, Reprinted 2014. Cheltenham: Elgar.
- Fasianos, Apostolos, Diego Guevara, and Christos Pierros. 2018. “Have We Been Here before? Phases of Financialization within the Twentieth Century in the US.” *Review of Keynesian Economics* 6 (January): 34–61. <https://doi.org/10.4337/roke.2018.01.03>.
- Milberg, William. 2008. “Shifting Sources and Uses of Profits: Sustaining US Financialization with Global Value Chains.” *Economy and Society* 37 (3): 420–51. <https://doi.org/10.1080/03085140802172706>.
- Milberg, William, and Deborah Winkler. 2010. “Financialisation and the Dynamics of Offshoring in the USA.” *Cambridge Journal of Economics* 34 (2): 275–93. <https://doi.org/10.1093/cje/bep061>.
- . 2013. “Outsourcing Economics.” Cambridge Books. Cambridge University Press. <https://econpapers.repec.org/bookchap/cupcbooks/9781107609624.htm>.
- Thévenot, Laurent. 1984. “Rules and Implements: Investment in Forms.” *Information (International Social Science Council)* 23 (1): 1–45. <https://doi.org/10.1177/053901884023001001>.

X. R-Code

Figures 1,2,5 and Residual Plot Figure 5

```
#Initialising-----

library(tidyverse)
library(scales)
library(ggplot2)
library(sqldf)
options(sqldf.driver = "SQLite")
library(cluster.datasets)
library(tibble)

setwd("/home/lukas/Google Drive/Studium/2.Master/Paris/Heterodox Econometrics/R2/Öko2")#change here to your
path
rm(list=ls()) #Remove all the objects in the environment

#Import and clean compustat-----
compustat<- read.csv("./compustat-1.csv", header=TRUE,sep=";",dec = ",")
head(compustat) #Show the headers of Canpustat

fig1.compustat<-as.tibble(compustat) %>% select(-c(X)) %>%
  rename(purchase.stock=prstk,sale.stock=sstk,
         dividends=dv,capital.expenditure=capx) %>%
  mutate(net.payout=purchase.stock+dividends-sale.stock,
         ratio=capital.expenditure/net.payout)
head(fig1.compustat)

ggplot(fig1.compustat,aes(x=year,y=ratio))+geom_point()+ #Displays the figure
  geom_line(color="red",size=2,alpha=0.3)+
  scale_y_continuous(labels = percent) +
  ggtitle("Investment to Net Payout Ratio")

#Import and clean Z1-----
Z1<- read.csv("./Z1.csv", header=F,
```

```

sep=";",dec = ",",skip=5)
# I don't import the first 5 lines with text
# so that R sees that the data is numeric
# but then I have to import again the column names (header)
names(Z1)<-names(read.csv("./Z1.csv", header=T,nrows=0, sep=";"))
head(Z1)

#here I select a set of variables by their column number and rename them
fig1.Z1<-as.tibble(Z1) %>% select(year=1,operating.surplus=9,
                                consumption.of.capital=3,18,equity.liability=67,fixed.capital.formation=29) %>%
  filter(year>1945) %>%
  mutate(net.payouts=dividends.paid-equity.liability,ratio=fixed.capital.formation/net.payouts)
head(fig1.Z1)

#Figure 1 final ----
fig1.offshoring<- read.csv("./Calculation_Offshoring.csv", sep=",") #import offshoring data
view(fig1.offshoring)

ggplot(fig1.Z1,aes(x=year,y=ratio))+geom_line(color="black",size=1,alpha=1)+
  geom_line(data=fig1.compustat,aes(x=year,y=ratio),color="grey",size=1,alpha=0.8)+
  geom_line(data=fig1.offshoring,aes(x=Year,y=Offshoring13*100),linetype="dotted",size=1)+
  scale_y_continuous(labels = percent, name="Investment to Net Payout Ratio", sec.axis = sec_axis(~./100, labels =
percent,name="Offshoring Intensity"))

#Figure 2 final ----

#from Z1 define US.economy=net.payouts/(operating.surplus+consumption.of.capital)

fig2.Z1<-as.tibble(Z1) %>% select(year=1,operating.surplus=9,
                                consumption.of.capital=3,18,equity.liability=67,fixed.capital.formation=29) %>%
  filter(year>1970) %>%
  mutate(net.payouts=dividends.paid-equity.liability,ratio=fixed.capital.formation/net.payouts,US.economy=net.payouts/
(operating.surplus+consumption.of.capital))

#from compustat, define US.listed.firms=net.payouts/income

fig2.compustat<-as.tibble(compustat) %>% select(-c(X)) %>%

```



```

rename(purchase.stock=prstk,sale.stock=sstk, capital.expenditure=capx,
profit=pi,taxes=txt,depreciation=dp,interest=Xint, dividends=dv) %>%
mutate(net.payout=purchase.stock+dividends-sale.stock,
ratio=capital.expenditure/net.payout,
Us.listed.firms=net.payout/(profit-taxes+depreciation+interest),
decade=floor(year/10)) %>%
group_by(decade) %>%
mutate(average=mean(Us.listed.firms)) %>%
ungroup(decade)

```

```

fig2.70s<-as.tibble(fig2.compustat) %>% select(year,average) %>% filter(year<1980)
fig2.80s<-as.tibble(fig2.compustat) %>% select(year,average) %>% filter(year<1990) %>% filter(year>=1980)
fig2.90s<-as.tibble(fig2.compustat) %>% select(year,average) %>% filter(year<2000) %>% filter(year>=1990)
fig2.20s<-as.tibble(fig2.compustat) %>% select(year,average,Us.listed.firms) %>% filter(year>=2000) %>%
mutate(average=mean(Us.listed.firms))

```

#They have no 2010 decade, why not? Maybe because they were tired

#Joels formel is $Us.listed.firms = net.payout / (profit + Interst + depreciation - taxes)$ - Economic explanation and we can also critique it and come to a different result (Justify it)

#At first we tried to use income before taxes, but in years 2000s we had a negative income and thus it skews the plot

```

ggplot(fig2.Z1,aes(x=year,y=US.economy))+
geom_line(color="black",size=1,alpha=1) +
scale_y_continuous(labels = percent) +
geom_line(data=fig2.compustat,aes(x=year,y=Us.listed.firms),color="grey",size=1,alpha=0.8)+
geom_line(data=fig2.70s,aes(x=year,y=average),color="black",linetype="dotted")+
geom_line(data=fig2.80s,aes(x=year,y=average),color="black",linetype="dashed")+
geom_line(data=fig2.90s,aes(x=year,y=average),color="black",linetype="twodash")+
geom_line(data=fig2.20s,aes(x=year,y=average),color="black",linetype="longdash")

```

#Figure 5-----

```
fig5<- read.csv("./fig5.csv", header=TRUE,sep=";",dec = ".") #no tibble needed in this case
```

```
head(fig5)
```

```
#install.packages("gridExtra")
```

```
library(gridExtra)
```

```
fig5.p1<-ggplot(fig5,aes(x= Payout.to.investment.ratio.....p50.,y=Non.core.non.energy.offshoring..mean.))+
geom_point()+
```

```

geom_smooth(method = "lm") #include the SE do have additional information as compared to Tristan

fig5.p2<-ggplot(fig5,aes(x= Payout.to.investment.ratio.....p75.,y=Non.core.non.energy.offshoring..mean.))+
  geom_point()+
  geom_smooth(method = "lm")

grid.arrange(fig5.p1,fig5.p2,ncol=2,nrow=1) #arrange multiple plots side by side
#Maybe we should polit it conversly in order to stay more with their narrative and then look at the results

fig5.res1<-lm(data=fig5,Non.core.non.energy.offshoring..mean.~Payout.to.investment.ratio.....p50.)
fig5.res2<-lm(data=fig5,Non.core.non.energy.offshoring..mean.~Payout.to.investment.ratio.....p75.)
summary(fig5.res1)
summary(fig5.res2)
confint(fig5.res1) #gives you the confidence intervall
confint(fig5.res2)

par(mfrow=c(2,2)) #Plots the LM-Function
plot(fig5.res1)

par(mfrow=c(2,2))
plot(fig5.res2)

#Figure 5 Barplots ----

Barplot<-read.csv("database.csv", header=TRUE,sep=",")

# Converting the SIC codes to industry codes

# The URL for the data.
ff.url <- paste("http://mba.tuck.dartmouth.edu",
               "pages/faculty/ken.french/ftp",
               "Industry_Definitions.zip", sep="/")

# Download the data and unzip it
f <- tempfile()
download.file(ff.url, f)
file.list <- unzip(f,list=TRUE)

```

```

trim <- function(string) {
  # Remove leading and trailing spaces from a string
  ifelse(grepl("^\\s*$", string, perl=TRUE), "",
    gsub("^\\s*(.*?)\\s*$", "\\1", string, perl=TRUE))
}

# Function to do the heavy lifting
extract_ff_ind_data <- function (file) {

  # Read in the data in a plain form
  ff_ind <- as.vector(read.delim(unzip(f, files=file), header=FALSE,
    stringsAsFactors=FALSE))

  # The first 10 characters of each line are the industry data, but only the first
  # row of the data for the SIC codes in an industry are filled in;
  # so fill in the rest.
  ind_num <- trim(substr(ff_ind[,1],1,10))
  for (i in 2:length(ind_num)) {
    if (ind_num[i]== "") ind_num[i] <- ind_num[i-1]
  }

  # The rest of each line is either detail on an industry or details about the
  # range of SIC codes that fit in each industry with a label for each group
  # of SIC codes.
  sic_detail <- trim(substr(ff_ind[,1],11,100))

  # If the line doesn't start with a number, it's an industry description
  is.desc <- grepl("^\\D", sic_detail, perl=TRUE)

  # Pull out information from rows about industries
  regex.ind <- "^((\\d+)\\s+(\\w+)).*$"
  ind_num <- gsub(regex.ind, "\\1", ind_num, perl=TRUE)
  ind_abbrev <- gsub(regex.ind, "\\2", ind_num[is.desc], perl=TRUE)
  ind_list <- data.frame(ind_num=ind_num[is.desc], ind_abbrev,
    ind_desc=sic_detail[is.desc])

  # Pull out information from rows about ranges of SIC codes

```

```

regex.sic <- "^(\\d+)-(\\d+)\\s*(.*)$"
ind_num <- ind_num[!is.desc]
sic_detail <- sic_detail[!is.desc]
sic_low <- as.integer(gsub(regex.sic,"\\1",sic_detail,perl=TRUE))
sic_high <- as.integer(gsub(regex.sic,"\\2",sic_detail,perl=TRUE))
sic_desc <- gsub(regex.sic,"\\3",sic_detail,perl=TRUE)
sic_list <- data.frame(ind_num, sic_low, sic_high, sic_desc)

return(merge(ind_list,sic_list,by="ind_num",all=TRUE))
}

# Extract the data of interest
#ind 10 table
ind_10_table <- extract_ff_ind_data("Siccodes10.txt")

#Step2----
#Create 2 vectors: 1-10 and names of industries
numind <- seq(1, 10, 1)
nameind <- c("Consumer Non-Durables","Consumer Durables", "Manufacturing", "Energy", "Business Equipment",
"Telecommunication", "Wholesale, Retail", "Healthcare", "Utilities", "Other")

#Join these to create dataframe with 2 columns and column names
industries <- cbind(numind, nameind)
colnames(industries) <- c("ind_num", "industry")

#Join this to ind_10_table
ind_10_table <- merge(ind_10_table, industries, by='ind_num')
view(ind_10_table)

#Join these industries to the main dataframe

# First, create intervals for "Others" (they are missing from the data)

table10new <- ind_10_table %>% select(sic_low, sic_high) %>% arrange(sic_low)

table10new <- table10new %>% mutate(sic_low_other = dplyr::lag((table10new$sic_high+1), 1, default=NA)) %>%
mutate(sic_high_other = sic_low-1)

```

```

#Convert low_other value to 0
table10new$sic_low_other[is.na(table10new$sic_low_other)] <- 0
#Convert high_other value to arbitrary large number (e.g. 10000000)
table10new$sic_high_other[is.na(table10new$sic_high_other)] <- 100000

#Take only right two columns
otherval <- table10new[, (3:4)]

#Add column with name: "Other -- Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance" and
reorder and rename columns
otherval$industry <- c(rep("Other", 59))
otherval <- otherval[, c(3, 1, 2)]
colnames(otherval) <- c("industry", "sic_low", "sic_high")
view(otherval)

#Change ind_10_table to be similar to this
ind_10_table1 <- ind_10_table %>% select(industry, sic_low, sic_high)

#Add otherval to bottom of ind_10_table
ind_10_table2 <- rbind(ind_10_table1, otherval)
view(ind_10_table2)

#Use the ind_10_table2 to classify the companys into sectors

df.temp <- sqldf("SELECT * from Barplot AS a LEFT JOIN ind_10_table2 AS b
                ON (a.PRIMARY_SIC_CODE BETWEEN b.sic_low AND b.sic_high)")

view(df.temp)

#BARPLOT NCNE----

fig5.Barplot <- as.tibble(df.temp) %>% select(year, industry, ncne) %>%
  group_by(year < 2003)

view(fig5.Barplot)

```

```
agg_mean<-aggregate(fig5.Barplot[3],by=list(fig5.Barplot$industry,fig5.Barplot$`year < 2003`),FUN=mean,  
na.rm=TRUE)
```

```
view(agg_mean)
```

```
agg_mean$colour <- ifelse(agg_mean$Group.2 == "FALSE", "orange", NA)
```

```
ggplot(agg_mean, aes(x=reorder(Group.1, -ncne),y=ncne,fill=colour))+geom_col(position="dodge2")
```

Table 6 and 7

```
#Initialisierung----

library(tidyverse)
library(scales)
library(ggplot2)
library(sqldf)
options(sqldf.driver = "SQLite")
library(cluster.datasets)
library(tibble)
library(plm)
library(plyr)
library(DescTools)

# Reduced Database -----

Table6.reduced<-read.csv('reduced2.csv', header=TRUE)
Table6.reduced <- Table6.reduced %>% arrange(Company_Name, year)
view(Table6.reduced)
#get column names in order to work quicker
colnames(Table6.reduced)

#run the plm model for large firms without offshoring

Table6.reduced.fe<-plm(w_ln_capex0~w_ln_capex0_l+w_ln_profits_l+ w_ln_sales_l+w_ln_tobinq_l+w_ln_longdebt+
      w_ln_interestexp_l+w_ln_interestinc0_l+ w_ln_div0_l+ln_stockissue0_l+ln_stockrep0_l+
      w_ln_netdebtissue0_l+w_ln_cash_l, data=Table6.reduced, effect="twoways", model="within",
      index=c("Company_Name", "year"))

# effect="twoways" creates dummies for both individuals and time
# effect="individual" or effect="time" instead creates dummies only for individuals or for time

summary(Table6.reduced.fe)

#Full Database ----
```

```

Table6 <- read.csv('database.csv', header=TRUE, dec = ".")

#rearrange data in years descending and by company name, as arranging with company name is impossible creating a subset
Table6<-Table6 %>% group_by(Table6$Company_Name) %>% arrange(Company_Name,desc(Company_Name),year)
%>%ungroup(Table6$Company_Name)

#Clean up the database / 2nd interpretation and realizing it is all done----
#From their report: "We also removed firms with no information for all years of capital expenditure,
#sales, net property plan and equipment, long-term debt, interest expenses,
#cash and short-term securities, total assets, total liabilities and equities

#Remove all firms with no information on the abovementioned, first check
#Clean-up LTDEBT
summary(Table6$IQ_LT_DEBT)# 2294 NAs in Long-term debt
Table6 <- mutate(Table6,clean=(ifelse(!is.na(Table6$IQ_LT_DEBT), "1", "0"))) # create a binary code based on the NAs
Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
as.table(Table6.count)
names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.count, by='Var1')
Table6<- mutate(Table6,clean=as.numeric(Table6$clean))
Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.clean, by='Var1')
Table6<-Table6 [!(Table6$x = 0),] #if clean equals 0 remove company
names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for clarification
rm(Table6.clean)
rm(Table6.count)
Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#Clean up Interest Expenses
summary(Table6$IQ_INTEREST_EXP) # 172 NAs in Interest Expenses
Table6 <- mutate(Table6,clean=(ifelse(!is.na(Table6$IQ_INTEREST_EXP), "1", "0"))) # create a binary code based on the NAs
Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
as.table(Table6.count)
names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.count, by='Var1')

```



```

Table6<- mutate(Table6,clean=as.numeric(Table6$clean))
Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.clean, by='Var1')
Table6<-Table6 [!(Table6$x = 0),] #if clean equals 0 remove company
names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for merging
rm(Table6.clean)
rm(Table6.count)
Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#Sales (Thus Revenue)
summary(Table6$IQ_REV)
Table6 <- mutate(Table6,clean=(ifelse(!is.na(Table6$IQ_REV), "1", "0"))) # create a binary code based on the NAs
Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
as.table(Table6.count)
names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.count, by='Var1')
Table6<- mutate(Table6,clean=as.numeric(Table6$clean))
Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.clean, by='Var1')
Table6<-Table6 [!(Table6$x = 0),] #if clean equals 0 remove company
names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for merging
rm(Table6.clean)
rm(Table6.count)
Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#NPPE
summary(Table6$IQ_NPPE) # no NAs

#Cash and short-term securities / investment
#Why surpress the firm when the variable is not being used?
summary(Table6$IQ_CASH_ST_INVEST)
Table6 <- mutate(Table6,clean=(ifelse(!is.na(Table6$IQ_CASH_ST_INVEST), "1", "0"))) # create a binary code based
on the NAs
Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
as.table(Table6.count)
names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging

```

```

Table6 <- merge(Table6, Table6.count, by='Var1')
Table6<- mutate(Table6,clean=as.numeric(Table6$clean))
Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.clean, by='Var1')
Table6<-Table6 [!(Table6$x = 0),] #if clean equals 0 remove company
names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for merging
rm(Table6.clean)
rm(Table6.count)
Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#Total Assets
summary(Table6$IQ_TOTAL_ASSETS)# No NAs

#Total Liabilities
summary(Table6$IQ_TOTAL_LIAB) # No NAs
#Equities
summary(Table6$IQ_TOTAL_EQUITY)# No NAs

#We also removed observations with no information on market cpaitalisation at the end of the year / done by Hugo
# with duplicate observations, negative values for interest income and positive values for interest expenses and
dividends
count(Table6$IQ_INTEREST_INCOME<0)
count(Table6$IQ_TOTAL_DIV_PAID_CF>0)
count(Table6$IQ_INTEREST_EXP>0)

#Their Final sample 2049 companies on average per year
#1995-2011 timeframe

count(Table6$year==1995)#1732
count(Table6$year==1996)#2275
count(Table6$year==1997)#2642
count(Table6$year==1998)#2708
count(Table6$year==1999)#2720
count(Table6$year==2000)#2753
count(Table6$year==2001)#2702
count(Table6$year==2002)#2540
count(Table6$year==2003)#2448

```

```

count(Table6$year==2004)#2407
count(Table6$year==2005)#2396
count(Table6$year==2006)#2306
count(Table6$year==2007)#2200
count(Table6$year==2008)#2128
count(Table6$year==2009)#1980
count(Table6$year==2010)#1851
count(Table6$year==2011)#1762

```

```
A<-c(1732,2275,2642,2708,2720,2753,2702,2540,2448,2407,2396,2306,2200,2128,1980,1851,1762)
```

```
B<-c(1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011)
```

```
Table6.apy <- data.frame(A,B)
```

```
mean(Table6.apy$A) #2326.471 firms yearly on average, which differs from their assumption of 2049 companies
average per year +~300 companies
```

```
rm(A)
```

```
rm(B)
```

```
rm(Table6.apy)
```

```
#Based on the Formula, create the important variables to log and lag them in the next step and winsorize (Both Versions
to calculate TQ)----
```

```
#It is imporant to treat the NAs for variables where it is logical as 0, because otherwise we will recieve more than
39000NAs for STKISSUE
```

```
temp.df <- data.frame(Table6) # copy the original
```

```
temp.df$IQ_PREF_ISSUED[is.na(temp.df$IQ_PREF_ISSUED)] <- 0 # logically we can assume that the NA represents
a 0
```

```
temp.df$IQ_PREF_REP[is.na(temp.df$IQ_PREF_REP)] <-0 #logically we can assume that NA represents 0
```

```
temp.df$IQ_COMMON_ISSUED[is.na(temp.df$IQ_COMMON_ISSUED)] <-0 #logically we can assume that NA
represents 0
```

```
temp.df$IQ_COMMON_REP[is.na(temp.df$IQ_COMMON_REP)] <-0 #logically we can assume that NA represents 0
```

```
temp.df$IQ_TOTAL_DIV_PAID_CF[is.na(temp.df$IQ_TOTAL_DIV_PAID_CF)] <-0 #logically we can assume that
NA represents 0
```

```
temp.df$IQ_TOTAL_DEBT_REPAID[is.na(temp.df$IQ_TOTAL_DEBT_REPAID)] <-0 #logically we can assume that
NA represents 0
```

```
temp.df$IQ_TOTAL_DEBT_ISSUED[is.na(temp.df$IQ_TOTAL_DEBT_ISSUED)] <-0 #logically we can assume that
NA represents 0
```

```
temp.df$IQ_INTEREST_INCOME[is.na(temp.df$IQ_INTEREST_INCOME)] <-0 #logically we can assume that NA
represents 0
```

```
#Capital expenditures
```

```
Table6$CapitalExpenditures <- ((temp.df$IQ_CAPEX*-1) / temp.df$IQ_NPPE) # or any other calculation
```

```

Table6$CapitalExpenditures <-
Winsorize(Table6$CapitalExpenditures,minval=quantile(Table6$CapitalExpenditures, .005,na.rm=TRUE),maxval=qua
ntile(Table6$CapitalExpenditures, .995,na.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$CapitalExpenditures)

count(Table6,c("CapitalExpenditures")) #counting

#Profits
Table6$Profits <- ((temp.df$IQ_OPER_INC) / temp.df$IQ_NPPE)

Table6$Profits <-
Winsorize(Table6$Profits,minval=quantile(Table6$Profits, .005,na.rm=TRUE),maxval=quantile(Table6$Profits, .995,n
a.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$Profits)

count(Table6,c("Profits"))

#Sales
Table6$Sales <- ((temp.df$IQ_REV) / temp.df$IQ_NPPE)

Table6$Sales <-
Winsorize(Table6$Sales,minval=quantile(Table6$Sales, .005,na.rm=TRUE),maxval=quantile(Table6$Sales, .995,na.rm
=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$Sales)

count(Table6,c("Sales"))

#Long-term debt
Table6$LTDEBT <- ((temp.df$IQ_LT_DEBT) / temp.df$IQ_NPPE)

Table6$LTDEBT <-
Winsorize(Table6$LTDEBT,minval=quantile(Table6$LTDEBT, .005,na.rm=TRUE),maxval=quantile(Table6$LTDEBT,
.995,na.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$LTDEBT)

count(Table6,c("LTDEBT"))

#Interest expenditure
Table6$INTEXP <- ((temp.df$IQ_INTEREST_EXP*-1) / temp.df$IQ_NPPE)

Table6$INTEXP <-
Winsorize(Table6$INTEXP,minval=quantile(Table6$INTEXP, .005,na.rm=TRUE),maxval=quantile(Table6$INTEXP, .
995,na.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$INTEXP)

count(Table6,c("INTEXP"))

#Interest and investment income
Table6$INTINC <- ((temp.df$IQ_INTEREST_INCOME) / temp.df$IQ_NPPE)

```

```

Table6$INTINC <-
Winsorize(Table6$INTINC,minval=quantile(Table6$INTINC, .005,na.rm=TRUE),maxval=quantile(Table6$INTINC, .
995,na.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$INTINC)
count(Table6,c("INTINC"))

#Dividends
Table6$DIV <- ((temp.df$IQ_TOTAL_DIV_PAID_CF*-1) / temp.df$IQ_NPPE) # or any other calculation
Table6$DIV <-
Winsorize(Table6$DIV,minval=quantile(Table6$DIV, .005,na.rm=TRUE),maxval=quantile(Table6$DIV, .995,na.rm=T
RUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$DIV)
count(Table6,c("DIV"))

#Stock issue
Table6$STKISSUE <- ((temp.df$IQ_COMMON_ISSUED + temp.df$IQ_PREF_ISSUED) / temp.df$IQ_NPPE)
Table6$STKISSUE <-
Winsorize(Table6$STKISSUE,minval=quantile(Table6$STKISSUE, .005,na.rm=TRUE),maxval=quantile(Table6$STK
ISSUE, .995,na.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$STKISSUE)
count(Table6,c("STKISSUE"))

#Stock repurchases
Table6$STKREP <- (((temp.df$IQ_COMMON_REP*-1)+(temp.df$IQ_PREF_REP*-1)) / temp.df$IQ_NPPE)
Table6$STKREP <-
Winsorize(Table6$STKREP,minval=quantile(Table6$STKREP, .005,na.rm=TRUE),maxval=quantile(Table6$STKREP,
.995,na.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$STKREP)
count(Table6,c("STKREP"))

#Net debt issue
Table6$NDEBTISSUE <- (((temp.df$IQ_TOTAL_DEBT_ISSUED)+(temp.df$IQ_TOTAL_DEBT_REPAID)) /
temp.df$IQ_NPPE)
Table6$NDEBTISSUE <-
Winsorize(Table6$NDEBTISSUE,minval=quantile(Table6$NDEBTISSUE, .005,na.rm=TRUE),maxval=quantile(Table
6$NDEBTISSUE, .995,na.rm=TRUE),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$NDEBTISSUE)
count(Table6,c("NDEBTISSUE"))

#Internal finance
Table6$IF <- ((temp.df$IQ_CASH_ST_INVEST) / temp.df$IQ_NPPE)

```

```

Table6$IF <-
Winsorize(Table6$IF,minval=quantile(Table6$IF, .005,na.rm=TRUE),maxval=quantile(Table6$IF, .995,na.rm=TRUE),
probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$IF)

count(Table6,c("IF"))

```

#Tobin's q RUN FIRST THIS

```

Table6$TQ <-
(((temp.df$IQ_MARKETCAP_end_of_year+temp.df$IQ_TOTAL_LIAB)/temp.df$IQ_TOTAL_ASSETS)) #Use the
end of year, as the authors used the same / makes large difference

```

```

Table6$TQ <-
Winsorize(Table6$TQ,minval=quantile(Table6$TQ, .005,na.rm=TRUE),maxval=quantile(Table6$TQ, .995,na.rm=TRU
E),probs=c(0.005,0.995),na.rm=TRUE)

summary(Table6$TQ)

count(Table6,c("TQ"))

```

#Tobin's q with the average market cap THEN THIS

```

#Table6$TQ <- (((temp.df$IQ_MARKETCAP_average+temp.df$IQ_TOTAL_LIAB)/temp.df$IQ_TOTAL_ASSETS))
#Use the end of year, as the authors used the same / makes large difference

```

```

#Table6$TQ <-
Winsorize(Table6$TQ,minval=quantile(Table6$TQ, .005,na.rm=TRUE),maxval=quantile(Table6$TQ, .995,na.rm=TRU
E),probs=c(0.005,0.995),na.rm=TRUE)

#summary(Table6$TQ)

#count(Table6,c("TQ"))

```

#Narrow or core offshoring

```

summary(Table6$noc)

count(Table6,c("noc")) # Assume it is already winsorized

```

#Non-core non-energy offshoring

```

summary(Table6$ncne)

count(Table6,c("ncne"))

remove('temp.df') # remove the temporary db

```

#log transformation----

#p. 18 footnote 8 every variable that can be positive and negative

#Important: Our log transformation avoids censorship of firms with variables equal or inferior to zero (those with

#negative earnings or without stock issues or financial payouts for example): for any variable var, we compute

$\ln(\text{var}) = -\ln(\text{var} + 1)$ if $\text{var} \leq 0$, and $\ln(\text{var}) = \ln(\text{var} + 1)$ if $\text{var} > 0$. - Mistake in their formula on p.18

```
#Capital expenditures
```

```
Table6 <- mutate(Table6, ln_CapitalExpenditures = ifelse((CapitalExpenditures>0), log(CapitalExpenditures+1),  
(log(abs(CapitalExpenditures)+1)*-1)))
```

```
#Profits
```

```
Table6 <- mutate(Table6, ln_Profits = ifelse((Profits>0), log(Profits+1),((log(abs(Profits)+1)*-1))))
```

```
#Creates NaNs/ Transform
```

```
Table6$ln_Profits[is.nan(Table6$ln_Profits)] <- NA
```

```
#Sales
```

```
Table6 <- mutate(Table6, ln_Sales = ifelse((Sales>0), log(Sales+1),(log(abs(Sales)+1)*-1)))
```

```
#Long-term debt
```

```
Table6 <- mutate(Table6, ln_LTDEBT = ifelse((LTDEBT>0), log(LTDEBT+1),(log(abs(LTDEBT)+1)*-1)))
```

```
#Interest expenditure
```

```
Table6 <- mutate(Table6, ln_INTEXP = ifelse((INTEXP>0), log(INTEXP+1),(log(abs(INTEXP)+1)*-1)))
```

```
#Interest and investment income
```

```
Table6 <- mutate(Table6, ln_INTINC = ifelse((INTINC>0), log(INTINC+1),(log(abs(INTINC)+1)*-1)))
```

```
#Dividends
```

```
Table6 <- mutate(Table6, ln_DIV = ifelse((DIV>0), log(DIV+1),(log(abs(DIV)+1)*-1)))
```

```
#Stock issue
```

```
Table6 <- mutate(Table6, ln_STKISSUE = ifelse((STKISSUE>0), log(STKISSUE+1),(log(abs(STKISSUE)+1)*-1)))
```

```
#Stock repurchases
```

```
Table6 <- mutate(Table6, ln_STKREP = ifelse((STKREP>0), log(STKREP+1),(log(abs(STKREP)+1)*-1)))
```

```
#Net debt issue
```

```
Table6 <- mutate(Table6, ln_NDEBTISSUE = ifelse((NDEBTISSUE>0), log(NDEBTISSUE+1),  
(log(abs(NDEBTISSUE)+1)*-1)))
```

```
summary(Table6$NDEBTISSUE)#NANS-Values transform them to NA
```

```
Table6$ln_NDEBTISSUE[is.nan(Table6$ln_NDEBTISSUE)] <- NA
```

```
#Internal finance
```

```

Table6 <- mutate(Table6, ln_IF = ifelse((IF>0), log(IF+1),(log(abs(IF)+1)*-1)))

#Tobin's q
Table6 <- mutate(Table6, ln_TQ = ifelse((TQ>0), log(TQ+1),(log(abs(TQ)+1)*-1)))

#Narrow or core offshoring
Table6 <- mutate(Table6, ln_NOC = ifelse((noc>0), log(noc+1),(log(abs(noc)+1)*-1)))

#Non-core non-energy offshoring
Table6 <- mutate(Table6, ln_NCNE = ifelse((ncne>0), log(ncne+1),(log(abs(ncne)+1)*-1)))

#lag the database---- This is not necessary and it takes a long time, as the calculation splits it into subsamples.
The results are lagged later on in the GMM function

#use ddply to split into subdataset, because we have to lag for every observation/firm individually

#Capital expenditures
Table6 <- ddply(Table6, .(Company_Name), transform, ln_CapitalExpenditures_l=c( NA, ln_CapitalExpenditures[-
length(ln_CapitalExpenditures)] ))

#Profits
Table6 <- ddply(Table6, .(Company_Name), transform, ln_Profits_l=c( NA, ln_Profits[-length(ln_Profits)] ))

#Sales
Table6 <- ddply(Table6, .(Company_Name), transform, ln_Sales_l=c( NA, ln_Sales[-length(ln_Sales)] ))

#Long-term debt
#Not lagged

#Interest expenditure
Table6 <- ddply(Table6, .(Company_Name), transform, ln_INTEXP_l=c( NA, ln_INTEXP[-length(ln_INTEXP)] ))

#Interest and investment income
Table6 <- ddply(Table6, .(Company_Name), transform, ln_INTINC_l=c( NA, ln_INTINC[-length(ln_INTINC)] ))

#Dividends
Table6 <- ddply(Table6, .(Company_Name), transform, ln_DIV_l=c( NA, ln_DIV[-length(ln_DIV)] ))

#Stock issue
Table6 <- ddply(Table6, .(Company_Name), transform, ln_STKISSUE_l=c( NA, ln_STKISSUE[-
length(ln_STKISSUE)] ))

```



```
#Stock repurchases
```

```
Table6 <- ddply(Table6, .(Company_Name), transform, ln_STKREP_l=c( NA, ln_STKREP[-length(ln_STKREP)] ))
```

```
#Net debt issue
```

```
Table6 <- ddply(Table6, .(Company_Name), transform, ln_NDEBTISSUE_l=c( NA, ln_NDEBTISSUE[-length(ln_NDEBTISSUE)] ))
```

```
#Internal finance
```

```
Table6 <- ddply(Table6, .(Company_Name), transform, ln_IF_l=c( NA, ln_IF[-length(ln_IF)] ))
```

```
#Tobin's q
```

```
Table6 <- ddply(Table6, .(Company_Name), transform, ln_TQ_l=c( NA, ln_TQ[-length(ln_TQ)] ))
```

```
#Narrow or core offshoring
```

```
Table6 <- ddply(Table6, .(Company_Name), transform, ln_NOC_l=c( NA, ln_NOC[-length(ln_NOC)] ))
```

```
#NCNE
```

```
Table6 <- ddply(Table6, .(Company_Name), transform, ln_NCNE_l=c( NA, ln_NCNE[-length(ln_NCNE)] ))
```

```
#PGMM Arellano-Bond tw-step difference----
```

```
Table6.pgmm.prelagged<- pgmm(ln_CapitalExpenditures ~ ln_CapitalExpenditures_l + ln_Profits_l + ln_Sales_l +  
ln_TQ_l+ln_LTDEBT+ln_INTEXP_l+ln_INTINC_l+ln_DIV_l+ln_STKISSUE_l+ln_STKREP_l+  
ln_NDEBTISSUE_l+ln_IF_l | dplyr::lag(ln_CapitalExpenditures, 2:99), data=Table6, model="twosteps",  
effect="twoways", index=c("Company_Name", "year"))
```

```
summary(Tbale6.pgmm.prelagged)
```

```
#Seeing that the pgmm was pre-lagged variables yields wrong results, lagging the variables in the pgmm function
```

```
Table6.pgmm<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+  
dplyr::lag(ln_Sales) +  
ln_LTDEBT +dplyr::lag(ln_TQ) + dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+  
dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) + dplyr::lag(ln_NDEBTISSUE)  
+  
dplyr::lag(ln_IF)| dplyr::lag(ln_CapitalExpenditures, 2:99), data=Table6, model="twosteps",  
effect="twoways", index=c("Company_Name", "year"), transformation = 'd')
```

```
summary(Table6.pgmm)
```

```

#Split dataframe into small and large companies, where they can change yearly in the categories V2----
# create list of company names (alphabetical)
Table6.year <- sort(unique(Table6$year))

#create the average yearly median
Table6.yearly.median<-aggregate(Table6$IQ_TOTAL_ASSETS,
by=list(year=Table6$year),FUN=median,na.rm=TRUE, na.action=NULL,c("median"))
names(Table6.yearly.median)[names(Table6.yearly.median) == "x"] <- "yearly.median" #renaming / taking the
individual median over

#Add median values to main dataframe

Table6 <- merge(Table6, Table6.yearly.median, by.x='year', by.y='year')

#Split dataframes after whether their assets are higher or lower than the median

Table6.small.f <- Table6 %>% filter(IQ_TOTAL_ASSETS < yearly.median )
Table6.large.f <- Table6 %>% filter(IQ_TOTAL_ASSETS >= yearly.median )
rm(Table6.year)
rm(Table6.yearly.median)

#Run GMM for small and large firms----

Table6.pgmm.large<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+
dplyr::lag(ln_Sales) +
ln_LTDEBT + dplyr::lag(ln_TQ) + dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) + dplyr::lag(ln_NDEBTISSUE)
+
dplyr::lag(ln_IF)| dplyr::lag(ln_CapitalExpenditures, 2:99), data=Table6.large.f, model="twosteps",
effect="twoways", index=c("Company_Name", "year"), transformation = 'd')

summary(Table6.pgmm.large)

Table6.pgmm.small<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+
dplyr::lag(ln_Sales) +
ln_LTDEBT + dplyr::lag(ln_TQ) + dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+

```

```

      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)| dplyr::lag(ln_CapitalExpenditures, 2:99), data=Table6.small.f, model="twosteps",
effect="twoways", index=c("Company_Name", "year"), transformation = 'd')

```

```
summary(Table6.pgmm.small)
```

```
#Including Financialisation and Offshoring-Model----
```

```

Table6.pgmm.off<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+
dplyr::lag(ln_Sales) +
      ln_LTDEBT + dplyr::lag(ln_TQ) +dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures,
2:99), data=Table6, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation = 'd')

```

```
summary(Table6.pgmm.off)
```

```
#Large
```

```

Table6.pgmm.large.off<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+
dplyr::lag(ln_Sales) +
      ln_LTDEBT + dplyr::lag(ln_TQ) + dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures,
2:99), data=Table6.large.f, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation =
'd')

```

```
summary(Table6.pgmm.large.off)
```

```
#Small
```

```

Table6.pgmm.small.off<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+
dplyr::lag(ln_Sales) +
      ln_LTDEBT + dplyr::lag(ln_TQ) + dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures,
2:99), data=Table6.small.f, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation
= 'd')

```

```
summary(Table6.pgmm.small.off)
```

```

#Split into High and low- non-core non-energy Offshoring----
#High- and low-offshoring sectors are those belonging upper and lower median of non-core non-energy offshoring / use
the sector splitting code

# Converting the SIC codes to industry codes

# The URL for the data.
ff.url <- paste("http://mba.tuck.dartmouth.edu",
               "pages/faculty/ken.french/ftp",
               "Industry_Definitions.zip", sep="/")

# Download the data and unzip it
f <- tempfile()
download.file(ff.url, f)
file.list <- unzip(f,list=TRUE)

trim <- function(string) {
  # Remove leading and trailing spaces from a string
  ifelse(grepl("^\\s*$", string, perl=TRUE), "",
        gsub("^\\s*(.*?)\\s*$", "\\1", string, perl=TRUE))
}

# Function to do the heavy lifting
extract_ff_ind_data <- function (file) {

  # Read in the data in a plain form
  ff_ind <- as.vector(read.delim(unzip(f, files=file), header=FALSE,
                                stringsAsFactors=FALSE))

  # The first 10 characters of each line are the industry data, but only the first
  # row of the data for the SIC codes in an industry are filled in;
  # so fill in the rest.
  ind_num <- trim(substr(ff_ind[,1],1,10))
  for (i in 2:length(ind_num)) {
    if (ind_num[i]== "") ind_num[i] <- ind_num[i-1]
  }
}

```

```

# The rest of each line is either detail on an industry or details about the
# range of SIC codes that fit in each industry with a label for each group
# of SIC codes.
sic_detail <- trim(substr(ff_ind[,1],11,100))

# If the line doesn't start with a number, it's an industry description
is.desc <- grepl("^\\D",sic_detail,perl=TRUE)

# Pull out information from rows about industries
regex.ind <- "^((\\d+)\\s+(\\w+).*$)"
ind_num <- gsub(regex.ind,"\\1",ind_num,perl=TRUE)
ind_abbrev <- gsub(regex.ind,"\\2",ind_num[is.desc],perl=TRUE)
ind_list <- data.frame(ind_num=ind_num[is.desc],ind_abbrev,
                      ind_desc=sic_detail[is.desc])

# Pull out information rows about ranges of SIC codes
regex.sic <- "^((\\d+)-(\\d+)\\s*(.*)$)"
ind_num <- ind_num[!is.desc]
sic_detail <- sic_detail[!is.desc]
sic_low <- as.integer(gsub(regex.sic,"\\1",sic_detail,perl=TRUE))
sic_high <- as.integer(gsub(regex.sic,"\\2",sic_detail,perl=TRUE))
sic_desc <- gsub(regex.sic,"\\3",sic_detail,perl=TRUE)
sic_list <- data.frame(ind_num, sic_low, sic_high, sic_desc)

return(merge(ind_list,sic_list,by="ind_num",all=TRUE))
}

# Extract the data of interest
#ind 10 table
ind_10_table <- extract_ff_ind_data("Siccodes10.txt")

#Create 2 vectors: 1-10 and names of industries
numind <- seq(1, 10, 1)
nameind <- c("Consumer Non-Durables","Consumer Durables", "Manufacturing", "Energy", "Business Equipment",
"Telecommunication", "Wholesale, Retail", "Healthcare", "Utilities", "Other")

#Join these to create dataframe with 2 columns and column names
industries <- cbind(numind, nameind)

```

```

colnames(industries) <- c("ind_num", "industry")

#Join this to ind_10_table
ind_10_table <- merge(ind_10_table, industries, by='ind_num')
view(ind_10_table)

#Join these industries to the main dataframe

# First, create intervals for "Others" (they are missing from the data)

table10new <- ind_10_table %>% select(sic_low, sic_high) %>% arrange(sic_low)

table10new <- table10new %>% mutate(sic_low_other = dplyr::lag((table10new$sic_high+1), 1, default=NA)) %>%
  mutate(sic_high_other = sic_low-1)
#Convert low_other value to 0
table10new$sic_low_other[is.na(table10new$sic_low_other)] <- 0
#Convert high_other value to arbitrary large number (e.g. 10000000)
table10new$sic_high_other[is.na(table10new$sic_high_other)] <- 100000

#Take only right two columns
othervals <- table10new[, (3:4)]

#Add column with name: "Other -- Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance" and
reorder and rename columns
othervals$industry <- c(rep("Other", 59))
othervals <- othervals[, c(3,1,2)]
colnames(othervals) <- c("industry", "sic_low", "sic_high")
view(othervals)

#Change ind_10_table to be similar to this
ind_10_table1 <- ind_10_table %>% select(industry, sic_low, sic_high)

#Add othervals to bottom of ind_10_table
ind_10_table2 <- rbind(ind_10_table1, othervals)
view(ind_10_table2)
rm(file.list)
rm(table10new)

```

```

rm(otherval)
rm(ind_10_table)
rm(ind_10_table1)
rm(industries)
#Use the ind_10_table2 to classify the companys into sectors

df.temp <- sqldf("SELECT * from Table6 AS a LEFT JOIN ind_10_table2 AS b
                ON (a.PRIMARY_SIC_CODE BETWEEN b.sic_low AND b.sic_high)")

Table6.median.off<-aggregate(df.temp$ncne, by=list(industry=df.temp$industry),FUN=median,na.rm=TRUE,
na.action=NULL,c("median"))

names(Table6.median.off)[names(Table6.median.off) == "x"] <- "industry.median" #renaming / taking the individual
median over

#Split and in order to do so find the median along which we will do so

Table6.median <-mutate(df.temp, median=median(df.temp$ncne,na.rm=TRUE,na.action=NULL),c("median")) %>%
select(median) #median for the whole dataframe

Table6.median <- Table6.median [1:10,1]

#dividing the industries

Table6.high <- Table6.median.off %>% filter(Table6.median.off$industry.median>=Table6.median) %>%
select(industry)

Table6.low <- Table6.median.off %>% filter(Table6.median.off$industry.median<Table6.median) %>% select(industry)

#Assign a binary code

Table6.high <- mutate(Table6.high, size=1)
Table6.low <- mutate(Table6.low, size=0)

#Joining the coded data

Table6.high.low <- rbind(Table6.high, Table6.low)
rm(Table6.high)
rm(Table6.low)

#Join the binarycode with the dataset

```

```
df.temp <- merge(df.temp, Table6.high.low, by='industry')
```

```
#filter
```

```
Table6.high.off <- df.temp %>% filter (df.temp$size>0)
```

```
Table6.low.off<- df.temp %>% filter (df.temp$size==0)
```

```
# Now for large and small assuming the sectors stay the same
```

```
Table6smallf<-Table6.small.f # Take out the . as SQL does not work with them
```

```
df.temp.small <- sqldf("SELECT * from Table6smallf AS a LEFT JOIN ind_10_table2 AS b  
ON (a.PRIMARY_SIC_CODE BETWEEN b.sic_low AND b.sic_high)")
```

```
rm(Table6smallf)
```

```
Table6largef <-Table6.large.f
```

```
df.temp.large <- sqldf("SELECT * from Table6largef AS a LEFT JOIN ind_10_table2 AS b  
ON (a.PRIMARY_SIC_CODE BETWEEN b.sic_low AND b.sic_high)")
```

```
rm(Table6largef)
```

```
df.temp.small <- merge(df.temp.small, Table6.high.low, by='industry')
```

```
df.temp.large <- merge(df.temp.large, Table6.high.low, by='industry')
```

```
#filter
```

```
Table6.high.off.small <- df.temp.small %>% filter (df.temp.small$size>0)
```

```
Table6.high.off.large <- df.temp.large %>% filter (df.temp.large$size>0)
```

```
Table6.low.off.small<- df.temp.small %>% filter (df.temp.small$size==0)
```

```
Table6.low.off.large<- df.temp.large %>% filter (df.temp.large$size==0)
```

```
#High-non-core non-energy Offshoring----
```

```
Table6.pgmm.high.off<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+  
dplyr::lag(ln_Sales) +
```

```
ln_LTDEBT + dplyr::lag(ln_TQ) +dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
```



```

      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures, 2:99),
data=Table6.high.off, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation = 'd')

```

```
summary(Table6.pgmm.high.off)
```

```
Table6.pgmm.low.off<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) + dplyr::lag(ln_Profits)+
dplyr::lag(ln_Sales) +
```

```

      ln_LTDEBT + dplyr::lag(ln_TQ) +dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures, 2:99),
data=Table6.low.off, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation = 'd')

```

```
summary(Table6.pgmm.low.off)
```

```
#small
```

```
Table6.pgmm.high.off.small<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) +
dplyr::lag(ln_Profits)+ dplyr::lag(ln_Sales) +
      ln_LTDEBT + dplyr::lag(ln_TQ) +dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures, 2:99),
data=Table6.high.off.small, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation
= 'd')
```

```
summary(Table6.pgmm.high.off.small)
```

```
Table6.pgmm.low.off.small<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) +
dplyr::lag(ln_Profits)+ dplyr::lag(ln_Sales) +
      ln_LTDEBT + dplyr::lag(ln_TQ) +dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)|
dplyr::lag(ln_CapitalExpenditures, 2:99), data=Table6.low.off.small, model="twosteps", effect="twoways",
index=c("Company_Name", "year"), transformation = 'd')
```

```
summary(Table6.pgmm.low.off.small)
```

```
#large
```

```
Table6.pgmm.high.off.large<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) +
dplyr::lag(ln_Profits)+ dplyr::lag(ln_Sales) +
```

```

      ln_LTDEBT + dplyr::lag(ln_TQ) +dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures, 2:99),
data=Table6.high.off.large, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation
= 'd')

```

```
summary(Table6.pgmm.high.off.large)
```

```

Table6.pgmm.low.off.large<- pgmm(ln_CapitalExpenditures ~ dplyr::lag(ln_CapitalExpenditures) +
dplyr::lag(ln_Profits)+ dplyr::lag(ln_Sales) +
      ln_LTDEBT + dplyr::lag(ln_TQ) +dplyr::lag(ln_INTEXP) + dplyr::lag(ln_INTINC)+
      dplyr::lag(ln_DIV) + dplyr::lag(ln_STKISSUE) + dplyr::lag(ln_STKREP) +
dplyr::lag(ln_NDEBTISSUE)+
      dplyr::lag(ln_IF)+ dplyr::lag(ln_NOC)+dplyr::lag(ln_NCNE)| dplyr::lag(ln_CapitalExpenditures, 2:99),
data=Table6.low.off.large, model="twosteps", effect="twoways", index=c("Company_Name", "year"), transformation =
'd')

```

```
summary(Table6.pgmm.low.off.large)
```

#Additional Content due to misinterpretation into their paper

#Clean up the database /1st interpretation----

```
#From their report: "We also removed firms with no information for all years of capital expenditure,
```

```
#sales, net property plan and equipment, long-term debt, interest expenses,
```

```
#cash and short-term securities, total assets, total liabilities and equities
```

```
#Remove all firms with no information on the abovementioned, first check
```

```
summary(Table6$IQ_CAPEX) # No NAs in Capx
```

```
#Clean-up LTDEBT
```

```
summary(Table6$IQ_LT_DEBT)# 2294 NAs in Long-term debt
```

```
Table6 <- mutate(Table6,clean=(ifelse(!is.na(Table6$IQ_LT_DEBT), "1", "0"))) # create a binary code based on the
NAs
```

```
Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
```

```
as.table(Table6.count)
```

```
names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging
```

```
Table6 <- merge(Table6, Table6.count, by='Var1')
```

```
Table6<- mutate(Table6,clean=as.numeric(Table6$clean))
```

```
Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
```

```
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
```

```
Table6 <- merge(Table6, Table6.clean, by='Var1')
```

```
Table6<-Table6 [!(Table6$x < Table6$Freq),] #if clean < freq remove company
```

```

names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for clarification
rm(Table6.clean)
rm(Table6.count)
Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#Clean up Interest Expenses
summary(Table6$IQ_INTEREST_EXP) # 172 NAs in Interest Expenses
Table6 <- mutate(Table6,clean=(ifelse(!is.na(Table6$IQ_INTEREST_EXP), "1", "0"))) # create a binary code based on
the NAs
Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
as.table(Table6.count)
names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.count, by='Var1')
Table6<- mutate(Table6,clean=as.numeric(Table6$clean))
Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.clean, by='Var1')
Table6<-Table6 [!(Table6$x < Table6$Freq),] #if clean < freq remove company
names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for merging
rm(Table6.clean)
rm(Table6.count)
Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#Sales (Thus Revenue)
summary(Table6$IQ_REV)
Table6 <- mutate(Table6,clean=(ifelse(!is.na(Table6$IQ_REV), "1", "0"))) # create a binary code based on the NAs
Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
as.table(Table6.count)
names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.count, by='Var1')
Table6<- mutate(Table6,clean=as.numeric(Table6$clean))
Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.clean, by='Var1')
Table6<-Table6 [!(Table6$x < Table6$Freq),] #if clean < freq remove company
names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for merging
rm(Table6.clean)
rm(Table6.count)

```

```

Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#NPPE
summary(Table6$IQ_NPPE) # no NAs

#Cash and short-term securities / investment
#Why surpress the firm when the variable is not being used?
summary(Table6$IQ_CASH_ST_INVEST)

Table6 <- mutate(Table6, clean=(ifelse(!is.na(Table6$IQ_CASH_ST_INVEST), "1", "0"))) # create a binary code based
on the NAs

Table6.count <- table(Table6$Company_Name) #count the appearances in the DB
as.table(Table6.count)

names(Table6)[names(Table6) == "Company_Name"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.count, by='Var1')

Table6<- mutate(Table6, clean=as.numeric(Table6$clean))

Table6.clean <- aggregate(Table6$clean, by=list(Category=Table6$Var1), FUN=sum)
names(Table6.clean)[names(Table6.clean) == "Category"] <- "Var1" #rename for merging
Table6 <- merge(Table6, Table6.clean, by='Var1')

Table6<-Table6 [!(Table6$x < Table6$Freq),] #if clean < freq remove company
names(Table6)[names(Table6) == "Var1"] <- "Company_Name" #rename for merging
rm(Table6.clean)
rm(Table6.count)

Table6 <- Table6 %>% select(-(69:71)) #remove the unnecessary columns

#Total Assets
summary(Table6$IQ_TOTAL_ASSETS)# No NAs

#Total Liabilities
summary(Table6$IQ_TOTAL_LIAB) # No NAs

#Equities
summary(Table6$IQ_TOTAL_EQUITY)# No NAs

#We also removed observations with no information on market cpaitalisation at the end of the year / done by Hugo
# with duplicate observations, negative values for interest income and positive values for interest expenses and
dividends
count(Table6$IQ_INTEREST_INCOME<0) # No negative values left
count(Table6$IQ_TOTAL_DIV_PAID_CF>0) # No positive values left
count(Table6$IQ_INTEREST_EXP>0) # No positive values left

```

```
#Their Final sample 2049 companies on average per year
#1995-2011 timeframe
```

```
count(Table6$year==1995)#1195
count(Table6$year==1996)#1526
count(Table6$year==1997)#1751
count(Table6$year==1998)#1776
count(Table6$year==1999)#1759
count(Table6$year==2000)#1721
count(Table6$year==2001)#1640
count(Table6$year==2002)#1575
count(Table6$year==2003)#1530
count(Table6$year==2004)#1517
count(Table6$year==2005)#1516
count(Table6$year==2006)#1494
count(Table6$year==2007)#1455
count(Table6$year==2008)#1437
count(Table6$year==2009)#1356
count(Table6$year==2010)#1304
count(Table6$year==2011)#1298
```

```
A<-c(1195,1526,1751,1776,1759,1721,1640,1575,1530,1517,1516,1494,1455,1437,1356,1304,1298)
```

```
B<-c(1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011)
```

```
Table6.apy <- data.frame(A,B)
```

```
mean(Table6.apy$A) #1520.588 firms yearly on average, which differs from their assumption of 2049 companies
average per year -529 companies less
```

```
rm(A)
```

```
rm(B)
```

```
view(Table6.apy)
```

```
#Removing firms itself is questionable
```

```
#Split dataframe into small and large companies V1----
```

```
#Large and small companies are those firms in the upper and lower median of total assets
```

```

Table6.median<-aggregate(Table6$IQ_TOTAL_ASSETS,
by=list(Company_Name=Table6$Company_Name),FUN=median,na.rm=TRUE, na.action=NULL,c("median"))
names(Table6.median)[names(Table6.median) == "x"] <- "individual.median" #renaming / taking the individual median
over

#Split and in order to do so find the median along which we will do so

Table6.median <-mutate(Table6.median, median=median(Table6.median$individual.median)) #median for the whole
dataframe

Table6.large <- Table6.median %>% filter(Table6.median$individual.median>=median) %>% select(Company_Name)
Table6.small <- Table6.median %>% filter(Table6.median$individual.median<median) %>% select(Company_Name)

#Assign a binary column

Table6.large <- mutate(Table6.large, size=1)
Table6.small <- mutate(Table6.small, size=0)

#Joining the coded data

Table6.small.large <- rbind(Table6.large, Table6.small)

#Join the binarycode with the dataset

Table6 <- merge(Table6, Table6.small.large, by='Company_Name')

#Do the final split of the data

Table6.small.f <- Table6 %>% filter(size == 0)
Table6.large.f <- Table6 %>% filter(size ==1)

#Clean-up
rm(Table6.small)
rm(Table6.median)
rm(Table6.large)
rm(Table6.small.large)

```