Global value chains: Can quality certifications improve local firms' integration? Evidence from Latin America and the Caribbean

Federico Bernini* Lucas Figal Garone†

Alessandro Maffioli‡ Andrés Sebastián Mena§

Work in Progress – July 2019 – Please do not cite

Abstract

This study evaluates the impact of internationally recognized quality certification (QC) adoption on firms' exports, local sales, productivity, and access to credit. Using a panel dataset of 5.410 firms from Latin America and the Caribbean we follow a two-stage identification strategy: (1) We estimate firms' probability of QC adoption using Random Forest technique, and (2) we use the estimated probabilities for apply a weighted Diff-in-Diff approach. Our findings show that acquiring a QC has a positive effect on firm export behavior. Interestingly, we find that this effect is driven by an increase in the extensive margin of indirect exports and the intensive margins of direct exports. QC also helps easing constraints in the access to finance. Conversely, no effect was found on various measures of firm productivity and local sales. This evidence is consistent with the idea that QC adoption reduces entry barriers to foreign markets which improves the integration of local firms into existing Global Value Chains via indirect exports and the access to pre-export or value-chain finance.

JEL Classification Codes: D22, D24, D82, L15, L25
Keywords: Quality certification, information asymmetry, impact evaluation, firm performance, panel data, Latin America and the Caribbean.

*Department of Economics, Universidad de San Andrés, Buenos Aires, Argentina E-mail: fg-bernini@gmail.com

†Strategy & Development Department, IDB Invest, Buenos Aires, Argentina, and Department of Economics, Universidad de San Andrés, Buenos Aires, Argentina. E-mail: lfigal@iadb.org.

‡Strategy & Development Department, IDB Invest, Washington D.C., United States. E-mail: alessandrom@iadb.org.

§Universidad Nacional de Tucumán, Tucumán, Argentina and CONICET, Argentina E-mail: asmena@face.unt.edu.ar

1 Introduction

The Latin America and Caribbean (LAC) is a developing, middle-income region that has managed to raise its per capita GDP over the last quarter century. However, the gap between LAC and developed countries in terms of income and well-being has yet to be bridged. (Crespi *et al.*, 2014a) find that this situation is not due to relatively smaller increases in the levels of physical or human capital, but rather to the productivity gap, which has been growing in recent decades.[1]

There are a number of explanations for this productivity gap. From a macroeconomic standpoint, severe and persistent economic instability in many LAC countries has discouraged long-term investment. This situation is aggravated by the preponderance of small and medium enterprises (SMEs), whose productivity gap with firms in developed countries is even greater. From a microeconomic perspective, the region has failed to correct certain market failures, such as information asymmetries. Typically, owners or managers of a firm know more about their internal operations and future prospects than external agents - i.e. investors, customers or creditors - can know. This information asymmetry can translate into severe barriers for firm's growth and often generates additional costs.

When a firm requests a loan, for example, the financial entity must bear the cost of gathering data about that firm's solvency and future prospects. To mitigate the costs of information asymmetry, firms need to seek mechanisms to demonstrate their desirable characteristics that are not easily observable by all agents. Quality standard certifications are effective for this purpose and represent a widely accepted solution, as they provide a guarantee that the firm implements high-quality business and manage practices. King *et al.* (2005) argue that managerial standards, such as those granted by the International Organization for Standardization, enable firms to demonstrate characteristics that are not typically observable by third parties.

Firms that obtain an internationally-recognized quality certification tend to have more desirable characteristics than those that do not. It is unclear, however, whether this relationship is causal, or whether firms deciding to adopt this certification, for example, are already close to (or have surpassed) the quality threshold required to obtain it. In any case, firms can use this certification as a mark of quality to address the information asymmetry between them and the external agents with whom they interact.For the purpose of this study, we classified these agents into three groups: (i) domestic customers, for whom the certification is a guarantee that the firm complies with certain quality standards; (ii) foreign customers, with whom greater information problems arise, and therefore who are more demanding with respect to process and product quality standards, which increases the importance of certification; and (iii)

[1] For this reason, many productive development policies (PDPs) has been recently supported in LAC. Recently, some of them have been rigorously evaluated. These include innovation policies (Crespi *et al.*, 2015) and their spillovers (Castillo *et al.*, 2016b), cluster development (Figal Garone *et al.*, 2015; Figal Garone & Maffioli, 2016), regional industrial policy (Castillo *et al.*, 2017), technical assistance for MSMEs (Castillo *et al.*, 2016a), and scientific research funds (Benavente *et al.*, 2012), among others.

financial institutions, which through the certification, also receive relevant information by the certification about the firm's economic situation and good business practices, thereby improving its credit risk assessment.

In this paper we analyze the role of acquiring international certification on removing informational barriers that hinder firm's development and growth. First we describe the firms in LAC that obtained an internationally-recognized quality certification [2]. Then we explore the effects of this certification on several variables of firm performance. The evidence on the impact of acquiring a quality certification on firm performance is scarce and it is specially important to firms in developing countries (as Latin America and Caribbean) where information asymmetries are larger and firms' growth were slow. We argue that analyzing several outcomes is important not only to understand the impact in different firm-level dimensions, but also because each outcome is related to a different external agent of the firm. Therefore, it allows us to explore why firms certificate and, at the same time, identify for which agents the quality certification really matters as signaling mechanism of relevant characteristics of the firm.

We use the World Bank Enterprise Survey (WBES) carried out in Latin America in 2006 and 2010 (with the exception of Brazil, where it was conducted from 2003 to 2009), and the WBES carried out in 2011, in combination with the new round carried out in 2014 (PROTEqIN) for Caribbean countries. Overall, we have data from 19,499 enterprises in 32 countries in the region, where 5,410 firms were surveyed in two years and allow us to construct a panel database. Our identification strategy for the impact of quality certification is a weighted difference-in-difference approach, which allows us to deal with potential endogeneity issues. Weights are introduced in order to conduct a double-robust evaluation assuring that treated and control groups' covariates are equal in mean at baseline. To avoid biased conclusions due to weights misspecification, we estimate two different Inverse Probability Weights using the traditional probit model and also random forest.

Results indicate that adoption of international quality certifications is more frequent among exporters, foreign firms and enterprises with greater sales and more experienced managers. Furthermore, obtaining international quality certification has a positive effect on the business objectives most affected by problems of information asymmetry, such as entry into international markets (the intensive margin of direct exports and the extensive margin of indirect exports) and access to financing. For local sales we find positive effects, but only significant for the Random Forest method, indicating that domestic customers give less importance to quality certifications than international ones. Finally, adopting certification seems to have no statistically significant effects on firm productivity. This study therefore finds evidence that the main benefit to firms that obtain quality certification is not linked to an improvement in productivity or local sales, but rather to quality signaling for foreign clients and credit institutions, by demonstrating certain desirable characteristics that are difficult to verify in the absence of such certification.

[2] We focus on process certification, such as ISO 9000 certificates.

To test the robustness of our results, we perform two falsification tests using those firms that initiated a quality certification process but not received it yet and also those firms that based on observed covariates are closer to treated ones[3]. No significant effects were found.

The paper is organized as follows: Section 2 contains a review of the literature on quality certification adoption and its effects. Section 3 presents the data and variables used, describes the sample and shows the differences in the variables between firms that adopt international standards and those that do not. Section 4 carefully details our identification strategy by indicating how we defined treated and control groups, the formal Diff-in-Diff specification, the Probit and Random Forest methods for propensity score estimation, and the results of the first stage propensity score estimation and the resulting baseline covariates mean balance. Section 5 analyzes the impact of adopting international standards on firmsâĂŹ local sales, export bahavior, finance restriction and productivity. Section ?? assess the robustness of the results using falsification tests. Section 6 concludes and offers some final considerations.

2 Literature Review

Firms that opt to obtain an international quality certification (e.g., ISO standards) must ensure that they are implementing best business practices by establishing a quality policy with measurable objectives, complying with certain requirements concerning customer satisfaction, and providing the necessary training for staff to reach the required level of competence, among other activities. Hudson & Orviska (2013) developed a model that shows that obtaining international quality certification allow firms to demonstrate higher quality and less uncertainty about their activities and products or services. Quality signaling is particularly important for firms in developing countries. As the general perception of the relationship between quality and the country's per capita income is already unfavorable, firms seeking to export need an alternative way of showing that their products can also be of high quality.

The empirical literature on the determinants of the adoption of international quality standards and its impact on firm's performance is quite recent. The evidence shows that exporters, larger firms and those with a higher share of foreign ownership are more willing to seek adoption of international quality standards (Fikru, 2014; Hudson & Orviska, 2013; Pekovic, 2010; Ullah *et al.*, 2014). Exporters are more likely to adopt a certification because there is a greater information asymmetry with their foreign customers than their local clients and, in addition, because there is a harder competition in foreign than local markers. At the same time, larger firms and companies with foreign ownership generally enjoy greater access to financial resources and have superior management, enabling them to achieve certification more readily.

The literature about the impact of the adoption of quality certifications on firm's

[3] We employed the Genetic-Matching (**?**) algorithm to find the nearest-neighbours of treatment group and falsified them as treated

performance has focus on two main variables: exports and productivity. In terms of the literature about the impact of certificate on exports, there are various studies that find significant effects on both the probability of exporting and the volume exported, that is, on both the intensive and extensive margin of trade (e.g., Otsuki, 2011; Sun & Outyang, 2014; Volpe Martincus *et al.*, 2010; Xiaoyang Chen *et al.*, 2008). These findings are important not only in terms of developing the export potential of firms in a given country, but also because there is evidence that firms learn lessons throughout the export process, which enables them to improve their productivity (De Loecker, 2007; Harrison & Rodriguez-Clare, 2010).

In contrast, the evidence about the effects on firm's productivity is less conclusive. While there is evidence that certified firms are more productive (Dick *et al.*, 2008; Starke & Rangamonhan, 2012; Trifkovic, 2017; Ullah *et al.*, 2014), it is not clear whether firms improve their productivity by obtaining the certification or they obtain the certification because they are more productive. Javorcik & Sawada (2018) shows that ISO 9000 has no effect on labor productivity and average wages on the short run, while there is a significant effect on both variables on the long run. This difference in the timing of the effect reflects that the improvements implemented to certify require maturing time to materialize in a more efficient production process. In addition, after examining data from manufacturing firms in 59 countries, Goedhuys & Sleuwaegen (2013) find that those firms that obtain international standard certification have greater productivity and that the effect is greater for the firms located in countries with weaker market institutions. This finding underscores the importance of exploring these issues in the LAC region, where the potential benefits may be considerable.

Finally, the impact of acquiring a quality certification on other outcomes of interest is almost non-existent. In particular, we will focus on the impact on the access to finance. Ullah *et al.* (2014) show, using the Enterprise Survey, that ISO certified firms exhibit significantly lower level of financial constraints, higher level of exports and productivity. Although the data they used are similar to the one used in this paper, they provide evidence in a cross-section setting for 31 LAC countries and therefore their results rest on the assumption that there is no firm-level unobserved heterogeneity that is not related to both the decision to certify and firm performance.In our case, we exploit panel data and apply a difference-in-difference approach that allow us to control for time-invariant heterogeneity. This approach reduce the potential heterogeneity cause by selection biases, since certified and non-certified firms are clearly not similar.

3 Data and descriptive statistics

This study is based on the World Bank Enterprise Survey (WBES). The WBES is a firm-level survey of a representative sample of an economy's private sector. The survey cover a broad range of business environment topics including access to finance, corruption, infrastructure, crime, competition, and performance measures. In this paper, we use the WBES carried out in Latin America in 2006 and 2010 (with the exception of

Brazil, where it was conducted in 2003 and 2009), and the WBES carried out in 2011 in combination with the new round carried out in 2014 (PROTEqIN) for Caribbean countries. Overall, data were obtained from 22,945 enterprises in 32 countries in the LAC region. Given that 5,410 firms were surveyed in two years, we can construct a panel database for our main estimations.

In Table 1, the main variables used in this study are described. Our main focus is on the variable "Quality certification" that identifies if the firm have an internationally-recognized quality certification. This variable is first used as an outcome variable to explore the determinants of certification adoption. Thereafter, the effect of certification on certain outcomes of interest linked to firm performance is analyzed using it as a "treatment variable".

We focus on a set of ten main outcome variables to evaluate firm's performance and how they are affected by the acquisition of an internationally-recognized quality certification. First, in order to explore the firm's performance in foreign markets, we focus on the export status of the firm and the amount of exports (direct and indirect). Second, we use local sales to assess firm's performance in national markets. Then, we focus on firm's access to finance (own perception as a barrier). Finally, we use labor productivity and TFP measures to evaluate firm's productivity.[4].

We also use a set of control variables to reduced potential biases. We control for firm's characteristics including firm age, number of employees, management experience and whether firm has foreign owners. As well, we have a group of three variables that are related to activities of the firm, such as whether the firm runs training programs; or uses a website to communicate with customers and suppliers; and the percentage of working capital financed by banks.

Table 2 compares the means values of outcomes and control variables between firms that achieved a quality certification and those that did not. In general, firms that have adopted international quality standards tend to perform better. They have greater sales in local markets compared to firms without certification. In terms of international insertion, 50% of certified firms are exporters, whereas only 21% of non-certified firms sell in foreign markets. Moreover, those that manage to get certification export a larger volume, and a higher proportion of their sales −14%, versus only 5% for non-certified firms. Although this preliminary evidence reveals a clear correlation between export orientation and performance for firms with international certification, it is still not sufficient to attribute causality. For example, it may be the case that firms incorporate knowledge and lessons by exporting that subsequently help them to acquire a certification.

Firm productivity is a relevant variable to observe, as it highly correlated with firm's survival, long-term growth, and entry into export markets (Syverson (2011)). The results show that firms that obtained international certification are significantly better in terms of labor productivity but the opposite seems to be the case for TFP.

[4] We computed productivity using other measures (e.g. value added per employee and measures of TFP using input shares in the total cost) and results remained unchanged

Table 1: Description of variables

Variable	Definition
Quality Certification	Binary variable that takes the value 1 if the firm has an internationally recognized quality certification and 0 otherwise
Exporting firm	Binary variable that takes the value 1 if the firm exports 0 otherwise
Direct Exporting firm	Binary variable that takes the value 1 if the firm exports directly and 0 otherwise
Indirect Exporting firm	Binary variable that takes the value 1 if the firm exports indirectly 0 otherwise
Exports	Total amount exported (US dollars)
Direct Exports	Total amount exported directly (US dollars)
Indirect Exports	Total amount exported indirectly (US dollars)
Local sales	Total amount sold in the local market (US dollars)
Financial barrier	Binary variable that takes the value 1 if the firm considers access to financing to be a significant or severe barrier and 0 otherwise
Labor productivity	Sales per employee (US dollars)
TFP	Total factor productivity estimated using the Levinsohn & Petrin (2003) methodology
Employees	Number of permanent full-time employees
Management experience	Years of management experience in the sector
Foreign	Binary variable that takes the value 1 if the firm is own by a private foreign individual/organization in more than 10%
Age	Number of years since the business was started
Training	Binary variable that takes the value 1 if the firm runs a formal training program for their employees
Website	Binary variable that takes the value 1 if the firm use a website to communicate with customers or suppliers
Bank	Binary variable that takes the value 1 if the firm is financed by banks

Table 2: Characteristics of firms with and without quality certification

	Mean			
	Without certification	With certification	Difference	Observations
Exporting Firm	0.23	0.51	0.28***	10685
Direct Exporting Firm	0.18	0.43	0.25***	10791
Indirect Exporting Firm	0.08	0.14	0.06***	10682
Ln Exports (USD)	11.93	13.57	1.64***	2862
Ln Direct Exports (USD)	12.04	13.78	1.74***	2316
Ln Indirect Exports (USD)	11.15	12.00	0.85***	937
Ln Local Sales (USD)	12.88	14.24	1.36***	9591
Financial barrier	0.29	0.23	-0.07***	10701
Ln Labor Productivity	9.78	10.25	0.47***	9841
TFP	1.42	1.36	-0.06*	4950
Manager Experience	20.47	21.69	1.22***	10577
Website	0.46	0.75	0.29***	10793
Foreign	0.10	0.29	0.19***	10507
Bank	0.48	0.54	0.06***	10692
Ln Sales (USD)	13.02	14.52	1.50***	9856
Age	24.28	31.27	6.99***	10749
Training	0.45	0.81	0.35***	7550
Employees	66.93	260.45	193.52***	10798
Micro: $\leq$ 10 employees	0.26	0.10	-0.16***	10798
Small: 11 to 50 employees	0.49	0.34	-0.14***	10798
Medium: 51 to 200 employees	0.19	0.33	0.14***	10798
Big: $\geq$ 200 employees	0.06	0.22	0.16***	10798

Source: Author's estimates based on WBES and PROTEqIN.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In addition, these firms enjoy easier access to financing; only 23% of certified firms considered access to finance as a barrier, compared to 29% of non-certified firms.

In addition, certified firms have more desirable characteristics as measured by several variables. They tend to be larger, with an average of 260 employees compared to 67 employees in firms without certification. This is understandable, since firms that have adopted quality standards are more likely to fall in the medium- (between 51 and 200 employees) and large-sized (more than 200 employees) categories, whereas the non-certified group are mostly micro (less or equal than 10 employees) and small (between 11 and 50) firms. They are also older firms, have managers with more years of experience and their owners tend to be foreigners in larger proportions.

Finally, certified firms have a greater probability to run a formal training program for their employees -81% versus 45%-. They also are more likely to have a website to communicate with their customers and suppliers and finance a greater proportion of their working capital with banks, which is another signal that they have a better access to finance.

4 Identification Strategy

Our identification strategy consists in a two step difference-in-difference estimation from a balanced repeated cross-section of 4376 firms. In the first step we estimate the treatment probability, i.e. baseline probability of obtaining a quality certification, using two alternative methods: Parametric Probit and Non-Parametric Random Forest. In the second step we use the estimated propensity scores for weighting the diff-in-diff model and estimating the average treatment effect on the treated (ATT). As stated by (Wooldridge, 2007, p.1293), under regular assumptions this identification strategy has a general "double robustness"[5] property. Details of each step are presented next.

4.1 Natural Experiment Setting

To estimate the impacts of interest we set-up a natural experiment framework which allow us to use difference-in-difference strategy. For that, we define two relevant periods, before treatment ($t = 0$) and after treatment ($t = 1$), and also two relevant groups, *Treatment group* and *Control group*. For each firm, before treatment period corresponds to the baseline survey year and after treatment period corresponds to the follow-up survey year[6]. Furthermore, treatment group is comprised by those firms who had not

[5] This implies that if at least one of the two models is correctly specified, i.e. diff-in-diff or propensity score, then the estimated ATT is consistent. This is partially true for our strategy because in order to avoid confusion between treatment effects and treatment determinants, time varying covariates were not included in the propensity score model while for construction all baseline characteristics are included into the fixed-effects DID.

[6] Each firm have only one observation by period but baseline survey was collected in year 2003 for Brazil, 2006 for remaining LA countries and 2011 for Caribbean. Follow-up survey for Brazil, LA and Caribbean were respectively conducted in 2009, 2010 & 2014.

a *Quality Certification* in $t = 0$ but obtained one before $t = 1$. We can see in table 4 that 433 firms conform the treatment group. Also, control group is comprised by those firms who had not a *Quality Certification* in $t = 0$ and remain the same in $t = 1$. We can see in table 4 that 3943 firms conform the control group while, as displayed in table 3, 1034 firms who already had a *Quality Certification* in $t = 1$ are omitted. This reduce our panel from 5410 firms to 4376 firms.

Table 3: Quality Certification

Has Quality Certification	Baseline (t=0)	Follow-Up (t=1)	Total
No	4376	4158	8534
Yes	1034	1252	2286
Total	5410	5410	10820

Note:Balanced Panel of 5410 firms

Table 4: Treatment Status

Treated	Before (t=0)	After (t=1)	Total
No	4376	3943	8319
Yes	0	433	433
Total	4376	4376	8752

Note:Balanced Panel of 4376 firms

To asses the impact of obtaining a *Quality Certification* on firm performance we use a weighted differences-in-differences (Dif-in-Dif) model. The resulting model is presented in equation 1 :

$$y_{icst} = \beta T_i + f_i + \delta_t + c_c * \delta_t + s_s * \delta_t + \gamma X_{it} + \varepsilon_{icst} \quad ; \quad \lambda_i^m \tag{1}$$

The coefficient β represents the effect of adopting the quality certification on any outcome. The variable T takes the value 0 if the firm lacked certification and continues without it, or if it had not acquired a certification yet; it takes the value 1 if the firm was not certified in the previous period, but is now. To endow the model with a Diff-in-Diff structure we use firm-level (f_i) and time (δ_t) fixed-effects, and to avoid bias coming from country and sector specific trends we also include Country-Year ($c_c * \delta_t$) and Sector-Year($s_s * \delta_t$) specific trends. Finally, we include a set of firm specific time-varying covariates in X_{it}. Since firms in the panel are surveyed in 2 periods, number of observations will double number of firms in all estimations.

The Diff-in-Diff method control for both observable and unobservable heterogeneity between firms that is constant over time (e.g. firm's sector, location, and other other firm intrinsic characteristics) which enables a significant reduction of the estimation bias. Nevertheless, the method has some limitations and its causal interpretation relies on treated and control groups satisfying the parallel trend assumption. This assumption may be implausible if pre-treatment characteristics that are thought to be associated with the dynamics of the outcome variable are unbalanced between the treated and the untreated. If that is the case, and pre-treatment characteristics of treatment and control groups are significantly different, then even in the absence of treatment they would perform differently over time. That scenario would violate the parallel trend assumption. That is why, in order to achieve pre-treatment balanced groups, we adjust the Diff-in-Diff using the weights λ_i^m. Given that models implemented to define weights are feasible of misspecification leading to biased conclusions, we present results from two alternative methods m. General approach and specific methods are detailed next.

4.2 Propensity Score

Weights λ_i^m are defined using Inverse Probability Weighting, a strategy from the family of propensity score (PS) methods commonly used to minimize selection bias in non-experimental studies (Austin, 2011). PS was first introduced by Rosenbaum & Rubin (1983) as a way for "*balancing*" treatment and control groups on a set of baseline characteristics; i.e., to make the groups as similar as possible with respect to those observed baseline characteristics. The PS itself is defined as the conditional probability of receiving the treatment of interest as a function of those covariates:

$$Pr(T_i = 1|X_i = x_i) = p_T(x) \tag{2}$$

Both, the function $p_T(.)$ mapping from covariates to treatment, as well as the conditional probability of treatment assignment $Pr(T_i = 1|X_i = x_i)$, are essentially unknown in natural experiments and different methods were used for their estimation. Most common methods are logit and probit (Imbens & Wooldridge, 2009). However, there is evidence that slight misspecification of parametric propensity score models can result in substantial bias of estimated treatment effects (Drake, 1993; A. Smith & E. Todd, 2005; King & Nielsen, 2019). In order to address this issue semi-parametric and non-parametric techniques can be employed for estimating it (Mccaffrey *et al.*, 2005; Li *et al.*, 2008; Imai & Ratkovic, 2014; Busso *et al.*, 2014). Among non-parametric methods, machine learning algorithms such as CART, Support Vector Machines and Random Forest are a promising alternative for PS estimation (Westreich *et al.*, 2010; Imbens & Rubin, 2015). Main reason for this is that ML algorithms usually achieve higher classification accuracy requiring fewer assumptions about functional forms and tuning parameters, specially in the presence of non-linearity and non-additivity of confounders (Lee *et al.*, 2010). Despite of the different possible approaches, if the PS is correctly estimated and the assumption of selection in observable holds, we can be sure that the true treatment effect will be estimated without bias. We employ two of the most popular among parametric and non-parametric methods for estimating PS, i.e. Probit and Random Forest. Methods and results are detailed next.

4.2.1 Probit Propensity Score

Treatment probability of firms conditional on observed variables is estimated using a probit probability model represented by Equation 3[7].

$$T_i = X_{i0}\beta + \varepsilon_{i0} \tag{3}$$

Where T_i takes the value of 1 if firms belongs to treatment group and 0 if belongs to control group, X_{i0} represents a set of characteristics of the firm i in the period $t = 0$[8] and $\varepsilon_{i0} \sim N(0, 1)$. This model allow us to compare characteristics between treatment and control groups before treatment occurred and consequently estimate the propensity score for each firm.

Despite we define our model specification not for predicting treatment status, but for achieving baseline covariates balance[9], results (showed in Table 5) of the model are still useful guides to understand which firms' characteristics are more important predictors of treatment status. Our findings point in the direction anticipated by the theory and the existing evidence. Firms with larger sales volume, have higher probability of obtaining an internationally-recognized quality certification. Larger firms may find it necessary to standardize processes in order to optimize performance. Direct Exporting firms also have higher probability of certifying than those that operate in the domestic market alone and the effect increase with export volume. This could be due to higher information asymmetries with foreign than local customers and harder competition in international markets. Meanwhile, foreign owned firms are 22% more likely to adopt quality certifications. This is typically because they enjoy easier access to financial resources and have superior management skills[10]. Finally, years of manager experience are significant predictors of quality certification adoption.

Figure 10 is the density of predicted propensity scores for treatment and control groups. As expected, control group distribution is placed to the left of treatment group with density peaks near 0.1, whereas treatment group distribution is flatter and less asymmetrical but still with peaks near 0.2. Given that the true propensity score is essentially unknown and that small misspecifications of the probit model can strongly

[7] More precisely, we suppose that propensity score takes the form $Pr(T_i = 1|X_i = x_i) = Pr(\hat{T}_i > 0) = Pr(X_i\beta + \varepsilon_i > 0) = \Phi(\beta X_i)$, where Φ is the Normal distribution. This method is called parametric because thanks to this assumption, we can reduce the problem of estimating the propensity score to that of estimating parameter β

[8] To avoid confusion between treatment outcomes and determinants we use only pre-treatment firms' characteristics observed in the baseline survey.

[9] This is an important aspect because our metric for including or not covariates of different order into the model was not the r^2, their predictive power, or even the significance of parameters, but the degree of baseline covariate imbalance between treated and control groups. Log Direct Exports was transformed using the *Inverse Hiperbolic Sin Transformation* to evaluate the extensive margin of exports without losing those observations from non-exporters

[10] However, in this model we are already controlling for years of manager experience. Another channel, as showed by Pekovic (2010), is that foreign owned firms have also more internal pressure to adopt international quality standards (e.g. requirement of the headquarters when the firm is inside an international group).

bias causal inference (Drake, 1993; Dehejia & Wahba, 2000), in order to evaluate the robustness of our estimations we also use an alternative non-parametric estimator of the propensity score using random forest.

Table 5: Probit Propensity Score

	(1) Prob. Treated
Log Total Sales	0.071**
	(0.031)
Log Employees	0.031
	(0.043)
Age	0.005
	(0.004)
Age^2	-0.000
	(0.000)
Direct Exporting Firm	-0.638
	(0.419)
Log Exports	0.063**
	(0.032)
Manager Experience	0.007**
	(0.003)
Bank	0.066
	(0.065)
Website	0.116
	(0.073)
Foreign	0.222**
	(0.103)
Observations	3,676

Source: Author's estimates based on WBES and PROTEqIN.
Dummy by Country and Sector included
Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

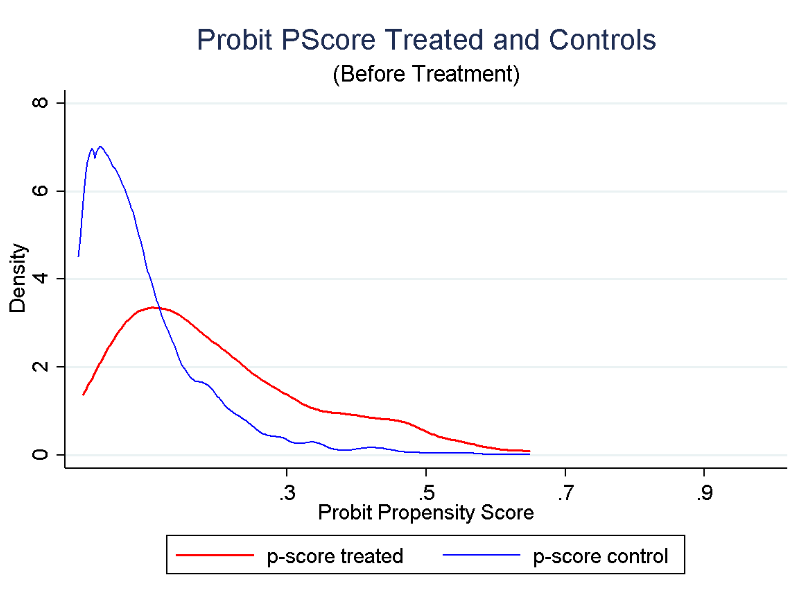

Figure 1: Probit Propensity Score

4.2.2 Random Forest Propensity Score

Random Forests (RF), firstly proposed for Breiman (2001), is one of the most successful machine (statistical) learning algorithms for practical applications[11] with growing use in econometrics and applied economics literature(Imbens & Wooldridge, 2009; Varian, 2014; Duflo *et al.*, 2017; Athey *et al.*, 2019). This algorithm is generally recognized for *"...its accuracy and its ability to deal with small sample sizes and high-dimensional feature spaces"* (Biau & Scornet, 2016, p.1). RF is a generic name for the process of *(i)*aggregating *(ii)*random *(iii)*decision-trees, its three main features:

i Aggregating feature comes from the process of *Bagging*, a contraction of bootstrap-aggreggating. It consists on drawing B bootstrap samples of size N from the data, constructing a decision-tree for each sample, and averaging results over the B samples. We took $B = 1000$ samples of size $N = 748$ without replacement. Sample size was defined for achieving, in average, size-balanced treated and control groups [12].

ii Randomness arise from the sequential selection of L regressors out of K possible covariates. This is a key aspect of RF accuracy because exogenous variation is

[11] For a detailed explanation of the algorithm functioning and recent developments see (Efron & Hastie, 2016, p.325) and Biau & Scornet (2016)

[12] Imbalanced groups in the classification dimension is a serious issue in this kind of algorithms because simple decision rules overestimates the probabilities of belonging to the majority group. In our database, after eliminating observations with missing covariates, we have a treatment group of size $N_T = 374$ and control group of size $N_C = 3302$. For achieving size-balance we set a vector assigning sampling-probabilities of $(\frac{1}{N_T}, \frac{1}{N_C})$ for treated and control groups respectively.

introduced into the classification process. We employed the same set of $K = 12$ covariates used in Probit model and followed the empirical rule of setting $L = \frac{K}{3} = 4$[13]

iii Decision-trees are the basic rule of classification employed for RF. In a general sense, it consists on repeated binary splits of the sample according to some criteria. Our main objective is to classify two groups into our two states of interest $T = 0/T = 1$. This classification is made at each decision node d conditional on the support of a single covariate chosen among the L randomly selected regressors. This covariate is selected according to its relative accuracy for splitting treated from controls, i.e. its relative predictive power into the subset of the L regressors. More precisely, in RF the criteria used for sample-splitting is CART[14], a non-parametric regression which is estimated in order to minimize the mean square error of node splitting d 4:

$$\begin{aligned} MSE(d) &= MSE_{\hat{C}} + MSE_{\hat{T}} \\ &= \sum_{i \in \hat{T}=0} (y_i - \overline{y}_{\hat{T}=0})^2 + \sum_{i \in \hat{T}=1} (y_i - \overline{y}_{\hat{T}=1})^2 \end{aligned} \tag{4}$$

Where $MSE(d)$ is the mean square error resulting from the splitting d, which is the sum of $MSE_{\hat{C}}$ and $MSE_{\hat{T}}$, i.e. the respective error resulting from control and treatment group classification. The outcome of each decision node d are two estimated groups ($\bar{T} = 0$ and $\bar{T} = 1$) split over the optimal cut-point value[15] of the preferred regressor among the L sampled. This process is repeated until no improvement on classification can be achieved or until the user-defined minimum terminal node-size (N_{min}) is achieved. In our case we set $N_{min} = 6$, the default recommendation for regression tress. An example of two sample decision trees are presented in figures 2 and 3. Figure 2 indicates that from the $L = 4$ randomly chosen regressors[16], the best classification performance was achieved by *loglabor* which is the natural logarithm of the number of employees. Big firms, those with *loglabor* > 3.4, were classified as belonging to treatment group, while small firms were classified as control. Among big firms, those with greater sales *logsales* $>$

[13] This rule was complemented by cross-validation using covariates balancing as a measure of algorithm performance. Regressors' sampling-probability were defined using the same rule. As expected after running the Probit model, more weight were required for those covariates with greater imbalance and treatment predictive power such as Sales, Exports Volume and Foreign ownership.

[14] Acronym for clasiffication and Regression Trees

[15] For continuous regressors, the algorithm creates a discrete grid, evalutes the resulting MSE at each possible cut-point, and finally chose that which minimize the MSE. We use the default grid of 10 possible cut-points for each splitting.

[16] From the plotted CART we can't know which were the contesting regressors at each node but only which are the winners. Also, the subset of regressors at each node is taken with replacement allowing the inclusion of the same covariate more than once in a single tree

14.547 were assigned to node 9 and classified as treated. Over the 124 firms in node 9, accuracy of the classification was 0.661 and consequently the error rate was 0.339. Another interpretation, more useful for us, is that conditional treatment probability of big firms with greater sales volumes is 0.66. Similarly, from figure 3, we can infer that non-foreign owned firms, from sectors $\{2, 15, 24...\}$, which are non-exporters, have probability 0.61 of being into the control group.

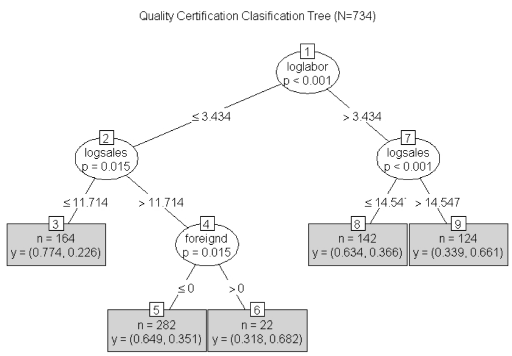

Figure 2: Sample Classification Tree using Continuous Predictors

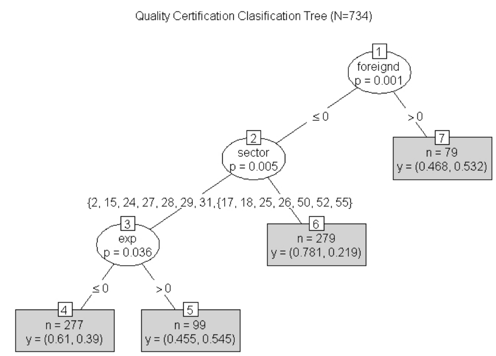

Figure 3: Sample Classification Tree using Categorical Predictors

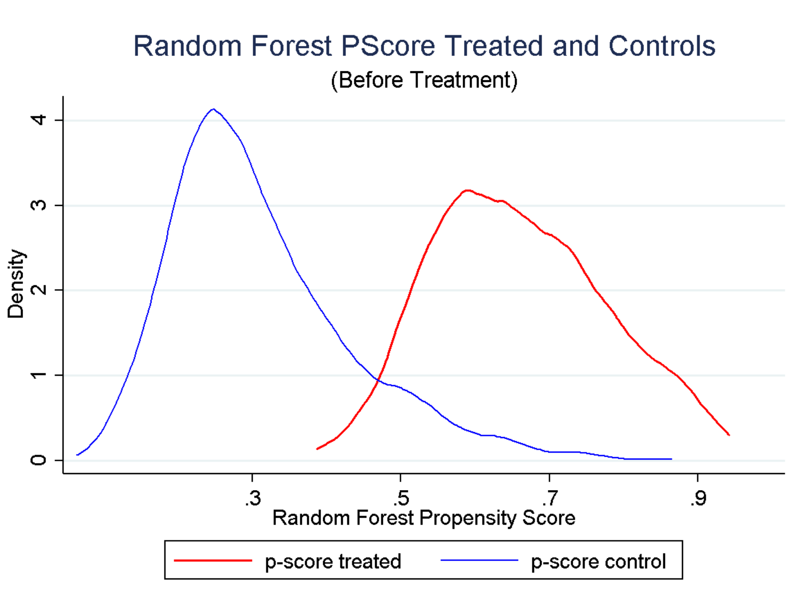

Figure 4: Random Forest Propensity Score

The resulting PS estimation via RF is presented on figure 4. Due to the superior classification power of RF over Probit, we can realize that RF-PS density of treated and control groups has fewer overlapping over the support than those observed in figure 3. Because we don't follow a local PS-Matching strategy but a complete sample re-weighting of a perfectly balanced panel, while covariate balancing between treated and control group is satisfied, reduced overlapping is not a treat for our identification strategy[17]. On the contrary, given the marked differences of methodology and results between the two competing strategies we propose for PS estimation, similar results on the estimated treatment effects will be an important proof of robustness.

4.2.3 Inverse Probability Weighting and Balancing results

Inverse Probability weighting (IPW) has a long tradition in statistics and was firstly employed in econometrics for consistent estimation under non-random attrition and censored data (Horvitz & Thompson, 1952; Rubin, 1976; Robins *et al.*, 1995; Robins & Rotnitzky, 1995). In the same line, Rotnitzky & Robins (1995) show that weighting observations by the inverse of a parametric estimate of the selection probability is more efficient than weighting by the inverse of the true selection probability. This result was extended by Wooldridge (2007) for estimations of probabilities using non-parametric methods and stating the "double robustness" (M. Robins & Ritov, 1997) property of IWP estimation. This property implies that when covariates affecting the selection and the outcomes of interest are the same, if either the estimated propensity score or the unweighted model are correctly especified, then the parameters of interest can be consistently identified[18].

The problem of selection on the treatment is a special case o selection problems described before. Rosenbaum (1987) and Hirano *et al.* (2003) studied the properties of IWP for estimation of treatment effects respectively proving its consistency and efficiency. Lee *et al.* (2010) conducted montecarlo simulations evaluating the performance of IPW using parametric and machine learning methods and discovered that random forest was among the best performers in reducing estimation bias of treatment effects[19]. Furthermore, they find that covariate balancing was one the best metrics predicting bias reduction on treatment estimations. We use IPW for estimating the average treatment effect on the treated (ATT). ATT requires weighting the control group using the odds-ratio of the estimated propensity scores and leaving the treatment group unchanged as

[17] We estimate the ATT (average treatment effect on the treated) which according to Imbens (2004) requires two weaker conditions for identification: i) Unconfoundness for control group $Y_0 \perp Pr(T_0 = 1|X)$, and ii) Weak-overlapping $Pr(T_0 = 1|X) < 1$

[18] Wooldridge (2007) points that IWP is not free of risks and indicates caution for two scenarios where weighting can introduce bias to the basic model. First scenario is a bad specification of the PS model. To avoid this possibility we propose two alternative methods for this estimation and provide tests of baseline covariates balance. The second scenario, more dangerous and hard to detect, may appear when variables affecting the outcome of interest can not be included on the propensity score estimation. This omission may introduce additional bias even if the propensity score is correctly estimated. This is not a treat for our strategy because all the baseline covariates used for PS estimation are also fixed effects in the DID model.

[19] This was specially true por data generating process with non-additivity and non-linearity. For DGP with linear specification parametric methods performed equally well.

showed in equation 5:

$$\lambda_i^m = \begin{cases} 1 & \text{if } T = 1 \\ \frac{\hat{p}_t(X_i)^m}{1-\hat{p}_t(X_i)^m} & \text{if } T = 0. \end{cases} \tag{5}$$

Results of baseline covariate mean balance before and after weighting with models $m = \{Probit; RF\}$ are presented in Table 6 and Figure 5[20]. Table 6 indicates that unweighted baseline differences between treatment and control group are statistically significant according to independent t-test for most of the covariates. This is reflected also on the multivariate F-test which reject the null of similar groups. More specifically, treated group is composed by firms: bigger (both in terms of sales and employees), mostly exporters (both in terms of exporting probability and volume), more productive, foreign owned, and with greater probability of having a website. However, after IPW all significant differences between groups disappears for both methods employed and RF slightly over-performs Probit balance. Some authors remarked the importance of evaluating covariates imbalance using statistics that are not affected by sample size (Ho *et al.*, 2007; Imai *et al.*, 2008; Austin, 2009) and the use of standardized mean differences (smd) became popular among propensity score literature. Main trouble with this measure is that don't provide a clear cut-off point to assess what means for two groups to be imbalanced. A rule of thumb, also popular among practitioners, is the threshold of 0.1 absoulte smd. Figure 5 displays graphically another perspective regarding baseline because thanks to standardization we can say that, before weighting, baseline imbalance was relative greater in Sales, Employees and Labor Productivity. Also, according to this criteria difference in age between groups was also significant. Another interesting feature is that despite the difference between estimated propensity scores, RF IPW produce similar balancing than Probit IPW, both performing well in achieving smd far below the 0.1 threshold[21].

[20] Due to space limitations, balance for country and sector are not presented. Both were included into both models as important covariates and a correct balance was achieved after weighting

[21] We can also visualize that RF forest adjusting is below 0.05 smd for most of the covariates, performing better than Probit for covariates with greater previous imbalance and worse for those with fewer previous imbalance. This is due to the flexibility of non-parametric methods for managing trade offs between dimensions.

Table 6: Adjusted Baseline Covariate Mean Balance

	Unweighted			Probit		Random Forest	
	Treated	Control	Diff	Control	Diff	Control	Diff
Ln Employees	3.551	3.143	0.408***	3.526	0.064	3.539	0.051
Ln Sales	13.659	12.811	0.848***	13.508	0.146	13.532	0.122
Ln Local Sales	13.193	12.527	0.666***	13.047	0.139	13.086	0.100
Direct Exporter	0.246	0.173	0.073***	0.249	0.021	0.249	0.021
Exporter	0.292	0.223	0.070***	0.299	0.008	0.301	0.007
Indirect Exporter	0.086	0.080	0.006	0.085	-0.005	0.088	-0.008
Ln Direct Exp.	3.338	2.085	1.252***	3.221	0.277	3.205	0.294
Ln Exports	3.846	2.672	1.174***	3.819	0.081	3.819	0.081
Ln Indirect Exp.	0.968	0.927	0.042	1.008	-0.118	1.032	-0.142
Finance	0.318	0.321	-0.004	0.282	0.042	0.305	0.020
M. Experience	18.451	19.044	-0.593	18.396	-0.014	18.913	-0.530
Age	23.587	22.215	1.372	23.941	0.466	24.276	0.130
Website	0.491	0.369	0.121***	0.461	0.017	0.443	0.036
Foreign	0.138	0.082	0.056***	0.138	-0.001	0.119	0.018
Bank	0.495	0.457	0.038	0.524	0.002	0.508	0.018
Ln Labor Prod.	10.087	9.661	0.426***	9.982	0.082	9.994	0.071
TFP	1.540	1.440	0.099	1.403	0.122	1.422	0.103
Multiv. F Test			5.51***		2.10		1.32

Source: Author's estimates based on WBES and PROTEqIN.
Balanced groups by country and sector also achieved after weighting.
Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

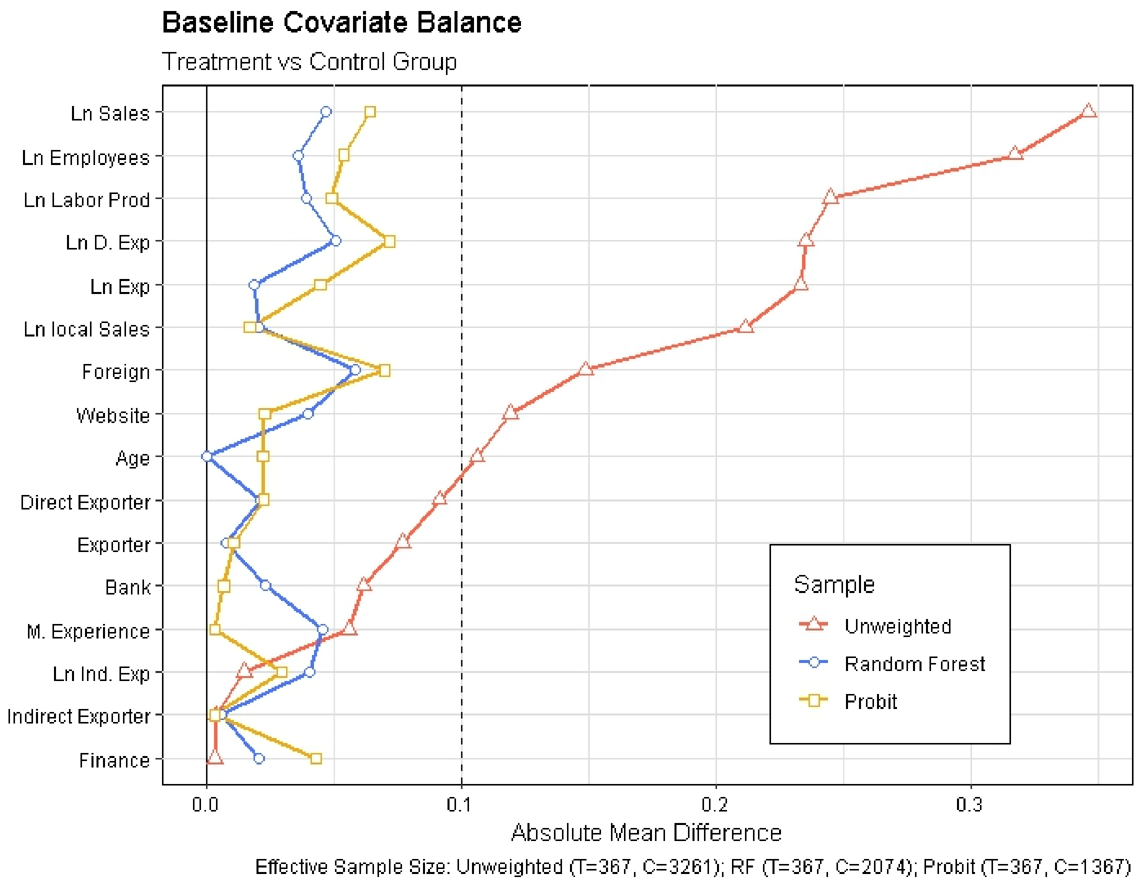

Figure 5: After weighting covariate balancing

Proven the effectiveness of the estimated weights λ_i^m in achieving baseline balance between groups , we can proceed to the second step of our identification strategy consisting on estimating the weighted differences-in-difference model presented in equation 1. We can state that the resulting ATT will be:

$$A\hat{T}T = \hat{\beta} = \frac{1}{\lambda_i^m N} \sum_{i=1}^{N} \lambda_i^m (\hat{Y}_{i1} - \hat{Y}_{i0}) \tag{6}$$

Where $\hat{Y}_{i1}, \hat{Y}_{i0}$ are the respective conditional expectations for the first and second period resulting from the unweighted DID model presented in equation 1. This identification strategy is in the spirit of Abadie (2005) and same assumptions for equating $A\hat{T}T = \hat{\beta}$ are required:

1. *Unconfoundedness for Control Trends*: That is, conditional on observed variables, treatment status is not related to trends in controls' outcomes.

$$E[Y_1(0) - Y_0(0)|X, T] = E[Y_1(0) - Y_0(0)|X] \tag{7}$$

2. *Weak Overlap*: This is a weaker assumption than the required for ATE estimation, needed for propensity score identification.

$$Pr(T = 1|X) < 1 \tag{8}$$

Results of the estimated ATT of firms' *Quality Certification* adoption on Local Sales, Export behavior, Finance restriction and Productivity are presented next.

5 Impacts on Firm Performance

The analysis of the results is divided into four parts. The first three subsections analyses the impacts on different performance variables in order to explore actors to whom the possession of an international certification -i.e. signaling certain quality standards- may matter: domestic customers, foreign customers, and financial entities. Furthermore, because firms use new productive processes and management control systems and implement a staff training agenda during the certification process, increases in productivity may be expected and are analyzed in last subsection.

5.1 Local Sales

Firms that seek to obtain benefits by signaling good business practices and quality control standards to their customers in the domestic market should have an increase in their local sales or national sales. Therefore, we estimate the Equation 6 using local sales of the firms as outcome variable. Results are shown in Table 7.

Table 7: Effect of internationally-recognized quality certification on local sales

	(1) Ln local Sales Probit	(2) Ln local Sales RF
Treated	0.136	0.173*
	[0.089]	[0.089]
Observations	6,328	6,328
Firms	3164	3164
Control	2846	2846
Treated	318	318

Source: Author's estimates based on WBES and PROTEqIN using IPW Diff-in-Diff.
Time varying controls for Foreign ownership, Manager Experience and own Website.
Firm-level fixed effects and Country and Sector specific trends included.
Clustered standard errors at firm-level in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Adopting an internationally-recognized quality certification seems to have weak rise in firm's local sales. Estimated $\hat{ATT}$ is positive for both methods but significant only for RF. There are probably two main reasons for this finding. First, local customers (specially in developing counties as Latin American and the Caribbean) may pay little or no attention to the firm's business practices and quality controls. Second, since local customers have closer dealings with the firm, they have alternative ways to judge quality for themselves without requiring a certification. For example, they could interact with other firm's customers or they could visit firm's factory.

5.2 Export behavior

Firms may seek to signal their quality to customers in international markets. In this case, the impact could be reflected by the firm's entry into foreign markets, either by starting in exporting (non-exporting firms) or by expanding exports (already exporting firms). Hence, we estimate the Equation 6 using export condition and the amount of exports. Results for the extensive and intensive margins of exports are shown in Tables 8 and 9.

Table 8: Effect of internationally-recognized quality certification on Export Probability

	(1) Export Probit	(2) Export RF	(3) Dir. Exp. Probit	(4) Dir. Exp. RF	(5) Ind. Exp. Probit	(6) Ind. Exp. RF
Treated	0.074***	0.063**	0.023	0.011	0.050**	0.042*
	[0.024]	[0.025]	[0.022]	[0.023]	[0.022]	[0.022]
Observations	6,984	6,984	7,088	7,088	6,984	6,984
Firms	3492	3492	3544	3544	3492	3492
Control	3137	3137	3186	3186	3137	3137
Treated	355	355	358	358	355	355

Source: Author's estimates based on WBES and PROTEqIN using IPW Diff-in-Diff.
Time varying controls for Foreign ownership, Manager Experience and own Website.
Firm-level fixed effects and Country and Sector specific trends included.
Clustered standard errors at firm-level in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Effect of internationally-recognized quality certification on Exports Volume

	(1) Ln Exp Probit	(2) Ln Exp RF	(3) Ln D. Exp Probit	(4) Ln D. Exp RF	(5) Ln I. Exp Probit	(6) Ln I. Exp RF
Treated	0.458*	0.441*	0.694***	0.665***	-1.905	-0.906
	[0.266]	[0.265]	[0.237]	[0.250]	[1.219]	[1.451]
Observations	1,244	1,244	940	940	326	326
Firms	622	622	470	470	163	163
Control	543	543	405	405	155	155
Treated	79	79	65	65	8	8

Source: Author's estimates based on WBES and PROTEqIN using IPW Diff-in-Diff.
Time varying controls for Foreign ownership, Manager Experience and own Website.
Firm-level fixed effects and Country and Sector specific trends included.
Clustered standard errors at firm-level in brackets. *** p<0.01, ** p<0.05, * p<0.1

Quality certification is found to increase both the extensive and intensive margins of trade. More specifically, adopting an internationally recognized quality certification increases the likelihood of entering into foreign markets between 6.3 and 7.4 percentage points. Furthermore, firms that were already exporting increase their exports by approximately 44 percent. The increase in the amount exported might have been achieved either by exporting higher amounts to the same destinations or by entering new markets or introducing new products. For firms already exporting, entering new markets would be an argument in favor of certifications as an instrument to remove informational barriers that prevented them from demonstrating the quality of their products and process. This is the mechanism proposed by Volpe Martincus *et al.* (2010). Unfortunately, our dataset don't have information about export destinations or products. However, thanks to the decomposition between indirect and direct exports, we can propose an alternative mechanism explaining increases in intensive and extensive margins.

Our findings indicate that quality certifications don't affect extensive margins of direct exports, but allows firms' introduction into the indirect exports market. Also, for those firms already exporting before obtaining a quality certification, only direct exports volume is increased due to certification adoption.

This evidence is consistent with the internationalization process proposed in the Global Value Chain literature (Gereffi *et al.*, 2001; Pietrobelli & Rabellotti, 2010; Hernández *et al.*, 2014; Gereffi & Fernandez-Stark, 2016). Obtaining a quality certification would promote indirect internationalization of local firms by reducing informational barriers among already-international firms operating in the local markets. However, this signaling effect coming from quality certification may be not enough to avoid other barriers for direct exports, such as language, paperwork, invoicing, or sales management (Crespi *et al.*, 2014b) affecting the capacity of firms of contacting customers abroad, identifying business opportunities in foreign markets, and learning about distribution channels for their products and bureaucratic procedures (Leonidou, 2004). This may be the reasons why obtaining a quality-certification promote only an indirect path to international markets.

The fact that only direct exports volume increase is consistent with the hypothesis of the existence of formal barriers impeding international access to firms' which are not certified. For firms already exporting directly, costs of entering into new international markets are smaller than those faced for local firms' to becoming direct exporters. Obtaining a quality certification may be helping those firms' to overcome formal barriers for introducing new highly-regulated products[22] into the regions they already operate, or for accessing to new regions with stronger regulation policies for the same products they already trade.

5.3 Finance restriction

We observe the impact of obtaining quality certification on access to financing by using the own perception of the firm on how access to finance represent an obstacle to grow. In the short term, it is expected that firms that achieve a quality certification manage to improve their financing due to an easing of credit restrictions. This might be due to the fact that certification is a criterion in credit institutions evaluation of a firm's creditworthiness, since it is associated with better future performance. We analyze this potential effect in Table 10.

[22] This hypothesis is also consistent with learning-by-exporting theory (Silva *et al.*, 2012) predicting an increase of sophistication in the production after entering international markets. Learning-by-exporting may be operating also among firms which start as indirect exporters and become direct exporters later. Our dataset indicates that 18% of firms exclusively exporting indirectly in the first period become direct exporters in the second. Also 35% stop exporting at all.

Table 10: Effect of internationally-recognized quality certification on perceived Finance restriction

	(1) Finance Probit	(2) Finance RF
Treated	-0.071**	-0.062**
	[0.031]	[0.032]
Observations	6,980	6,980
Firms	3490	3490
Control	3141	3141
Treated	349	349

Source: Author's estimates based on WBES and PROTEqIN using IPW Diff-in-Diff.
Time varying controls for Foreign ownership, Manager Experience and own Website.
Firm-level fixed effects and Country and Sector specific trends included.
Clustered standard errors at firm-level in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Adopting an internationally-recognized quality certification leads to a reduction between 6 and 7 percentage points in the likelihood that firms consider access to credit as a barrier to their growth. Improving access to credit could be an important source of growth for firms because it enables them to deploy long-term investments, for example, in projects to improve (or expand) productive infrastructure or increase R&D spending leading to product and process innovations (Brito & Mello, 1995; Schiavo & Musso, 2008). Moreover, this positive finding regarding access to credit can, at the same time, enhance the effects on exports due to the importance of pre- and post-export financing (Bellone *et al.*, 2010).

5.4 Effects on Productivity

The effects analyzed up to this point have been mainly referred to reduced informational barriers with certain agents external to the firm. Another hypothesis could be that firms adopting quality certification become more productive due to the implementation of internationally standardized process. Hence, we estimate Equation 1 for labor productivity and TFP. Table 11 presents the results.

Table 11: Effect of internationally-recognized quality certification on productivity

	(1) Labor Prod Probit	(2) Labor Prod RF	(3) TFP Probit	(4) TFP RF
Treated	0.002	0.043	-0.089	-0.019
	[0.079]	[0.081]	[0.114]	[0.121]
Observations	6,516	6,516	1,890	1,890
Firms	3258	3258	945	945
Control	2927	2927	841	841
Treated	331	331	104	104

Source: Author's estimates based on WBES and PROTEqIN using IPW Diff-in-Diff.
Time varying controls for Foreign ownership, Manager Experience and own Website.
Firm-level fixed effects and Country and Sector specific trends included.
Clustered standard errors at firm-level in brackets.
*** p<0.01, ** p<0.05, * p<0.1

We find no statistically significant effects of acquiring a quality certification on either labor productivity or TFP. However, certain aspects of these estimates are worth highlighting. Since questionnaire question on raw materials has a lower response rate, the number of observations for these estimates is a third of the total, which generates higher variance of the parameters and hampers identification of impacts. This is particularly in the case for the TFP measure, which also suffers the low rate of response to the questions about capital.

Overall, we find statistically significant positive effects on local sales, export performance (both for firms that previously exported and non-exporters) and on access to credit, but not or firm productivity (both labor productivity and TFP). However, effects on local sales are only significant at 90% under one of the specifications.Positive effects seems to be stronger on sales to foreign customers indicating that the information asymmetries might be different with regard to each actor. The higher costs foreign customers faced in seeking out and identifying firms with desirable characteristics and signing contracts and verifying their compliance imply that the firm faces bigger obstacles in its commercial relationship with the former. In this case, obtaining a certification is more important for revealing information about the existence of such desirable characteristics. An alternative explanation could be that firms that acquire internationally-recognized quality certification can improve their products and, thus, export more. But if this hypothesis were true, it would also generate an increase in its local sales, which is not so clear.

Moreover, we find effects on short-term outcomes but not on long-term outcomes, such as productivity. There are at least two possible explanations why no statistically significant effect on the latter was found. Obtaining an internationally-recognized quality certification may not, in fact, lead to increased productivity, the improvement may be too weak to be captured by the estimate, or need more time to materialize. This argument reinforces the idea that the main benefit of certification is the signaling of desirable firm's characteristics for certain agents, especially to foreign customers and credit agencies.

However, given the positive effect of certification on a firm's export potential, it may set in motion a learning-by-exporting process, in which the firm can improve its productivity, albeit indirectly, based on the knowledge acquired after starting to export, or through access to credit. Firms may decide to install better machinery, provide training for its employees, increase expenditure on R&D, and so forth. That is, it is expected that the increase in productivity is more an indirect consequence of the export behavior and access to credit due to certification adoption than a direct effect of this.

Therefore, the temporal dynamic of the impact need to be taken into account. In our database, the time gap between one period and another is only four years. However, certification can occur at any intermediate point, including immediately after responding to the first questionnaire or immediately before responding to the second. It is therefore possible that the effect on firm productivity is not registered because it requires a longer window of time to materialize. Since our time-lapse is a short-run period, our results are consistent with (Javorcik & Sawada, 2018) in which productivity effects are not seen until the third year of certification.

6 Falsification tests

For evaluating the robustness of our findings, we propose two alternative experiments imitating our identification strategy but using different firms falsely classified as treatment group. The purpose of these experiments is to try to falsify our previous findings. The impossibility of reproducing the estimated effects with these methodologies will be evidence favoring our conclusions[23].

Experiment A consists on creating a false treatment group conformed by those firms who didn't had a Quality Certification in the first period, but in the follow up survey declared having a Quality Certification in process (not awarded yet). Results provided in Table 13 confirms that none of our significant previous findings can be reproduced using in-process certification as treatment indicator. We believe this is an almost ideal group for mimic treatment because those firms also self-selected into a certification process. Finding no significant effects provides further evidence in favor of the causal effect of effectively certifying by international quality standards. However, it can be argued that non-significant coefficients are due to the relatively smaller number of firms with quality certification in process. For this reason, we propose a second falsification

[23] We present results only for Random Forest methodology and for previous significant effect. However, we checked that results can not be falsely reproduced for any method or outcome

test.

Experiment B consists on selecting into the false treatment group firms who didn't receive a quality certification, but are as similar as can we know to firms who actually received the treatment. For that, we firstly matched 1-to-1 quality certified firms to false treatment group using Genetic Matching algorithm (Diamond & Sekhon, 2013) [24] and later repeated our identification strategy. Table 13 confirms that no significant effects can be falsely reproduced using similar firms providing further support on the validity of our previous findings.

Table 12: Falsification test A. Effect of In-Process Quality Certification on Firm Performance

	(1) Ln Loc. Sales RF	(2) Export RF	(3) Ind. Export RF	(4) Ln Exp. RF	(5) Ln. D. Exp RF	(6) Finance RF
In-Process	0.186	-0.030	-0.047	0.221	0.437	-0.098
Quality Cert.	[0.123]	[0.059]	[0.052]	[0.369]	[0.414]	[0.068]
Observations	6,324	6,980	6,978	1,244	942	6,978
Firms	3162	3490	3489	622	471	3489
Control	3095	3418	3417	607	459	3418
Falsely Treated	67	72	72	15	12	71

Source: Author's estimates based on WBES and PROTEqIN using Random Forest IPW DID.
Time varying controls for Foreign ownership, Manager Experience and own Website.
Firm-level fixed effects and Country and Sector specific trends included.
Clustered standard errors at firm-level in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 13: Falsification test B. 1-to-1 Genetic Matching

	(1) Ln Loc. Sales RF	(2) Export RF	(3) Ind. Export RF	(4) Ln Exp. RF	(5) Ln. D. Exp RF	(6) Finance RF
1-to-1 Genetic	-0.045	0.012	-0.004	-0.249	-0.241	-0.043
Match	[0.068]	[0.021]	[0.020]	[0.178]	[0.186]	[0.031]
Observations	6,324	6,980	6,978	1,244	942	6,978
Firms	3162	3490	3489	622	471	3489
Control	2855	3157	3156	551	413	3147
Falsely Treated	307	333	333	71	58	342

Source: Author's estimates based on WBES and PROTEqIN using Random Forest IPW DID.
Time varying controls for Foreign ownership, Manager Experience and own Website.
Firm-level fixed effects and Country and Sector specific trends included.
Clustered standard errors at firm-level in brackets. *** p<0.01, ** p<0.05, * p<0.1

[24] Aware of the recent findings of King & Nielsen (2019) indicating the risks of matching over the propensity score, we used genetic matching proposed for Diamond & Sekhon (2013). We forced an exact match on binary dimensions, and asked the algorithm to iteratively reduce mahalanobis distance of continuous covariates until no extra gains can be achieved

7 Conclusion

This study presents empirical evidence at the firm level for the LAC region on the determinants of adopting an internationally-recognized quality certification and their effects on firm performance using the WBES and PROTEqIN, surveys from 32 countries in the LAC region. The results indicate that exporting firms, foreign firms, and firms with higher sales volume have the highest ex-ante probabilities of obtaining a quality certification.

We then find that obtaining a quality certification produce a positive effect by signaling desirable characteristics to firm's external agents for whom the relationship has more information asymmetries. Thus, firms that obtained this certification achieved to start exporting indirectly. Furthermore, firms already exporting directly increased their volume of direct exports.

Effects on locals sales seems to be less important and we find positive significant effects only in one of our strategies.

Firms also improve their credit situation, as they reported that access to financing was easier after certification.

We do not find statistically significant effects of the certification on either of the measures of firm productivity rejecting Ullah *et al.* (2014) previous findings for LAC .

Our findings provide several useful considerations for productive development policymakers. First, the study provides information about the types of firms most likely to seek and achieve a quality certification. The fact that firms with more experienced managers and more internationalized profile (i.e., exporters and foreign owned) are most likely to certify may be a signal of informational barriers regarding the benefits and requirements for quality certifications. This indicates that public intervention can play and important role eliminating this barrier by providing public information about international business opportunities for certified firms, and also offering training programs on quality process, reducing then informational costs for less experienced firms operating locally.

Regarding public policy design, policies or programs aimed to support certification for firms that require quality signaling to successfully entry (or expand) into foreign markets can be designed and implemented. For this purpose, those firms most hampered to information asymmetries and those less capable for facing certification's fixed costs must be identified.

For instance, policies could be designed with focus on local firms that wish to incorporate to Global Value Chains with high differentiation. For the export process to be successful, these firms must demonstrate that they can produce efficiently and with high quality standards, reducing the risks of failures and delivery delays. This seem to be an important tool facilitating integration between firms operating locally and more internationalized-firms, fostering first-time indirect exports, and providing an indirect path for local firms to entering into global value chains.

Furthermore, small and finance restricted firms may be impeded for covering sunk costs of quality certification process despite the benefits they could perceive after that.

Facilitating credit access or co-financing certifications could be an important policy tool for promoting exports.

Finally, the findings regarding productivity are not conclusive. The effect on firm performance may depend on the time periods during which the firm implements process improvements. In other words, it is possible that the immediate improvements are the ones observed in this paper and that, over a longer period, by starting to export and reducing the credit barrier, improvements in productivity can be achieved.

Acknowledgements

We wish to thank Roberto Alvarez, Jose Miguel Benavente, Gabriel Casaburi, Gustavo Crespi, and and participants at LACEA-LAMES 2017 for useful comments on a previous version of this paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank Group, its respective Boards of Directors, or the countries they represent. The usual disclaimers apply.

References

A. Smith, J. & E. Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2), 305–353.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies*, 72(1), 1–19.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Ann. Statist.*, 47(2), 1148–1178.

Austin, P. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28, 3083–107.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. PMID: 21818162.

Bellone, F., Musso, P., Nesta, L., & Schiavo, S. (2010). Financial constraints and firm export behaviour. *The World Economy*, 33(3), 347–373.

Benavente, J. M., Crespi, G., Figal Garone, L., & Maffioli, A. (2012). The impact of national research funds: A regression discontinuity approach to the chilean fondecyt. *Research Policy*, 41(8), 1461–1475.

Biau, G. & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Brito, P. & Mello, A. S. (1995). Financial constraints and firm post-entry performance. *International Journal of Industrial Organization*, 13(4), 543 – 565. The Post-Entry Performance of Firms.

Busso, M., DiNardo, J., & McCrary, J. (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *The Review of Economics and Statistics*, 96(5), 885–897.

Castillo, V., Figal Garone, L., Maffioli, A., & Ohaco, M. (2016a). *Asistencias técnicas y competitividad de las MiPyMEs: Evidencia para Argentina*. Technical report, Inter-American Development Bank.

Castillo, V., Figal Garone, L., Maffioli, A., Rojo, S., & Stucchi, R. (2016b). The effects of knowledge spillovers through labor mobility. *MPRA Paper No. 69141*.

Castillo, V., Figal Garone, L., Maffioli, A., & Salazar, L. (2017). The causal effects of regional industrial policies: a synthetic control approach. *Regional Science and Urban Economics*, 67, 25–41.

Crespi, G., Fernández-Arias, E., & Stein, E. (2014a). Rethinking productive development. In *Rethinking Productive Development* (pp. 3–31). Palgrave.

Crespi, G., Fernandez-Arias, E., & Stein, E., Eds. (2014b). *Rethinking Productive Development*. Washington: Palgrave Macmillan US.

Crespi, G., Garone, L. F., Maffioli, A., & Melendez, M. (2015). Long-term productivity effects of public support to innovation in colombia. *Emerging Markets Finance and Trade*, 51(1), 48–64.

De Loecker, J. (2007). Do exports generate higher productivity? evidence from slovenia. *Journal of international economics*, 73(1), 69–98.

Dehejia, R. & Wahba, S. (2000). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94, 1053–1062.

Diamond, A. & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *The Review of Economics and Statistics*, 95(3), 932–945.

Dick, G. P., Heras, I., & Casadesús, M. (2008). Shedding light on causation between iso 9001 and improved business performance. *International Journal of Operations & Production Management*, 28(7), 687–708.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*, 49(4), 1231–1236.

Duflo, E., Chernozhukov, V., Chetverikov, D., Demirer, M., Hansen, C., & Newey, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5), 261–65.

Efron, B. & Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Figal Garone, L. & Maffioli, A. (2016). Impact evaluation of cluster development programs: An application to arranjos produtivos locais policy in brazil. In C. Pietrobelli, R. Stucchi, & A. Maffioli (Eds.), *The Impact Evaluation of Cluster Development Programs: Methods and Practices*. Washington, DC: Inter-American Development Bank.

Figal Garone, L., Maffioli, A., de Negri, J. A., Rodriguez, C. M., & Vázquez-Baré, G. (2015). Cluster development policy, sme's performance, and spillovers: evidence from brazil. *Small Business Economics*, 44(4), 925.

Fikru, M. G. (2014). Firm level determinants of international certification: evidence from ethiopia. *World Development*, 64, 286–297.

Gereffi, G. & Fernandez-Stark, K. (2016). Global value chain analysis: A primer (second edition). *Center on Globalization, Governance & Competitiveness, Duke University.*

Gereffi, G., Humphrey, J., Kaplinsky, R., & Sturgeon, T. (2001). Introduction: globalisation, value chains and development. *IDS Bulletin*, 32(3), 1–8.

Goedhuys, M. & Sleuwaegen, L. (2013). The impact of international standards certification on the performance of firms in less developed countries. *World Development*, 47, 87–101.

Harrison, A. & Rodriguez-Clare, A. (2010). Trade, foreign investment, and industrial policy for developing countries. In D. Rodrik & M. Rosenzweig (Eds.), *Handbook of Development Economics 5* (pp. 4039–214). Elsevier.

Hernández, R. A., Martínez, J. M., & Mulder, N. (2014). *Global value chains and world trade: Prospects and challenges for Latin America*. ECLAC Books. Economic Commission for Latin America and the Caribbean (ECLAC).

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4), 1161–1189.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(03), 199–236.

Horvitz, D. G. & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), 663–685.

Hudson, J. & Orviska, M. (2013). Firm's adoption of international standards: One size fits all? *Journal of Policy Modeling*, 35(2), 289–306.

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society Series A*, 171(2), 481–502.

Imai, K. & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 243–263.

Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *The Review of Economics and Statistics*, 86(1), 4–29.

Imbens, G. W. & Rubin, D. B. (2015). *Estimating the Propensity Score*, (pp. 281–308). Cambridge University Press.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.

Javorcik, B. & Sawada, N. (2018). The iso 9000 certification: Little pain, big gain? *European Economic Review*, 105, 103–114.

King, A. A., Lenox, M. J., & Terlaak, A. (2005). The Strategic Use of Decentralized Institutions: Exploring Certification with the ISO 14001 Management Standard. *Academy of Management Journal*, 48(6), 1091–106.

King, G. & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, (pp. 1–20).

Lee, B. K., Lessler, J., & Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3), 337–346.

Leonidou, L. (2004). An analysis of the barriers hindering small business export development. *Journal of Small Business Management*, 42, 279 – 302.

Levinsohn, J. & Petrin, A. (2003). Estimating Production Functions using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2), 317–41.

Li, Q., Racine, J. S., & Wooldridge, J. M. (2008). Estimating average treatment effects with continuous and discrete covariates: The case of swan-ganz catheterization. *American Economic Review*, 98(2), 357–62.

M. Robins, J. & Ritov, Y. (1997). Robins jm, ritov y. toward a curse of dimensionality appropriate (coda) asymptotic theory for semi-parametric models. *Statistics in Medicine - STAT MED*, 16, 285–319.

Mccaffrey, D., Ridgeway, G., & Morral, A. (2005). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, 9, 403–25.

Otsuki, T. (2011). Effect of International Standards Certification on Firm-Level Exports: An Application of the Control Function Approach. *The Empirical Economics Letters*, 10(7).

Pekovic (2010). The Determinants of ISO 9000 Certification: A Comparison of the Manufacturing and Service Sectors. *Journal of Economic Issues*, 44(4), 895–914.

Pietrobelli, C. & Rabellotti, R. (2010). Upgrading to compete global value chains, clusters, and smes in latin america. *Harvard University Press*.

Robins, J. M. & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429), 122–129.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429), 106–121.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82(398), 387–394.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

Rotnitzky, A. & Robins, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika*, 82(4), 805–820.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.

Schiavo, S. & Musso, P. (2008). The impact of financial constraints on firm survival and growth. *Journal of Evolutionary Economics*, 18, 135–149.

Silva, A., Afonso, O., & Africano, A. P. (2012). Learning-by-exporting: What we know and what we would like to know. *The International Trade Journal*, 26(3), 255–288.

Starke, F. & Rangamonhan, E. (2012). Impact of ISO 9000 Certification on Firm Performance: Evidence from Brazil. *Management Research Review*, 35(10), 974–97.

Sun, Y. & Outyang, W. (2014). International Standards for Exporting Firms: Evidence from China. *The Journal of Applied Business Research*, 30(6), 1753.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, 48(2), 326–365.

Trifkovic, N. (2017). Spillover Effects of International Standards: Working Conditions in the Vietnamese SMEs. *World Development*, 97(C), 79–101.

Ullah, B., Wei, Z., & Xie, F. (2014). ISO certification, financial constraints, and firm performance in Latin American and Caribbean countries. *Global Finance Journal*, 25(3), 203–28.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.

Volpe Martincus, C., Castresesana, S., & Castagnino, T. (2010). ISO Standards: A Certificate to Expand Exports? Firm-level Evidence from Argentina. *Global Finance Journal*, 18(5), 896–912.

Westreich, D., Lessler, J., & Jonsson Funk, M. (2010). Propensity score estimation: Neural networks, support vector machines, decision trees (cart), and meta-classifiers as alternatives to logistic regression. *Journal of clinical epidemiology*, 63, 826–33.

Wooldridge, J. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141(2), 1281–1301.

Xiaoyang Chen, M., Wilson, J., & Otsuki, T. (2008). Standards and Export Decisions: Firm-level Evidence from Developing Countries. *The Journal of International Trade & Economic Development*, 17(4), 501–23.