University of Nottingham

Applied Microeconometrics

Group Project A

Insert Title

Spring Term 2020

Supervisor
Professor Sourafel Girma

Authors
Nelly Lehn (20214338)
Yonesse Paris (20115536)
Thea Zoellner (20216019)
Georg Schneider (20214032)
Emilie Bechtold (20214031)

Contents

List of Tables

1 Introduction

2 Theoretical Background/Literature Review

2.1 FDI

2.2 PSM

Since (I guess) we will be focussing on ATE rather than ATT, we need to satisfy the following two assumptions:

1. Assumption: **Unconfoundedness (CIA)**
   "*[G]iven a set of observable covariates X which are not affected by treatment, potential outcomes are independent of treatment assignment*" (**?**, p. 35)

2. Assumption: **Overlap**
   "*persons with the same X values have a positive probability of being both participants and nonparticipants*" **?**, p. 35

–> if Assumption 1 holds, all biases due to observable components can be removed by conditioning on the propensity score (Imbens, 2004).

Binary Treatment

Difference between logit and probit lies in the link function. Logit assumes a log-distribution of residuals, probit assumes a normal distribution. Heteroskedastic probit models can account for non-constant error variances –> Check for heteroskedasticity?

Multiple Treatments

The multinomial probit model is the preferable option compared to logit. Alternatively, just run several binary ones (more complicated but also more robust to errors).

Variable selection

- outcome variable must be independent of treatment conditional on the pscore (CIA)

- Only variables that influence simultaneously the participation decision and the outcome variable should be included (based on theory and empirical findings)

- variables should either be fixed over time or measured before participation (include only variables unaffected by participation)

- choice of variables should be based on economic theory and previous empirical findings

Tests for variable selection

Strategies for the selection of variables to be used in estimating the propensity score: ...

3 Data and Descriptive Analysis

Our analysis is based on observational firm-level data. The dataset comprises 11,323 firms, of which 4,460 received FDI in 2016. The FDI can be divided into three subcategories. Table 1 shows the frequencies of each type of FDI in our sample. Among the recipients of FDI, most firms (1,965) received domestic market seeking FDI. 1,555 firms received technology intensive FDI and the remaining 640 firms received exports oriented FDI. The TFP outcome was measured in 2017. We standardize [1] the outcome variable to make its interpretation more intuitive.

Table 1: Frequency of FDI Types

FDI type	Abs. Freq.	Rel. Freq.
No FDI	6,863	61%
Exports oriented FDI	940	8%
Technology intensive FDI	1,555	14%
Domestic market seeking FDI	1,965	17%
Total	11,323	100%

A set of categorical and continuous control variables was measured in 2015 (one year prior to the firms receiving FDI). Table 2 provides an overview of the categorical variables and the frequencies of each category in our sample. The categorical variables are included as factor variables in the subsequent analysis. The port variable indicates whether a firm has access to a port within 500km. The legal ownership of a firm is captured in the ownership variable, where state owned firms represent the base group. The technology intensity of the industry the respective firm is operating in, is measured in four categories from low- to high-tech, low-tech being the base group. The R&D dummy indicates whether a firm has invested in Research and Design in 2015.

---

[1] Using Z-score standardization

Table 2: Summary Statistics of Categorical Covariates

	Abs. Freq.	Rel. Freq.
Port[a]		
No	7,366	65.05
Yes	3,957	34.95
Ownership		
Listed company	909	8.03
Subsidiary	2,630	23.23
Independent	4,593	40.56
State owned	3,191	28.18
Technology Intensity		
Low-tech	4,194	37.04
Medium low-tech	1,685	14.88
Medium high-tech	3,539	31.25
High-tech	1,905	16.82
R&D[b]		
No	9,951	87.88
Yes	1,372	12.12

[a] Indicates whether a firm has access to a port within 500km.
[b] Indicates whether a firm has invested in R&D in 2015.

The summary statistics of the continuous variables wages, Total Factor Productivity (TFP), firm size (measured in number of employees), debts and the firms' export intensity are displayed in Table 3. The variables wages, employment and, to a lesser extent, debts show large differences between their mean and median values, hinting at the existence of outliers in the sample. Log transforming the skewed variables is an easy way of reducing the influence of extreme values.

While there is no reason not to do so in the case of wages and debts, the covariate balance after estimating the propensity score when including the logged employment variable does not yield satisfying results [SEE SOMEWHERE???], which is why we prefer to include the untransformed employment variable in our estimation model. A visual analysis of the employment variable's distribution in our sample (see Figure 1) shows that there is at least one extreme value which could potentially bias our results. In section [ADD SECTIONNUMBER] we therefore test the sensitivity of our findings to excluding observations with extremely high values and find no evidence of distortions caused by outliers in the employment variable.

Table 3: Summary Statistics of Continuous Covaraites

	Mean	Median	Sd	Min	Max
Wages	1,967[a]	1,538	50,990[a]	0.00065	5,519,000[a]
TFP	3.041	3.032	2.047	-5.359	11.36
Employment	7,111	81.39	117,155	0.00197	8,824[a]
Debt	1.762	1.649	0.634	0.819	3.668
Export intensity	0.159	0.154	0.0798	0.0103	0.483

Note: All variables in levels.
[a] In Thousands

To further motivate the use of matching estimators in estimating the effect of FDI on a firm's TFP, the differences in means between the firms that received FDI and the firms that did not are displayed in Table 4. The t-tests show significant differences in all observable characteristics, meaning that there is in fact selection into treatment.

4 Empirical Specification

-Overlap??

In order to evaluate the causal effect of FDI on Total Factor Productivity (TFP) we estimate our main model described in equation (1) using nearest neighbour matching with one neighbour and with replacement. Second, we also estimate a nearest neighbour model with 5 neighbors with replacement and a caliper of 0.05. Further, we use inverse probability weighting (IPW) and augmented-inverse probability weighting (AIPW) estimators. Since we have a large sample, the different estimators should all yield similar results.

For the matching, the weighing and regression adjustment models we use the same specification including the beforementioned categorical and continuous variables. We unlog employment to improve covariate balance. We do not include the export variable as a matching covariate, since together with technology it predicts treatment too well and nearly eliminates overlap. As propensity score matching requires that all covariates which affect both outcome and likelihood of treatment, we can exlude export since it does not influence the outcome variable. As Girma et al (2004) show, productivity increases exports, not the other way around. We do not include the port variable for the same theoretical reason.

4

Table 4: Difference in Pre-Treatment Covariate Means

	(1) Control	(2) Treatment	T-test Difference (1)-(2)
Technology intensity	2.565	1.838	0.728***
	(0.014)	(0.015)	
Access to port	0.273	0.467	-0.194***
	(0.005)	(0.007)	
Log wages	7.529	7.031	0.498***
	(0.046)	(0.057)	
TFP	3.185	2.821	0.364***
	(0.025)	(0.030)	
Log employment	3.766	5.405	-1.639***
	(0.037)	(0.041)	
Log debts	0.511	0.493	0.019***
	(0.004)	(0.005)	
Export intensity	0.131	0.204	-0.073***
	(0.001)	(0.001)	
R&D dummy	0.117	0.128	-0.012*
	(0.004)	(0.005)	
Observations	6863	4460	

Notes: Columns (1) and (2) show the pre-treatment covariate means of the control and treatment group respectively. Standard errors are displayed in paratheses. The values displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 1: Outliers in Employment Variable
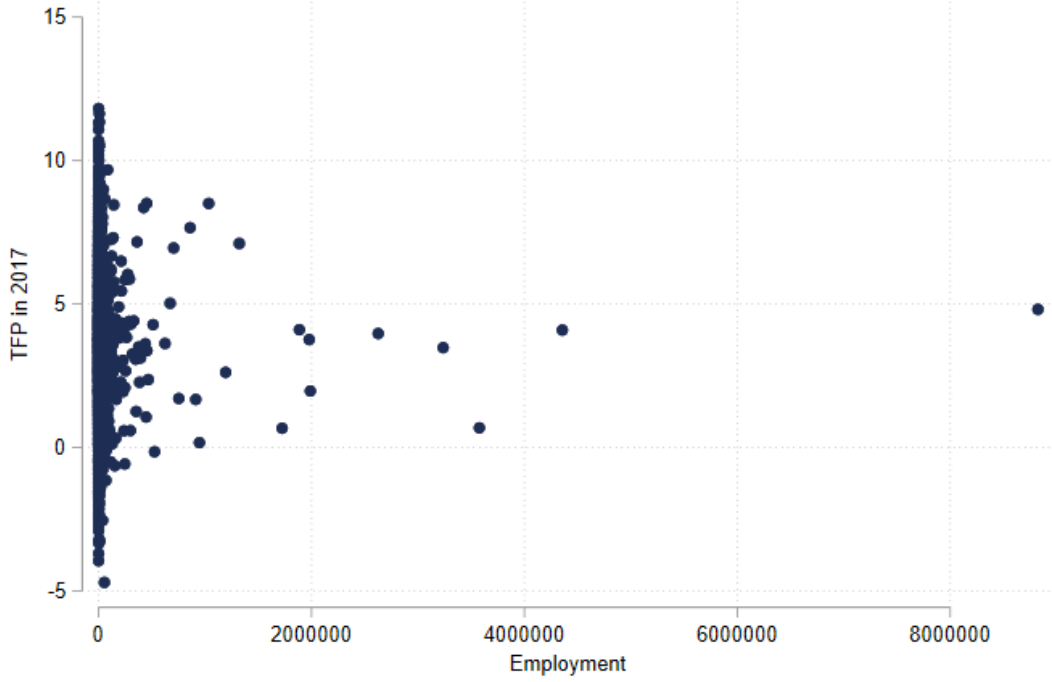
$$TFP = \beta_0 + \beta_1 FDI + \beta_2 X_{i2015} + \epsilon_i$$

where

- *FDI* is a non-random treatment variable. It takes value 1 in case of treatment, 0 otherwise;

- X_{i2015} describes control variables and consists of following pre-treatment covariates: Ownership, Technology Intensity, Research&Development, logarithm of Wages, Total Factor Productivity, Employment and Debts

Our matching model is thus a logit regression of TFP in 2017 on Ownership, Technology Intensity, Research&Development, logarithm of Wages, TFP in 2015, Employment and Debts. This very good overlap and covariate balances (the latter discussed in more detail down below). Figure two shows proof of sufficient overlap for a matching analysis.

6

Figure 2: Overlap in Main Model
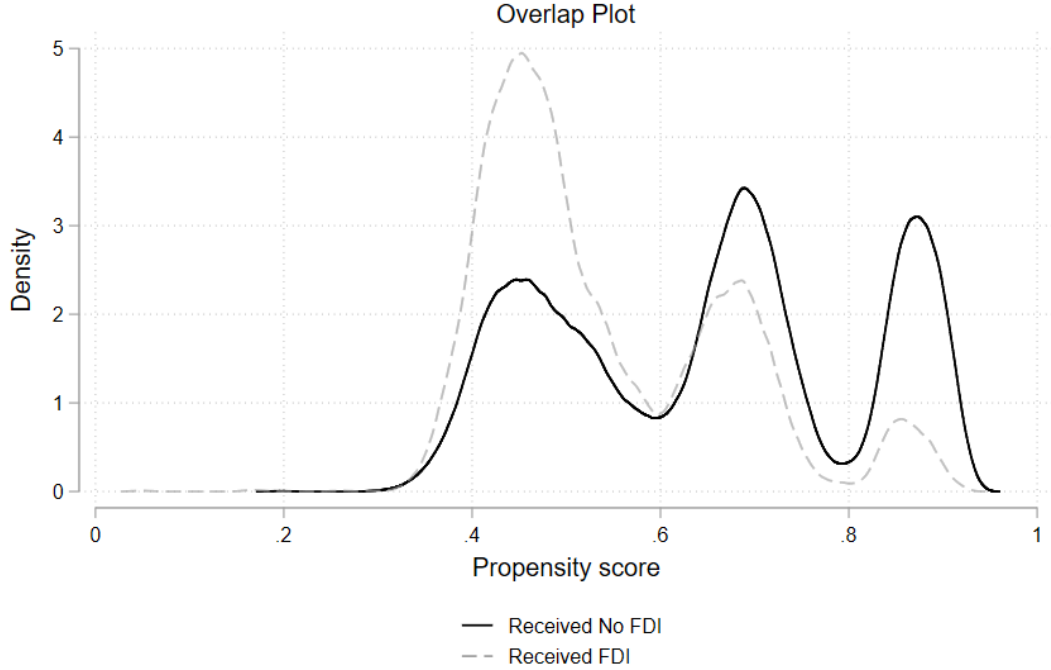
The main findings of this paper are displayed in Table 5. It reports the Average Treatment Effects (ATE) for TFP. Across estimators, we find large and highly significant coefficients showing that receiving FDI increases TFP of companies on average. Nonsurprisingly, the reported coefficients differ only slighlty in magnitude. This suggests that our model is not sensitive to different model estimations, which supports our chosen model specification. Column (1) shows the results of a one-to-one propensity score matching with replacement. Receiving FDI leads for the average company to an increase in TFP of 0.13 standard deviations. Sligthly lower results are obtained from a propensity score matching with five nearest neighbors and a caliper of 0.05 as well as for the invers IPW and AIPW-estimators.

The main findings of this paper are displayed in Table 5. It reports the average treatment effects (ATE) for TFP. Independent of the estimator, we find highly significant coefficients showing that receiving FDI increases TFP of companies on average. Nonsurprisingly, the reported coefficients differ only slighlty in magnitude. This suggests that our model is not sensitive to different model estimations, which supports our chosen model specification. Column (1) shows the results of a one-to-one propensity score

matching with replacement. Receiving FDI leads for the average company to an increase in TFP of 0.13 standard deviations. Sligthly lower results are obtained from a propensity score matching with five nearest neighbors and a caliper of 0.05 as well as for the invers IPW and AIPW-estimators. Taking the covariate balances into consideration, we prefer the one-to-one propensity score matching as it gives us better covariate balance as compared to the alternative estimators *see Appendix*. Following, we propbe the sensitivity of your findings in more detail for one-to-one propensity schore matching.

4.1 Effect of FDI on TFP

Table 5: ATE of FDI on TFP

	NN1	NN5[a]	IPW	AIPW
	(1)	(2)	(3)	(4)
FDI2016	0.130***	0.114***	0.122***	0.142***
	(0.015)	(0.011)	(0.007)	(0.003)
0.FDI2016				
P0 Means			-0.068***	-0.057***
			(0.010)	(0.009)
Observations	11,323	11,318	11,323	11,323

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Covariates include, unless otherwise specified: Ownership, Technology Intensity, Research&Development, logarithm of wages, Total Factor Productivity, Employment and Debts.
All shown coefficients are normalized for better interpretation.
[a] We use the Propensity Score Matching method with five nearest neihbors and replacement as well as a 0.05 caliper.

4.2 Robustness of Results

Table 6 reports several alternative model specifications that confirm the robustness of our main results. The positive and significant effect of foreign investment on total factor productivity persists. In column (1), we add interaction terms of the dummy variables with the continuous regressors to our set of covariates. The average treatment effect of FDI on productivity slightly increases by 0.027 standard deviations. The covariate balance does not improve with the inclusion of interaction terms, suggesting that interactions do not increase the quality of matching.[2] One possible source of bias are the outliers of

---

[2]The same holds true when interacting only dummy variables, only continuous variables or all variables.

employment. While most of the firms' number of employees is concentrated around a mean of 7,111, we are concerned about two observations with extreme values: one firm with over eight million employees, and another one that apparently hired more than 4 million people in 2015.[3] To check whether these outliers influence our main findings, we exclude the two extreme observations. The results reported in column (2) show no significant change in the average treatment effect. In our main specification, we have further assumed that a nearby port does not influence total factor productivity. Column (3) confirms that this is indeed the case, as the results do not change with the inclusion of the dummy variable port in our set of covariates.

Table 6: Robustness of Results[a]

	Including Interactions[b] (1)	Excluding Outliers[c] (2)	Including Port[d] (3)	Effect on the Treated (4)
ATE	0.152*** (0.016)	0.127*** (0.015)	0.125*** (0.019)	
ATT				0.127*** (0.017)
Observations	11,323	11,321	11,323	11,323

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
[a] All specifications use the Propensity Score Matching method with one nearest neighbor and replacement. Covariates include, unless otherwise specified: Ownership, Technology Intensity, Research&Development, logarithm of Wages, Total Factor Productivity, Employment and Debts.
[b] Interacting dummy variables (Ownership, Technology Intensity, Research & Development) with continuous variables (Logarithm of wages, Total Factor Productivity, Employment and Debts).
[c] Two observations with extreme values of Employment 2015 above four million are excluded.
[d] Adding a dummy variable indicating whether a port lies within 500km of the firm to the set of covariates.

Finally, column (4) reports the average treatment effect on the treated (ATT) of the nearest neighbor matching with one neighbor and replacement. Because we assume that treatment is not exogenous, we might find stronger effects of treatment on the treated. This would be the case if those firms receiving treatment are also the ones benefiting

---

[3] Given the limited information our dataset contains, we cannot be sure in which unit employment is measured. We therefore suppose that the common definition of employment being the number of hired employees holds.

more from it. However, our estimate in column (4) reports an ATT that is very similar to the average treatment effect for the hypothetical case that all firms have received FDI. This suggests that the propensity score matching performed very well in randomizing treatment and control groups.

One shortcoming of our analysis is the missing capability of explaining the mechanisms behind the observed treatment effects. We do not distinguish between within-firm effects and spillover effects from firms receiving FDI on other firms. Due to the limited information on a firm's sector and location, we are unable to measure spillover effects. That is, the literature on spillover effects typically suggests spillovers within sectors rather than across sectors (SOURCE) and higher spillover effects on nearby firms or firms that are trading factor inputs with foreign-owned companies (SOURCE).

5 Analysis by Type

To further test the robustness of our results we continue our analysis by looking at potential heterogeneity of the treatment effect across types of FDI. Doing so we can test the possibility that one specific type of Investment drives our previous results. From the data we can distinguish between three different types of FDI: (i) exports-oriented FDI, (ii) technology intensive FDI and (iii) domestic market seeking FDI. Table 5 shows their absolute and relative frequencies. It is possible, that for example only exports-oriented increased factor productivity while the other two types had little or no impact.

To test for this possibility we estimate an augmented IPW model with multi-valued treatment effects. The matching covariates are the same as in the previous model and the regression adjustment specification is the same as that for propensity score estimation. The covariate balance is good, with the highest standardized differences of 5.7% and a minimum variance ratio of 0.55 for employment (though all other variance ratios are close to 1).

Second, we estimate an IPW model to check if it yields similar estimates without regression adjustment. The covariate balance in this model is practially the same. Finally, we estimate a set of AIPW models where we restrict the sample to one type of treatment each. This allows for the IIA assumption to be relaxed which is required for the Mulitnomial Logit model. The first logit model has a maximum standardized difference of 12% for employment and and a maximum variance ratio of 1.36 for the same variable. The second model fares better in terms of standardized differences but has has a variance ratio of 1.88 for employment. The third model has great balance in all variables. The overlap assumption is satisfied for all treatment levels as can be seen in graph 2.

In Table 7 the results from the typewise analysis are shown. In the AIPW Multinomial specification the differences in ATE between different types of FDI are below half a percent of a standard deviation. This suggests, that all types of FDI increase factor productivtiy by essentially the same margin. The estimated effect size is close to the one estimated for FDI in tabl 2. In the IPW specification the differences are a bit larger but no bigger than 4.5% of a standard deviation. The seperate logit models also yield essentially the same effect sizes as the multinomial specification. We thus conclude that all types of FDI increase productivity by

The Inverse Probability Weighting Model gives us somewhat bigger differences in effect sizes between the types. The difference between the bigger effect of Exports oriented FDI and the smaller effect of Technology intensive FDI amounts to 4% of a standard deviation. Including interaction terms between continuous and categorical regressors

With the AIPW being a doubly robust estimators and rather small differences in the IPW estimator we take these results to suggest that all types of FDI have similar positive impacts on factor productivity.

Table 7: ATE by Type of FDI

	(1) AIPW Mlogit	(2) IPW Mlogit	(3) AIPW Logit	(4) AIPW Logit	(5) AIPW Logit
Exports-oriented FDI	0.144***	0.157***	0.140***		
	(0.006)	(0.032)	(0.007)		
Technology intensive FDI	0.139***	0.112***		0.139***	
	(0.005)	(0.018)		(0.005)	
Domestic market seeking FDI	0.143***	0.134***			0.143***
	(0.004)	(0.011)			(0.004)
PO Means	-0.057***	-0.068***	-0.012	-0.025**	-0.017
	(0.009)	(0.010)	(0.011)	(0.011)	(0.011)
Observations	11,323	11,323	7,803	8,418	8,828

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

To account for the possibility that the choice of the FDI type does not satisfy the IIA assumption we further estimate separate logit models for the two estimator types. The results, reported in table 7, are very similar to those obtained from a multinomial specification, suggesting that the IIA assumption holds.

11

Figure 3: Propensity Score by Treatment Level
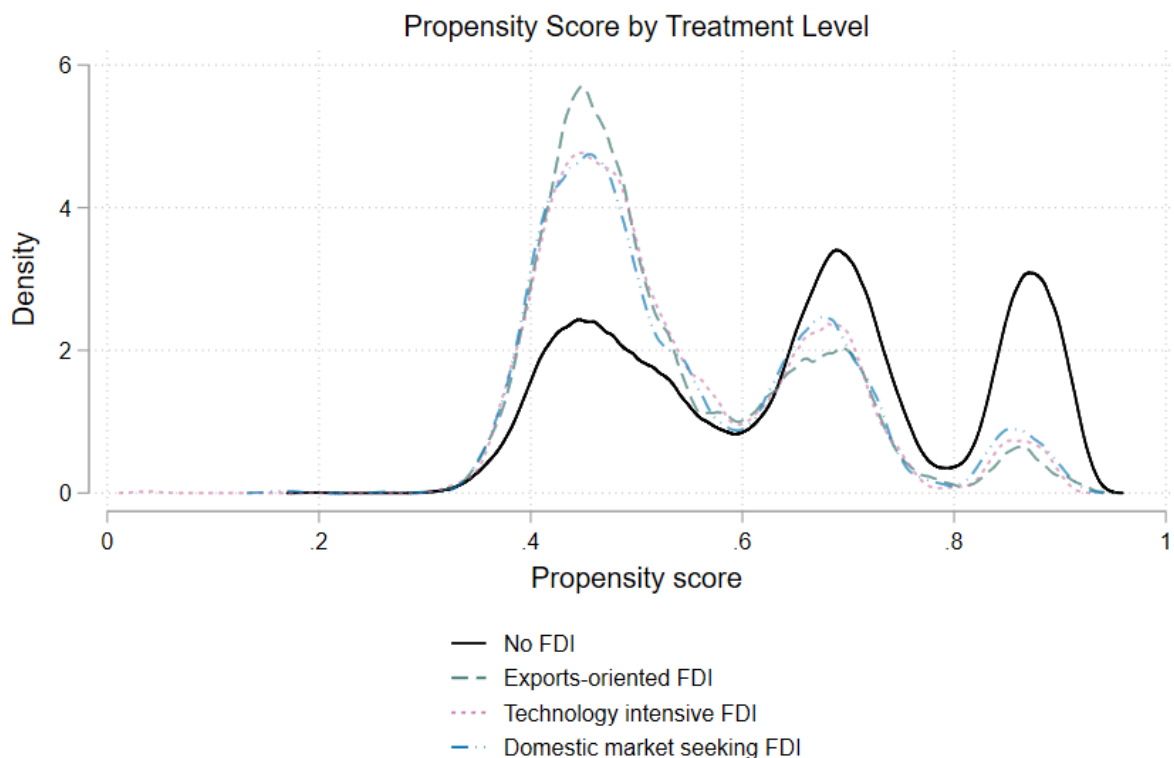
6 Discussion/Conclusion

For citation:

you have to add your reference firstly in bibCG. After having done so you can always include the reference in the actual file as follows:

?

(**?**, p. 35)

Thoughts on what we could write for discussion/limits of our study:

1. Do not know much about the context of the treatment (so cannot really rule out anticipation-effects?)

2. Would have been interesting to extend the study to several years after the treatment. Do effects persist? Do they vanish?

3. Might depend on firm size (see Aitken & Harrison 1999 → will include citation): find positive within-plant effects and spillover effects on TFP for small firms only (less than 50 employees)

4. Do not measure spillovers on plants that have not received FDI

5. Do not have sector-specific data → TECH variable has only 4 categories; e.g. in order to measure spillover effects from other firms in sector this would be necessary (i.e. if a foreign firm is more innovative)

Appendix