RIETI Discussion Paper Series 17-E-096

The Pollution Outsourcing Hypothesis: An empirical test for Japan

Matthew A. COLE
University of Birmingham

Robert R.J. ELLIOTT
University of Birmingham

OKUBO Toshihiro
Keio University

Liyun ZHANG
University of Birmingham

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 17-E-096

July, 2017

The Pollution Outsourcing Hypothesis: An empirical test for Japan

Matthew A. COLE (University of Birmingham)

Robert R.J. ELLIOTT (University of Birmingham)

OKUBO Toshihiro (Keio University)

Liyun ZHANG (University of Birmingham)

Abstract

This paper investigates whether firms that engage in outsourcing improve their environmental performance using Japanese firm-level data for the period 2009-2013. To identify the causal effect of production outsourcing on firm carbon dioxide (CO_2) emission intensities, we employ a non-parametric approach combining propensity score matching (PSM) and difference-in-differences (DiD). Our results show that, relative to the control group, the growth in CO_2 emission intensities (relative to the year before treatment) of new production outsourcers is 5.1% lower in the year when they start outsourcing, and 6.6% and 9.5% lower one and two years after outsourcing, respectively. When we decompose firms' outsourcing activities into domestic and foreign according to the destination of the outsourced production, we find that the effects on emission intensity growth are driven by overseas outsourcing. Firms that outsource part(s) of their production overseas have a 7.3% lower emission intensity growth when they start outsourcing and a 7.7% reduction in the following year. We also investigate whether the decision to import or export has an impact on firm level environmental performance as predicted by the more traditional pollution halo hypothesis (PHH) literature. Firms are found to have a 3.3% lower growth rate of CO_2 emission intensity when they start to import, but no significant impact is found for exporting.

Keywords: Outsourcing, Environment, Propensity score matching, Difference-in-differences.

JEL classification: F18, F23, L51, L60, Q56, R3

This study is conducted as a part of the project "Regional Economies in the New Era of Globalization and Informatization" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study utilizes the micro data of the questionnaire information based on the "Basic Survey of Japanese Business Structure and Activities" which is conducted by the Ministry of Economy, Trade and Industry (METI). The author is grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI.

1. Introduction

In recent years there has been increasing interest from both academics and policymakers in the relationship between globalisation and environmental outcomes with a popular view being that the international activities of firms are likely to be detrimental to the natural environment. The premise that polluting firms may relocate to countries or regions with low environmental regulations or that regulations affect trade flows through changes in the competitive environment is known as the pollution haven hypothesis (PHH). As such, a growing literature has examined the link between a firm's international activities, such as exporting, importing and FDI and the environment (see for example, Kellenberg, 2009; Cui *et al.*, 2012; Batrakova and Davies, 2012; McAusland and Millimet, 2013; Chung 2014; Rezza, 2015; Girma and Hanley, 2015; Forslid *et al.*, 2015 and Holliday, 2016). Although, recent studies have refined the methodological approach to deal with endogeneity and reverse causality issues, conclusive evidence of a pollution haven consistent effect remains elusive. A related but relatively small literature investigates the interaction between international outsourcing and the environment, that we call the pollution outsourcing hypothesis (POH), whereby outsourcing provides a channel by which domestic firms may shift the dirtiest part(s) of the production process abroad to reduce the average emission intensity of the firm (Clark *et al.*, 2000; Levinson, 2010; Cole *et al.*, 2014; Brunel, 2016 and Cole *et al.*, 2017 who review the recent literature on the POH). However, to the best of our knowledge no empirical studies have provided a direct examination of the effect of international and domestic outsourcing on the environment.[1]

This purpose of this paper is therefore to test whether outsourcing affects the environmental performance of firms (CO_2 emissions as a percentage of total sales) utilizing data on a range of firm level characteristics for around 4,000 Japanese firms between 2009 and 2013. More specifically, the contribution of this paper is four-fold. First, we employ a non-parametric approach combining propensity score matching (PSM) and difference-in-differences (DiD) to investigate whether there is a causal effect between the decision to outsource and the subsequent CO_2 emission intensity of firms in the following two years. The PSM-DiD approach allows us to control for time invariant unobservables and reverse causality as we are able to examine the impact on CO_2 emissions for firms that start production outsourcing conditioning on previous trends in CO_2 emission intensities.

[1] See Cheriwchan *et al.* (2016) for a recent review of the trade and environment literature and refers to the pollution offshoring hypothesis (which is similar to the pollution outsourcing hypothesis but excludes the possibility of domestic outsourcing).

Second, in contrast to the majority of studies of outsourcing, we are able to distinguish between domestic outsourcing and foreign outsourcing (often referred to as offshoring). This distinction allows us to investigate whether Japanese firms are using foreign outsourcing as a mechanism by which they are able to avoid more stringent environmental regulations in Japan so as to appear "greener" which would be consistent with a pollution haven type effect. Although it is difficult to compare the stringency or laxity of the environmental regulations between Japan and the destination countries (as we have limited information on the destination of the offshored production) it is reasonable to assume that regulations are lower, or at best as strict, as those of Japan. An alternative to foreign outsourcing is that a Japanese firm uses a specialist domestic supplier as their outsourcing partner. Given that all emitting firms in Japan are regulated by the same environmental policies and enforcement is strong, there is little incentive for the domestic firms to engage in outsourced dirty production from other dirty firms. On the other hand, domestic outsourcing could provide a quick, although potentially expensive, solution to reducing firm level emissions. Third, we contribute to the more traditional PHH literature and investigate whether the decision to import or export has an impact on firm level CO_2 emission intensity where importing dirty goods would be another way to avoid stringent domestic regulations. Finally, we investigate the impact of production outsourcing on other firm level characteristics using our PSM-DiD methodology and compare the results with those of the existing offshoring literature.

We believe that Japan represents an ideal country to test for evidence consistent with the POH. Japan has an established network of overseas suppliers to its domestic industries via the flying geese model of global value chains so has a range of options for outsourcing. Second, Japan established the Environmental Agency and started to implement environmental policies to combat serious industrial pollution in the early 1970s (OECD, 1977). The environmental legislation was further strengthened in the 1990s (OECD, 2002; 2010). During this time Japan made significant improvements in its environmental quality and is widely recognized as an environmental technology and policy leader (Imura and Schreurs, 2005). As part of the regulatory environment, Japan introduced environmental regulations that imposed extra costs for manufacturers. Firms can respond to environmental regulations in several ways. They can purchase pollution abatement equipment, upgrade their production line by investing on technological innovation, or simply outsource the dirtiest part(s) of their production. However, adoption of advanced technology or investment in innovation is costly and time-consuming for the manufacturers.[2]

---

[2] . Hamamoton (2006) finds a positive relationship between regulatory stringency in Japanese manufacturing industries and R&D expenditure. For details on the ecological modernisation of Japan see Elliott and Okubo (2016).

Although relatively few papers have examined the impact of outsourcing on the environmental performance of firms there are a small number of papers that examine the relationship between exporting and environmental outcomes. For example, Batrakova and Davies (2012) and Girma and Hanley (2015) both show that exporters improve their environmental performance (become greener). Batrakova and Davies (2012) develop a theory of the decision to export and the adoption of environmentally-friendly technology using data on Irish manufacturing firms and show that exporting firms with low energy intensity increase their energy use while those with high energy intensity reduce their energy use. Girma and Hanley (2015) using UK firm-level data show that exporters are greener than non-exporters where exporting is associated with the introduction of more energy and materials saving innovations. Cui *et al.* (2012) model the relationship between firm productivity, exports and the emission of various air pollutants using US facility-level data and find a negative correlation between export status and emission intensity. In a related paper, McAusland and Millimet (2013) identify the channels through which trade impacts the environment. Using data on intra and international trade and environmental outcomes for the US and Canada they find that international trade has a statistically and economically beneficial causal effect on environmental quality, while intra-national trade has a harmful impact.[3]

In a more recent paper, Forslid *et al.* (2015) develop a model of trade and environmental emissions in a heterogeneous firm setting. Confirming the predictions of the model they show for Swedish manufacturers that exporters have between 10 and 30 percent lower CO_2, SO_2 and NO_x intensities and that exporting leads to a 62-73% increase in abatement activities for firms in non-energy intensive industries. For the US, Holladay (2016) assesses the relationship between international trade, productivity and environmental performance and finds that exporters pollute 9-13% less than non-exporters. Finally, using Japanese data, Jinji and Sakamoto (2015) study the relationship between exports and environmental performance. Taking a PSM approach they find that exporting reduces total CO_2 emissions of most industries although it increases CO_2 emissions for the iron and steel industry.

Turning to the limited environmental outsourcing literature, in the research most closely related to our own, Levinson (2010) examines whether the US increased the extent to which firms offshored pollution-intensive goods and shows that, on the contrary, the US imported a greater proportion of relatively clean goods and not an increased proportion of newly outsourced dirty goods as one might expect. More recently, Michel (2013) investigates the role of offshoring on

[3] The general trade literature provides considerable evidence that more productive firms are more likely to engage in international trade and are more likely to innovate (Bustos, 2011, Hanley and Pérez, 2012) and that engaging in international trade has a positive impact on productivity (De Leocker, 2007 and Elliott *et al.*, 2016).

the reduction in air emissions for a sample of Belgian manufacturing firms. Results from a decomposition analysis shows that offshoring contributed to a 17% reduction in greenhouse gas emissions, 6% in acid rain and 7% in tropospheric precursor emissions. Li and Zhou (2017) also study offshoring and firm's environmental strategies using trade, production and pollution data for the US. They find that US plants shift production to less pollution-intensive industries, produce less waste and spend less on pollution abatement when their parent company imports more from low-wage countries.[4]

Other papers that have investigated the environmental impact of international outsourcing include for example, Cole *et al.* (2014) who develop a theoretical model of international environmental outsourcing in which heterogeneous firms can either pay an abatement cost or a fixed cost to offshore their polluting activity. They find evidence of an 'environmental outsourcing' effect using Japanese firm-level data with pollution intensive and high regulation cost firms being more likely to outsource. Although this paper uses Japanese firm-level data it says nothing about the impact of outsourcing on the environmental performance of firms. Other examples include Antonietti *et al.* (2016) who analyse a survey of 684 Italian firms for 2011 to show that environmental policy stringency does not have a significant impact on firms' FDI decisions but that it does increase the probability that a firm outsources to less developed countries. Finally, Lyu (2016) examines the link between different offshoring tasks in China and CO_2 emissions for 12 industry sectors in 2010. The results show that polluting industries, such as iron and steel, nonferrous metals and chemicals, generate the most CO_2 emissions during the production of processed goods and that higher energy consumption industries have higher CO_2 emissions induced by offshoring. In each case these papers only consider one year of data.[5]

To briefly summarise our results, for our analysis of Japanese firms between 2009 to 2013 we find that relative to a control group, the CO_2 emissions intensity growth rate of new production

[4] The traditional outsourcing literature has tended to examine the relationship between outsourcing, foreign ownership and the productivity (Girma and Gorg, 2004). For example, Tomiura (2005) investigates the link between the decision to outsource overseas and firm level characteristics for a large sample of Japanese manufacturers and shows that firms that are larger or more skill-intensive are more likely to start foreign outsourcing. A second series of papers examines the impact of outsourcing on firm performance. Tomiura (2007) compares the productivity differences of firms that conduct foreign outsourcing, exporting and FDI and shows that foreign outsourcers and exporters are less productive than firms engaged in FDI while Gorg *et al.* (2008) shows how international outsourcing affects productivity and shows that outsourcing of service inputs has a positive impact on productivity for exporters only.

[5] An alternative approach is taken by Cadarso *et al.* (2010) who proposes a new methodology to quantify CO_2 emissions linked to the increase in international transport that is the result of offshoring. Using the Spanish emissions data, they find that the total CO_2 emissions as a result of international freight transportation increase by around 4.16% between 1995 and 2000. Another study to consider the environmental implications of outsourcing is Leoncini *et al.* (2016) who examine the link between CO_2-reducing innovations and outsourcing for Italian manufacturing firms in two green industries. They find that outsourcing tangible assets increases the propensity of a firm to implement CO_2-reducing innovations while the opposite holds for intangibles outsourcing.

outsourcing firms is 5.1% lower in the year that they start outsourcing. In the next one and two years the CO_2 emissions intensity growth rate is 6.6% and 9.5% lower than the growth rate the year before the firm started outsourcing, respectively. After decomposing firms' outsourcing activities into domestic and foreign outsourcing we find that the outsourcing effect on emission intensity growth is driven by foreign offshoring with firms that start overseas outsourcing experiencing a 7.3% lower emission intensity growth in the year they start outsourcing and 7.7% a year later compared to the year before they start outsourcing. For domestic-only outsourcing we find no evidence that firms have a significantly lower emissions intensity growth when they start outsourcing or in the subsequent two years. Hence, we only find evidence consistent with the POH for foreign outsourcing. In other outcomes we find that domestic-only outsourcing increases the growth rate of average wages and productivity a year after they decision to start outsourcing and R&D intensity two years after starting outsourcing compared to the year before outsourcing started. For overseas outsourcing, firms are found to increase both exports and imports when they start outsourcing compared to those that never outsource their production. When we investigated whether the decision to import had an impact on firm level environmental performance we found that firms had a lower growth rate of CO_2 emission intensity when they started to import which is also consistent with a PHH type effect.

The remaining of the paper is organized as follows. Section 2 presents the methodology. Section 3 describes the data. Results are reported in Section 4 and finally Section 5 concludes.

2. Methodology

Our methodological approach is to employ propensity score matching (PSM) to identify the potential causal effects between outsourcing and firm environmental performance. PSM allows us to identify whether there is any change in a particular outcome following a treatment by comparing the outcome for firms with the treatment to those where the treatment did not apply. In our case the treatment is the decision to outsource and the outcome is firm's subsequent environmental performance measured by that firm's CO_2 emissions intensity (CO_2 emissions divided by total sales).[6]

[6] Matching methods have also been used to evaluate effects of environmental regulations on plant births (List *et al.*, 2003, 2004), the land value effects of agriculture land preservation programmes on farmland loss (Liu and Lynch, 2011) and the effectiveness of the emission trading programme (Fowlie *et al.*, 2012). More generally, PSM methods have been widely adopted in studies evaluating the effects of various policies or programs, such as labour market programmes (Heckman and Todd, 1997; Sianesi, 2008; Lechner and Wunsch, 2013), the impact of Kyoto Protocol

We define y_{it} as firm *i*'s CO_2 emissions intensity in period *t* and $y_{i(t+s)}$ as the emissions intensity *s* period(s) later ($s \geq 0$). The causal effect of outsourcing on emissions intensity of firm *i* at *t+s* is given by:

$$y^1_{i(t+s)} - y^0_{i(t+s)} \tag{1}$$

where the superscripts denote outsourcing behaviour which is equal to 1 if firm *i* outsources at *t* and zero otherwise. Hence, $y^0_{i(t+s)}$ represents the emissions intensity of firm *i* at period *t+s* as if it had never outsourced since *t*.

The population average treatment effect (*ATE*) is calculated as the difference in the expected outcomes of the treatment groups and their counterfactual counterparts:

$$ATE \ = \ E[y^1_{i(t+s)} - y^0_{i(t+s)}] \tag{2}$$

In order to identify whether there are differences in firms' emissions intensity following the decision to outsource, we focus on the new outsourcers. The average effect on CO_2 intensity that outsourcing starters would have experienced, i.e. the average treatment effect on the treated (*ATT*), is given by:

$$\begin{gathered} ATT \ = \ E\left[y^1_{i(t+s)} - y^0_{i(t+s)} \middle| newOS_{it} = 1\right] = \\ E\left[y^1_{i(t+s)} \middle| newOS_{it} = 1\right] - E\left[y^0_{i(t+s)} \middle| newOS_{it} = 1\right] \end{gathered} \tag{3}$$

where *newOSit* is a dummy variable which equals 1 if firm *i* begins to outsource at time *t* and zero otherwise. The fundamental evaluation problem with equation (3) is that although we can estimate $E\left[y^1_{i(t+s)} \middle| newOS_{it} = 1\right]$, we cannot estimate $E\left[y^0_{i(t+s)} \middle| newOS_{it} = 1\right]$ and thus a direct estimation is not possible.

The literature on the causal effects of observational studies (for example, Rosenbaum and Rubin, 1983, 1985; Heckman *et al.*, 1998 and Dehejia and Wahba, 2002) suggests that the solution is to find an *appropriate* control group of observations so that, in our case each new outsourcer is matched with a comparison observation, i.e. a firm with similar (ideally identical) observable characteristics but that it has never participated in outsourcing related activities:

on bilateral trade flows (Aichele and Felbermayr, 2013), effect of international market participation on firm performance (Yasar and Rejesus, 2005; Elliott *et at.*, 2016) and the impact of foreign acquisition on employment and wages (Huttunen, 2007; Girma and Gorg, 2007; Heyman *et al.*, 2007).

$$E\left[y^0_{i(t+s)}\middle|newOS_{it}=1,X\right]=E\left[y^0_{i(t+s)}\middle|newOS_{it}=0,X\right] \tag{4}$$

where X is a vector of covariates of firm characteristics. Equation (3) can then be rewritten as:

$$ATT\ =\ E\left[y^1_{i(t+s)}\middle|newOS_{it}=1,X\right]-E\left[y^0_{i(t+s)}\middle|newOS_{it}=0,X\right] \tag{5}$$

However, finding observations with identical values for all covariates in X is not practical. Hence, Rosenbaum and Rubin (1983) propose a way to find a control group conditioning on $P(X)$ which is equivalent to conditioning on X:

$$P(X)=Pr(newOS_{it}=1|X)=\Phi(X_{i(t-1)},D_j,D_t) \tag{6}$$

where P denotes the propensity of firm i to start outsourcing at time t, and $\Phi(.)$ is the normal cumulative distribution function. X is a vector of firm characteristics including age (*logage*), size (*logemp*), average employee wages (*logwage*), labour productivity (*logLP*), export activity (*EXP, logexp* or *EXPshare*), import paritcipation (*IMP, logimp* or *IMPshare*), foreign ownership (*FOR* or *FORshare*), R&D activity (*RD* or *RDshare*) and foreign direct investment activity (*FDI*).[7] A full set of industry dummies (D_j) and year dummies (D_t) are also included to capture industry and time effects respectively. All time-variant explanatory variables are lagged by one year in order to mitigate simultaneity concerns. Furthermore, we include the pre-treatment growth of CO_2 emissions intensity (*pregrowth*) in the estimation. It is important to include pre-treatment growth as there exists the possibility that firms that start to outsource were already on a permanently different growth rate of CO_2 intensity (either higher or lower) than those firms that never outsource and failure to control this could result in this difference mistakenly capturing the decision to start production outsourcing.[8] Table A1 in the Appendix provides detailed definitions of our control variables. Since the treatment *newOS* is binary we estimate Equation (6) using a Probit regression.

Once we have estimated the propensity scores, the next step is to test the balancing properties of the propensity scores across treated and control groups. Following Dahejia and Wahba (2002), Imbens (2004) and Garrido *et al.* (2014), we split the sample into k blocks of the propensity scores and test within each block whether the mean propensity score is equivalent in the treatment and control groups. If the balancing test fails, we split the interval into smaller blocks and test again. We continue this process until equality holds for every interval. After the propensity score is balanced within blocks across the treated and control groups, we check for the balance of each

[7] We are unable to estimate firm level total factor productivity (TFP) due to a lack of information on firm-level intermediate inputs.

[8] There is an ongoing debate in the literature as to which variables to include in the propensity score model. See for example, Heckman *et al.* (1998), Imbens (2004), Ho *et al.* (2007) and Garrido *et al.* (2014).

observed covariate within blocks of the propensity score. If the balancing test is rejected, we modify the covariates in the propensity score estimation equation, for example by replacing continuous variables with categorical variables or including higher order terms or splines of the variables.

After creating a balanced propensity score, we match each outsourcing starter i with a non-outsourcing firm j so that the distances between the estimated propensity scores of a treated and control observation are within a range:

$$\left|P_{i(t+s)} - P_{j(t+s)}\right| = \min\left\{\left|P_{i(t+s)} - P_{j(t+s)}\right|\right\} < \lambda \tag{7}$$

where λ is a pre-defined scalar (e.g., bandwidth or calliper) and P_i and P_j are the estimated propensity scores for treatment and control observations.

Several matching algorithms have been developed. The most popular include nearest neighbour matching, calliper and radius matching, kernel matching and stratification matching.[9] In this paper, we adopt kernel matching and we impose the common support condition by dropping the outsourcing starters whose propensity scores are higher than the maximum or lower than the minimum of those persistent non-outsourcers. Kernel matching is shown to maximize precision as more information is used than with other matching algorithms as the sample size is maintained because only observations outside the range of common support are discarded (Garrido *et al.* 2014). The choice of bandwidth is important as it leads to a trade-off between bias and variance (Silverman 1998 and Garrido *et al.* 2014). High bandwidth values yield a smoother estimated density function which should lead to a better fit and a decrease in the variance between the estimated and the true underlying density function. However, bias may be induced as a result of selecting a wide bandwidth as potentially interesting and important features of the population regression function may be smoothed away in response to the weakness of common support (Caliendo and Kopeinig 2008).[10]

In this paper we perform matching with replacement so that a comparison observation can be matched more than once to the treated observations. The benefit of this approach is to reduce the bias as matching without replacement may mean that treated units are forced to match with units that are quite different when there are few comparison units that are similar to the treated

[9] See Stuart (2010) for a review of propensity score matching and Austin (2013) for a comparison of 12 different algorithms for matching on the propensity score.
[10] See Silverman (1986), Chiu (1991) and Sheather (2004) for more discussion on density estimation and bandwidth selection.

units (Dehejia and Wahba, 2002). Rather than matching across the entire manufacturing sector, our matching is performed within each 2-digit-sector-year group. In this way we create control groups within narrowly defined industries in the same year. This is important as firms in different industries face different technological and market conditions and the propensity to start outsourcing of these firms may differ substantially between different industries. Similarly, if matching is not done within the same year, an outsourcing starter in the treatment year can be matched with a control firm in any year.[11]

A drawback of the traditional PSM method is that it cannot control for any unobserved time invariant firm characteristics that may influence the decision to outsource or the outcome. Having constructed the control group of firms (C) that are similar to the treated firms (T) by using a PSM approach, we then use a difference-in-differences (DiD) method to estimate the causal effect of outsourcing on emissions intensities. The advantage of a combined PSM-DiD is that it improves the accuracy of the estimates as we are able to control for time-invariant unobserved firm characteristics.

A DiD estimator first measures the difference in the emissions intensity before and after a firm starts outsourcing for the treated firms conditioned on past performance and a set of dummy variables. However, such differences in emissions intensities cannot be exclusively attributed to outsourcing behaviour as post-entry emissions intensity growth might be caused by factors that are contemporaneous with entry into outsourcing. The second step is to difference the differences obtained for the outsourcing starters with the corresponding difference for non-outsourcing firms. Since DiD estimates the difference before treatment it removes the effects of common shocks and hence provides a more accurate estimate.

As Blundell and Costa Dias (2000, p.438) point out, a non-parametric approach that combines propensity score matching with difference-in-differences has the potential ". . . to improve the quality of non-experimental evaluation results significantly". Hence, we combine PSM with DiD such that the selection on unobservable determinants can be allowed when the determinants lie on separable firm and /or time-specific components of the error-term. Hence, imbalances in the distribution of covariates between the treated and control groups account for time varying unobserved effects that influence both the decision to outsource and the emissions intensity of firms. Our PSM-DiD estimator based on a sample of matched firms is then given by:

[11] In practice we create a number of bins for each 2-digit sector-year combination and assign each observation to a bin. Matching is then performed within each of the bins depending on the estimated propensity scores of each observation. Similar approach is also followed by Elliott *et al.* (2016).

$$\mathrm{ATT}^{\mathrm{PSM-DiD}} = \frac{1}{N_T} \sum_{i \in T} [\Delta y_{i(t+s)} - \sum_{j \in C} w_{ij} \Delta y_{j(t+s)}] \tag{8}$$

where *T(C)* denotes the treatment (control) group, N_T is the number of firms in the treatment group on the common support, *t* is the time period when treatment occurs (a firm starts outsourcing), $\Delta y_{i(t+s)}$ and $\Delta y_{j(t+s)}$ are the differences in emission intensities between *s* periods ($s \geq 0$) after treatment at *t* and pre-treatment period (*t*-1) for firms in treated group and control group respectively, i.e., $\Delta y_{i(t+s)} = y^T_{i(t+s)} - y^T_{i(t-1)}$ and $\Delta y_{j(t+s)} = y^C_{j(t+s)} - y^C_{j(t-1)}$, and w_{ij} is the weight placed on the matched control firm *j* when constructing the counterfactual estimation for treated firm *i*.

An important step after matching is to assess the quality of the matching procedure. We perform several of the balancing tests that are suggested in the literature (Rosenbaum and Rubin, 1985, Smith and Todd, 2005, Caliendo and Kopeinig, 2008 and Austin, 2009). The first stage is to compare the situation before and after the matching to check if there are any differences in the means of the observable characteristics for firms from both the treatment and control groups after conditioning on the propensity score. Differences in the means between the groups should be reduced significantly after matching. A formal two-sample *t*-test between the treated and control groups for each variable is also performed to ensure that no significant bias exists.

The second stage is to examine the standardized difference (SD) (or percentage bias) between the treated and control samples for all variables used in the PSM. The lower the SD, the more balanced the treated and control groups will be in terms of the variable being considered. The standardized difference for comparing means between groups for continuous variables is given by:

$$SD = 100 \frac{\overline{X_T} - \overline{X_C}}{\sqrt{\frac{S_T^2 + S_C^2}{2}}} \tag{9}$$

where $\overline{X_T}$ and $\overline{X_C}$ denote the sample means of the variable *X* in treated and control groups, respectively, while S_T^2 and S_C^2 are the sample variances of the variable in treated and control groups, respectively.

For dichotomous variables, the standardized difference is defined as:

$$SD = 100 \frac{\widehat{P_T} - \widehat{P_C}}{\sqrt{\frac{\widehat{P_T}(1-\widehat{P_T}) + \widehat{P_C}(1-\widehat{P_C})}{2}}} \tag{10}$$

where $\widehat{P_T}$ and $\widehat{P_C}$ denote the mean of the dichotomous variable P in treated and control groups, respectively. Unlike *t*-tests, the standardized difference is not influenced by sample size (Austin, 2007). Thus, the use of the standard difference allows us to compare the balance in measured variables between treated and the control in the matched sample with those in the unmatched sample (Flury and Riedwyl, 1986 and Austin, 2009).

There are no formal criteria specified in the literature for when a standardized difference is considered too large. Rosenbaum and Rubin (1985) suggest that a value of 20% of standardized difference is large. Sianesi (2004) suggests that the propensity score is re-estimated for the matched sample and the *pseudo*-R^2s are compared before and after matching given that the *pseudo*-R^2 indicates how well the variables X explain the participation probability. After matching there should be no systematic differences in the distribution of covariates between both groups and therefore the *pseudo*-R^2 should be fairly low. Finally, we also perform a likelihood-ratio test on the joint insignificance of all variables in the Probit model where we would like to see the test reject before matching but not afterwards.

3. Data Description

For our empirical analysis we use two datasets, namely the annual surveys of Japanese firms and data on CO_2 emissions from the Mandatory Greenhouse Gas Accounting and Reporting System, which is provided by the Ministry of Economy, Trade and Industry (METI) and the Ministry of the Environment (MOE). This dataset includes an estimation of the emissions of greenhouse gases (GHGs) at the firm level. The CO_2 data covers all firms where total energy use is greater than 1,500kl per year which means that our sample consists of relatively large firms. All other firm-level data such as employees, wage, capital, sales, R&D, exports and a measure of outsourcing are taken from the Basic Survey of Japanese Business Structure and Activities, which is an annual firm-level survey conducted by METI. The data includes all manufacturing firms with more than 50 regular employees and have at least 30 million Yen (approximately US$275,000) of capital assets. The two datasets are matched using firm name and address.

As part of the data cleaning process we drop observations with missing values for tangible assets, total employment, CO_2 or sector code. In addition we drop those observations with zero values for CO_2, wages or tangible assets and with negative values for R&D. We also remove observations where the export value is greater than total sales (total sales includes exports). After cleaning we have an unbalanced panel with 19,503 observations for the period 2009 to 2013. All nominal values are in 2005 prices using a GDP deflator. See Table A1 of the appendix for a description of the variables that we use in our estimations.

To generate our variable of interest we measure firm-level environmental performance using firm-level CO_2 emission intensity (*co2sales*) which is defined as the CO_2 emissions of a firm divided by total sales. The reason we use CO_2 emissions intensity is that improvements in energy efficiency or changes in the production process should translate into changes in a firm's CO_2 intensity per unit of sales. One way to think about how this measure can capture the impact of outsourcing is to think about a firm that manufactures a final good that combines three intermediate inputs (one that is relatively energy intensive and is part of the production process) and two service inputs (for example, IT and human resources which are relatively clean) that are all internal to the firm. If outsourcing is motivated by more stringent regulations then one might expect the firm to outsource the most pollution intensive of the three intermediate inputs which should lead to the greatest reduction in the emissions intensity per unit of sales. On the other hand, if a firm decides to outsource its IT department this should have relatively little impact on the CO_2 intensity of the final good.

Table 1 presents the summary statistics of our variables of interest for the full sample and for a range of different sub-samples. The bottom row of Table 1 shows that about 80% of the observations in the sample engage in some form of outsourcing activity (column 3). We also make a distinction between service and production outsourcing and find that the majority of outsourcers (90%) outsource some part of their production process (columns 4 and 6). Around half of outsourcing firms (7,011 firms out of 15,632 outsourcers in Column 4) engage in both service and production outsourcing while 11% (1,651 firms out of 15,632 in Column 5) only outsource some element of their services. Table 1 shows that outsourcers are on average smaller in terms of sales and employment but are more likely to export, import and engage in R&D.

Of primary interest for our analysis is the environmental performance of firms. Table 1 (columns 2 and 3) shows that non-outsourcers have larger sales and greater CO_2 emissions. However, outsourcers, on average, have lower CO_2 intensity (1.86 vs 2.38). Comparing outsourcers (Columns 4, 5 and 6) we find, as expected, that firms that outsource part(s) of their production are cleaner than those that only outsource services (1.70 vs 3.30). Moreover, among firms who only outsource part of their production process rather than services (columns 7, 8 and 9), those firms that outsource their production overseas have a lower average CO_2 intensity than those outsourcing domestically (1.77 for domestic outsourcing only vs. 1.24 for foreign outsourcing or 0.93 for both). The descriptive evidence supports our priors. In the following analysis we discard firms that only outsource their services and concentrate on production outsourcing where the impact of environmental performance is most likely to be driving the outsourcing decision.

4 Results

4.1 Environmental impact of outsourcing

Before we present the treatment effects of outsourcing, we first need to make sure of the quality of our matching procedure. Tables A2 and A3 in the Appendix present the balancing test results on Kernel matching for production outsourcing on firm's environmental performance.[12] Table A2 compares the individual covariate included in the matching process between treated and control samples before and after matching is performed. Standardized differences, *t*-tests and variance ratios for continuous variables of treated group over non-treated show that differences exist in some of the covariates between two groups before matching, but no statistical difference in the matched samples. Table A3 provides information on overall measures of covariate imbalance before and after matching including Pseudo-R^2, likelihood-ratio tests and mean and median bias. Statistics indicate imbalance in the unmatched sample, but sufficient balance in the matched sample. The quality of our matching is thus satisfactory.

In Table 2 presents our PSM-DiD results for two difference matching procedures. We present the results for Kernel matching with a bandwidth of 0.03 and radius matching with a calliper of 0.03.[13] As part of our sensitivity checks we re-estimated our results using bandwidths of 0.01 and 0.06 for the Kernel matching and for radius matching we tried a range of different callipers e.g., 0.01, 0.03, 0.06 and 0.2 of the standard deviation of the estimated propensity score (Austin, 2009, 2010). Matching is done with common support and replacement. There was no discernible difference in the results and hence are not presented in the paper but are available from the authors upon request. Given our relatively short time period we are only able to consider changes in emission intensities for up to 2 years after the treatment.

[12] We perform balancing tests for each matching procedure in the subsequent estimations and ensure that matching is of good quality using methods described in Section 3. Balancing test results for the quality of the match for each estimation and for other outcomes are not presented in the paper for reasons of space but are available from the authors upon request.

[13] Austin (2010) compares the performance of different calliper widths for propensity score matching and concludes that calliper of 0.02, 0.03 or 0.2 of the standard deviation of estimated propensity score perform best.

Table 2 presents the results from both matching estimators and indicates a consistently negative and significant treatment effect for up to two years. The estimates from Kernel matching show that, relative to the control group, the growth in CO_2 emissions intensity for new production outsourcing firms is 5.1% lower in the year that they start outsourcing than comparable firms that did not start outsourcing. In the first and second years after the initial outsourcing decision is made, CO_2 emissions growth is 6.6% and 9.5% lower respectively. Given that both matching methods produce similar results on the treatment effects of production outsourcing on firms' environmental performance, we only present our PSM-DiD results from Kernel matching with bandwidth of 0.03 for the subsequent analysis.

[Table 2 about here]

In Table 3 we make a distinction between those firms that outsource part of their production process domestically only and those that outsource overseas (also known as offshoring).[14] The first row of Table 3 shows that for firms that start production outsourcing domestically only there is no significant impact on their environmental performance in the year of the treatment or over the following two years. The growth in CO_2 intensity is 8.6% lower after two years (significant at the 10% level). When we look at foreign outsourcing we find, relative to the control group, firms experience a significant reduction in the growth of CO2 emissions with a fall of 7.3% in the year of the decision to outsource which increases to reduction of 7.7% one year later. There is no significant effect two years after a firm starts to outsource overseas although the sample size is now rather small which reduces our confidence in the reliability of that coefficient. The results indicate that the impact of outsourcing is mainly driven by firms that carry out foreign outsourcing which supports the underlying premise of the POH.

[Table 3 about here]

4.2 Environmental impact of importing and exporting

[14] The two categories of outsourcing firms here are domestic-only outsourcers and overseas outsourcers. Domestic-only outsourcers are those that outsource production to domestic firms rather than foreign countries while overseas outsourcers are those outsource all or part(s) of its production abroad regardless of domestic outsourcing activities. Overseas outsourcers thus include firms that outsource to foreign countries only and those outsource domestically and overseas at the same time. Ideally distinguishing between domestic-only and foreign-only outsourcers would enable us to better identify the source or channel of the effects of outsourcing on environment. However as showed in previous section, most of the foreign outsourcers also outsource domestically and only a few firms outsource overseas only. There are only 25 new foreign-only outsourcers can be used as treated for the PSM-DiD estimation which is too small a sample to provide reliable estimates.

The second stage of our analysis is to consider the impact of other forms of international engagement that are more closely related to the previous literature on trade and the environment. Table 4 presents the PSM-DiD estimates on the causal effects of importing and exporting on CO_2 emission intensity. If a firm starts to import for the first time and is motivated to import an intermediate good (that is relatively dirty) to substitute for domestic production then one would expect a negative effect of importing on the growth of a firm's CO_2 intensity relative to the control group of non-importers. If a firm starts to export there is no reason to expect that, relative to the control group of non-exporters, this will impact CO_2 emission intensity at least not immediately. Hence the treatment in Table 4 is whether a firm is a new importer or a new exporter and has never previously imported or exported, respectively. The results show that new importers experience a 3.3% reduction in CO_2 intensity growth compared, to the control group, in the year of the treatment although this effect appears to be temporary. In terms of exporting, we find no significant effect on firms' emission intensity growth when firms enter the exports market and for the subsequent two years.[15]

Combined with the outsourcing results it is possible to infer that both overseas outsourcing and importing can improve the environmental performance of firms in Japan which suggests, especially in terms of production outsourcing, that firms are outsourcing the relatively pollution intensive part of their production process. The import and export results suggest that overseas outsourcing is a more important part of the story than international trade although our time period may be too short to capture a learning from exporting effect (Girma and Hanley, 2015; Forslid *et al.*, 2015 and Holliday, 2016) whereby over time exporting results in greater productivity and which in turn leads to greater investment in energy saving capital.

[Table 4 about here]

4.3 Impact of international activities on other performance outcomes

Finally, as part of a series of robustness checks, we return to the traditional outsourcing literature that suggests that outsourcing may affect other aspects of firm performance. In the following series of tables we examine the treatment effects of outsourcing on a range of other performance

[15] The literature on exporting and environmental performance argues that there is a learning effect from exporting where exposure to international markets may lead to technological spillovers that reduce emissions. Likewise, if a firm is exporting an intermediate good that is part of the global supply chain it is possible that the company being supplied will insist of certain environmental standards being met. In the case of Japan which is both a world leading in terms of the stringency of environmental regulations and in terms of being at the technological frontier.

outcomes, including sales, imports, exports, capital intensity, average wages, labour productivity and R&D intensity. Again, we concentrate on production outsourcing and distinguish between domestic-only outsourcing and overseas outsourcing. The results are presented in Table 5. We find that domestic-only outsourcing had no significant effect on sales, exports, imports or capital intensity although we do find a small positive and significant effect on average wages (a 6.7% growth compared to the control group) one year after the treatment, labour productivity (a 6.8% growth compared to the control group) and R&D intensity (a 0.3% growth compared to the control group and significant at the 10% level) two years after a firm starts to outsource to domestic firms. On the other hand, overseas outsourcing is found to have an immediate effect on the firms' exports and imports (a 40.5% growth in export value and 60.1% growth in import value compared to the control group at the time of treatment). This finding is consistent with Wagner (2011) who shows that compared to non-offshoring firms, German manufacturing enterprises that started offshoring in the years 2001-2003 had a higher share of exports in total sales in 2004 (41.7% for enterprises with offshoring vs 30.58% for those without offshoring). No significant treatment effects of overseas outsourcing is observed for other performance indicators such as sales, capital intensity, average wages, labour productivity or R&D intensity.

We also test the causal effects of importing and exporting on these outcomes. The results are reported in Table 6. In general, relative to the control group, we find no statistical significant treatment effects on these outcomes for importing or exporting except for a 3% increase of capital intensity for importing when firms start importing and a 2.8% increase in wages for the new exporters the year after entering the export market.

Finally, in order to improve their environmental performance, firms could import a greater share of parts or components from foreign countries instead of producing them by themselves. To test whether Japanese firms shift pollution to a developing countries with more lax environmental regulations Table 7 presents the PSM-DiD estimates for importing from China. Relative to the control group, no significant effect of importing from China on emission intensity growth is found in the year of treatment and up to two years later. This suggests that our importing results is driven by imports from a country other than China. Unfortunately data limitations means it is not possible to look at this for other developing country regions.

[Tables 5, 6 and 7about here]

5 Conclusions

In this paper we examine the impact of outsourcing on the emissions intensity of Japanese firms for 2009-2013. There are a number of reasons why a firm may choose to outsource part of the production process either domestically or overseas including to reduce firm costs, to gain access to better technologies or to access overseas markets. One little researched motivation is that firms use outsourcing to reduce their domestic production of the dirtier parts of their production process. The result would be for a firm to reduce its emissions on CO_2 per unit of output which may enable the firm to meet certain government targets or thresholds or simply to appease other stakeholders who want the firm to appear cleaner and "greener" and thus avoid any bad publicity associated with being a large polluter.

Using a PSM-DiD approach to control for a number of endogeneity concerns we investigate whether firms that start to outsource in a given year experience a reduction in their CO_2 emissions intensities in the year they start outsourcing and in the following two years in comparison to the year before they started outsourcing. Our results show that outsourcing does appear to reduce the emissions intensity per unit of sales for all firms in the year of outsourcing and the following years. When we make the distinction between domestic outsourcing and foreign outsourcing we find that domestic-only outsourcers experience no significant decrease in emissions intensities in the year of the treatment and the following year relative to the control group of non-outsourcers. However, we do find that foreign outsourcers experience a significant decrease in emissions intensity. For this group of firms the CO_2 emissions growth rate is 7.3% and 7.7% lower than the treatment group in the year of outsourcing and the following year relative to the control group of non-outsourcers.

Although we are not able to pinpoint the overall impact of outsourcing on aggregate pollution levels our results suggest that some Japanese firms are using foreign outsourcing as a mechanism by which they are able to reduce their overall emissions for a given value of sales and that this seems to be a larger driver of emissions reduction that either importing or exporting. Hence, this paper provides tentative evidence that is consistent with the pollution outsourcing hypothesis and this may be one explanation for a lack of convincing evidence for the more traditional pollution haven result that have traditional looked only at imports or FDI flows. One possible implication of our result is that if the stringency of environmental regulations in Japan is causing some firms to outsource the production of the relatively dirty parts of their production process to less regulated countries to meet the environmental quality target, then global air quality might be lower than if the firm had retained production locally and instead introduced cleaner technologies.

From a policy perspective one could argue that Japan or other developed countries could encourage the development of specialist domestic firms that are able produce pollution intensive intermediates efficiently and at scale which could lead to an overall reduction in global pollution without the need for firms to relocate or to outsource dirty production overseas. Such a firm is also likely to require skilled workers and to use relatively advanced technologies and thus enable Japan to maintain a leading position in eco-innovation.

References

Aichele, R. and Felbermayr, G. (2013). Estimating the effects of Kyoto on bilateral trade flows using matching econometrics, *World Economy*, 36(3), 303-330.

Antonietti, R., De Marchi, V., and Di Maria, E. (2016). Governing offshoring in a stringent environmental policy setting: Evidence from Italian manufacturing firms. *Journal of Cleaner Production*. http://dx.doi.org/10.1016/j.jclepro.2016.11.106.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25), 3083-3107.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25), 3083-3107.

Austin, P.C. (2009a). Some Methods of Propensity-Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo simulations. *Biometrical Journal*, 51(1), 171-184.

Austin, P.C. (2010). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10(2), 150-161.

Austin, P.C. (2013). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33(6), 1057–1069.

Austin, P. C., Grootendorst, P. and Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study. *Statistics in Medicine*, 26, 734-753. doi:10.1002/sim.2580.

Batrakova, S. and Davies, R.B. (2012). Is there an environmental benefit to being an exporter? Evidence from firm-level data. *Review of World Economics*, *148*(3), 449-474.

Brunel, C. (2016). Pollution offshoring and emission reductions in EU and US manufacturing, *Environmental and Resource Economics*, forthcoming.

Bustos, P. (2011). Trade liberalization, exports, and technology upgrading: Evidence on the impact of MERCOSUR on Argentinian firms. *American Economic Review*, 101(1), 304-340.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72.

Cadarso, M. Á., López, L. A., Gómez, N. and Tobarra, M. Á. (2010). CO 2 emissions of international freight transport and offshoring: measurement and allocation. *Ecological Economics*, *69*(8), 1682-1694.

Cherniwchan, J., Copeland, B. R. and Taylor, M. S. (2016). Trade and the environment: New methods, measurements, and results. *National Bureau of Economic Research Working Paper No. 22636*.

Chiu, S-T. (1991). Bandwidth selection for Kernel density estimation. *Annuals of Statistics*, 19(4), 1883-1905.

Clark, D. P., Serafino, M. and Simonetta, Z. (2000). Do dirty industries conduct offshore assembly in developing countries? *International Economic Journal*, 14(3), 75-86.

Chung, S. (2014). Environmental regulation and foreign direct investment: Evidence from South Korea. *Journal of Development Economics*, 108, 222-236.

Cole, M.A., Elliott, R.J.R. and Shimamoto, K. (2006). Globalization, firm-level characteristics and environmental management: A study of Japan. *Ecological Economics*, 59(3), 312-323.

Cole, M.A., Elliott, R.J.R., Okubo, T. and Zhou, Y. (2013). The carbon dioxide emissions of firms: A spatial analysis. *Journal of Environmental Economics and Management*, 65(2), 290-309.

Cole, M.A., Elliott, R.J.R. and Zhang, L. (2017). Foreign direct investment and the environment: A review of the literature. *Annual Review of Environment and Resources*, 42.

Cui, J., Lapan, H. and Moschini, G. (2012). Are exporters more environmentally friendly than non-exporters? Theory and evidence. *Economics Working Papers (2002–2016) No.* 77. http://lib.dr.iastate.edu/econ_las_workingpapers/77.

Cui, J., Lapan, H. and Moschini, G. (2015). Productivity, export, and environmental performance: Air pollutants in the United States. *American Journal of Agricultural Economics*, 1-21; doi: 10.1093/ajae/aav066.

Dehejia, R.H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1), 151-161.

Elliott, R.J.R., Jabbour, L. and Zhang, L. (2016). Firm productivity and importing: Evidence from Chinese manufacturing firms. *Canadian Journal of Economics*, 49(3), 1086-1124.

Elliott, R.J.R. and Okubo, T. (2016), Ecological Modernization in Japan: The Role of Interest Rate Subsidies and Voluntary Pollution Control Agreements, *Asian Economic Papers*, 3(15), 66-88.

Flury, B. K. and Riedwyl, H. (1986). Standard distance in univariate and multivariate analysis. *The American Statistician*, 40(3), 249-251.

Forslid, R., Okubo, T. and Ulltveit-Moe, K. H. (2015). Why are firms that export cleaner? International trade, Abatement and Environmental Emissions. *CEPR Discussion Paper No. 8583*.

Fowlie, M., S.P. Holland and E.T. Mansur. (2012). What do emissions markets deliver and to whom? Evidence from Southern California's NOx trading program. *American Economic Review*, 102(2), 965-993.

Garrido, M. M., Kelley, A.S., Paris, J., Roza, K., Meier, D.E., Morrison, R.S. and Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health Services Research*, 49(5), 1701-1720.

Girma, S. and Gorg, H., 2004. Outsourcing, foreign ownership, and productivity: evidence from UK establishment-level data. *Review of International Economics*, 12 (5), 817-832.

Girma, S., and Gorg, H. (2007). Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *Journal of International Economics*, 72, 97-112.

Girma, S., and Hanley, A. (2015). How Green are Exporters? *Scottish Journal of Political Economy*, 62 (3), 291-309.

Gorg, H., Hanley, A. and Strobl, E., 2008. Productivity effects of international outsourcing: evidence from plant-level data. *Canadian Journal of Economics*, 41(2), 670-688

Hamamoto, M. (2006). Environmental regulation and the productivity of Japanese manufacturing industries. *Resource and Energy Economics*, 28(4), 299-312.

Hanley, A. and Pérez, J. M. (2012). Are newly exporting firms more innovative? Findings from matched Spanish innovators. *Economics Letters*, 116(2), 217-220.

Heyman, F., F. Sjöholm and P. G. Tingvall (2007). Is there really a foreign ownership wage premium? Evidence from matched employer–employee data. *Journal of International Economics*, 73(2), 355-376.

Heckman, J. J., Ichimura, H. and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4), 605-54.

Heckman, J, J., Ichimura, H. and Todd, P. E. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2), 261-294.

Ho, D. E., Imai, K., King, G. and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199-236.

Holladay, J.S. (2016). Exporters and the environment. *Canadian Journal of Economics*, 49(1), 147-172.

Huttunen, K. (2007). The effect of foreign acquisition on employment and wages: Evidence from Finnish establishments. *Review of Economics and Statistics*, 89(3), 497-509.

Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economic and Statistics*, 86(1), 4-29.

Imura, H. and Schreurs, M.A. (Eds.). (2005). *Environmental policy in Japan*. Cheltenham: Edward Elgar Publishing.

Ito, B. and Tanaka, A. (2016). External R&D, productivity, and export: evidence from Japanese firms. *Review of World Economics*, 152(3), 577-596.

Jinji, N. and Sakamoto, H. (2015). Does exporting improve firms' CO2 emissions intensity and energy intensity? Evidence from Japanese manufacturing. *RIETI Discussion Paper Series 15-E-130*.

Kellenberg, D.K. (2009), An empirical investigation of the pollution haven effect with strategic environment and trade policy, *Journal of International Economics*, 78(2), 242-255.

Lechner, M. and Wunsch, C. (2013). Sensitivity of matching-based program evaluations to the availability of control variables. *Labour Economics*, 21, 111-121

Leoncini, R., Montresor, S. and Rentocchini, F. (2016). CO2-reducing innovations and outsourcing: Evidence from photovoltaics and green construction in North-East Italy. *Research Policy*, 45(8), 1649–1659.

Levinson, A. (2010). Offshoring pollution: Is the United States increasingly importing polluting goods? *Review of Environmental Economics and Policy*, 4(1), 63-83.

Li, X. and Zhou, Y. M. (2017). Offshoring pollution while offshoring production? *Strategic Management Journal*. DOI: 10.1002/smj.2656

Liu, X. and Lynch, L. (2011). Do agricultural land preservation programs reduce farmland loss? Evidence from a propensity score matching estimator. *Land Economics*, 87(2), 183-201.

List, J.A., Millimet, D.L., Fredriksson, P.G. and McHone, W.W. (2003). Effects of environmental regulations on manufacturing plant births: evidence from a propensity score matching estimator. *Review of Economics and Statistics*, 85(4), 944-952.

List, J.A., McHone, W.W. and Millimet, D.L. (2004). Effects of environmental regulation on foreign and domestic plant births: is there a home field advantage? *Journal of Urban Economics*, 56(2), 303-326.

Lyu, Y. (2016). Evaluating carbon dioxide emissions in undertaking offshored production tasks: the case of China. *Journal of Cleaner Production*, 116, 32-39.

McAusland, C. and Millimet, D.L. (2013) Do national borders matter? Intranational trade, international trade, and the environment. *Journal of Environmental Economics and Management*, 65 (3), 411-437.

Michel, B. (2013). Does offshoring contribute to reducing domestic air emissions? Evidence from Belgian manufacturing. *Ecological Economics*, *95*, 73-82.

OECD. (1997). Environmental Policies in Japan. OECD, Paris

OECD. (2002). OECD Environmental Performance Reviews: Japan 2002. OECD, Paris

OECD. (2010). OECD Environmental Performance Reviews: Japan 2010. OECD, Paris

Rezza, A.A. (2015). A meta-analysis of FDI and environmental regulations. *Environment and Development Economics*, 20(02), 185-208.

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

Rosenbaum, P. and Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity. *American Statistician*, 39(1), 33–38.

Rubin, D.B. (2001). Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services and Outcomes Research Methodology*, 2, 169-188.

Sheather, S.J. (2004). Density estimation. *Statistical Science*, 19(4), 588–597.

Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics*, 15:370–399

Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1-21.

Tomiura, E. (2005). Foreign outsourcing and firm-level characteristics: Evidence from Japanese manufacturers. *Journal of the Japanese and International Economies*, 19(2), 255-271.

Tomiura, E. (2007). Foreign outsourcing, exporting, and FDI: A productivity comparison at the firm level. *Journal of International Economics*, 72(1), 113-127.

Villanti A.C., Cullen, J., Vallone D.M. and Stuart, E.A. (2011). Use of propensity score matching to evaluate a national smoking cessation media campaign. *Evaluation Review*, 35(6), 571-591

Wagner, J. (2011). Offshoring and firm performance: self-selection, effects on performance, or both? *Review of World Economics*, *147*(2), 217-247.

Yasar, M. and Rejesus, R.M. (2005). Exporting status and firm performance: Evidence from a matched sample. *Economic Letters*, 88, 397-402.

Table 1: Summary statistics for our key variables by outsourcing activity

	(1) Full sample	(2) Non-OS	(3) OS	(4) OS: service & production	(5) OS: service only	(6) OS: production only	(7) OS: domestic & foreign production	(8) OS: domestic production only	(9) OS: foreign production only
CO2	107,493.54	166,373.54	92,912.91	99,895.43	298,719.47	37,139.44	28,626.38	38,160.64	17,914.98
	(1,555,896.01)	(2,609,516.63)	(1,154,687.88)	(641,249.55)	(3,186,267.64)	(400,011.93)	(83,907.10)	(420,455.64)	(22,957.65)
rsales	60,473.06	64,032.90	59,591.52	79,367.45	72,680.09	36,598.95	48,792.98	35,347.91	39,037.53
	(284,680.53)	(297,149.56)	(281,510.35)	(365,688.56)	(248,428.35)	(166,307.82)	(174,687.50)	(166,038.34)	(52,438.48)
co2sales	1.96	2.38	1.86	1.70	3.30	1.69	0.93	1.77	1.24
	(6.68)	(5.65)	(6.91)	(8.00)	(9.03)	(4.80)	(1.23)	(5.03)	(2.14)
age	50.77	50.06	50.95	50.98	45.29	52.26	56.84	51.79	53.00
	(22.23)	(22.08)	(22.26)	(23.37)	(22.68)	(20.76)	(20.51)	(20.69)	(25.08)
emp	901.72	974.93	883.58	1,109.96	909.16	649.82	918.29	622.29	701.64
	(2,792.55)	(3,405.89)	(2,618.38)	(3,158.85)	(2,367.19)	(1,976.66)	(1,955.66)	(1,983.83)	(811.51)
wage	5,121.89	5,564.95	5,012.18	6,634.14	4,968.33	3,391.06	4,914.94	3,227.12	4,597.43
	(19,776.57)	(23,672.81)	(18,685.92)	(23,969.44)	(13,852.30)	(12,431.01)	(12,709.50)	(12,430.60)	(6,589.13)
KL	16.73	16.83	16.70	16.95	24.06	14.71	12.64	14.92	14.09
	(22.48)	(19.43)	(23.17)	(19.45)	(43.21)	(19.11)	(9.90)	(19.85)	(9.03)
LP	53.90	51.29	54.55	56.62	88.80	44.35	39.62	44.77	50.41
	(88.77)	(63.53)	(93.97)	(80.82)	(220.01)	(37.30)	(24.62)	(38.30)	(38.84)
EXP	0.43	0.30	0.46	0.52	0.32	0.42	0.78	0.38	0.85
	(0.49)	(0.46)	(0.50)	(0.50)	(0.47)	(0.49)	(0.42)	(0.49)	(0.36)
IMP	0.35	0.24	0.38	0.45	0.30	0.32	0.79	0.27	0.85
	(0.48)	(0.42)	(0.49)	(0.50)	(0.46)	(0.47)	(0.40)	(0.45)	(0.36)
FOR	0.02	0.03	0.02	0.02	0.04	0.01	0.03	0.01	0.02
	(0.15)	(0.18)	(0.15)	(0.15)	(0.20)	(0.12)	(0.18)	(0.11)	(0.14)
RD	0.60	0.51	0.62	0.67	0.54	0.59	0.74	0.57	0.81
	(0.49)	(0.50)	(0.49)	(0.47)	(0.50)	(0.49)	(0.44)	(0.50)	(0.39)
EXPshare	0.07	0.06	0.08	0.10	0.06	0.06	0.14	0.05	0.25
	(0.16)	(0.15)	(0.17)	(0.19)	(0.14)	(0.15)	(0.20)	(0.14)	(0.20)
IMPshare	0.03	0.03	0.03	0.04	0.05	0.03	0.06	0.02	0.13
	(0.09)	(0.09)	(0.09)	(0.09)	(0.13)	(0.08)	(0.11)	(0.07)	(0.13)
FORshare	0.04	0.05	0.04	0.05	0.05	0.03	0.05	0.03	0.03
	(0.15)	(0.17)	(0.14)	(0.15)	(0.17)	(0.12)	(0.16)	(0.11)	(0.14)
RDshare	0.02	0.01	0.02	0.02	0.03	0.01	0.02	0.01	0.03
	(0.17)	(0.03)	(0.19)	(0.04)	(0.57)	(0.03)	(0.03)	(0.03)	(0.10)
Obs	19,503	3,871	15,632	7,011	1,651	6,970	634	6,283	53
(% of total)	100.00	19.85	80.15	39.95	8.47	35.74	3.25	32.22	0.27

Notes: We report the means and standard deviations in parentheses. See Table A1 in Appendix for detailed definition of the variables. OS: Outsourcing.

Table 2: Effects of production outsourcing on firms' CO_2 intensity (PSM-DiD estimates)

	s=0	*s=1*	*s=2*
Gaussian Kernel matching			
ATT	-0.051**	-0.066*	-0.095*
	(0.026)	(0.034)	(0.052)
N(T)	173	113	54
N(C)	1060	619	263
Radius matching			
ATT	-0.052*	-0.062*	-0.095*
	(0.027)	(0.036)	(0.054)
N(T)	165	105	51
N(C)	848	484	202

Notes: Standard errors in parentheses. N(T) and N(C) are the numbers of observations for the treated and control groups respectively. ***, ** and * denote significance at 0.01, 0.05 and 0.1 respectively.

Table 3: Effects of production outsourcing on firms' CO_2 intensity by outsourcing destination

Treatment	*s=0*	*s=1*	*s=2*
Domestic-only outsourcing			
ATT	-0.013	-0.025	-0.086*
	(0.025)	(0.038)	(0.049)
N(T)	225	147	73
N(C)	1071	627	253
Foreign outsourcing			
ATT	-0.073***	-0.077*	-0.007
	(0.024)	(0.041)	(0.075)
N(T)	114	76	31
N(C)	915	512	156

Notes: Standard errors in parentheses. N(T) and N(C) are the numbers of observations for the treated and control groups respectively. ***, ** and * denote significance at 0.01, 0.05 and 0.1 respectively.

Table 4: Effects of importing/exporting on firms' CO_2 intensity (PSM-DiD estimates)

Treatment	*s=0*	*s=1*	*s=2*
Importing			
ATT	-0.033*	-0.032	0.022
	(0.018)	(0.028)	(0.055)
N(T)	150	107	50
N(C)	4243	2544	1033
Exporting			
ATT	0.004	0.010	0.013
	(0.020)	(0.034)	(0.059)
N(T)	160	94	40
N(C)	4645	2529	1027

Notes: Standard errors in parentheses. N(T) and N(C) are the numbers of observations for the treated and control groups respectively. ***, ** and * denote significance at 0.01, 0.05 and 0.1 respectively.

Table 5 Effects of production outsourcing on firms' other performance indicators

Treatment	Domestic-only outsourcing			Overseas outsourcing		
Outcome	*s=0*	*s=1*	*s=2*	*s=0*	*s=1*	*s=2*
1) logsales						
ATT	0.004	0.0004	0.015	0.027	0.002	-0.023
	(0.019)	(0.031)	(0.030)	(0.019)	(0.033)	(0.053)
N(T)	226	147	73	114	76	31
N(C)	1072	639	263	911	508	152
2) logexp						
ATT	0.173	0.153	0.318	0.405*	0.332	0.400
	(0.157)	(0.188)	(0.244)	(0.212)	(0.318)	(0.287)
N(T)	226	147	73	114	76	31
N(C)	1074	633	258	918	513	158
3) logimp						
ATT	0.145	0.207	0.485	0.609**	0.490	0.902
	(0.149)	(0.219)	(0.364)	(0.290)	(0.421)	(0.620)
N(T)	226	147	73	114	76	31
N(C)	1072	640	264	931	526	169
4) logKL						
ATT	-0.014	-0.002	0.070	0.0005	-0.041	0.085
	(0.033)	(0.044)	(0.046)	(0.031)	(0.103)	(0.092)
N(T)	226	147	73	114	75	31
N(C)	1073	652	275	928	519	155
5) logwage						
ATT	0.022	0.067**	0.054	-0.034	-0.045	0.067
	(0.024)	(0.032)	(0.044)	(0.037)	(0.060)	(0.101)
N(T)	226	147	73	114	76	31
N(C)	1073	641	265	941	534	177
6) logLP						
ATT	0.021	0.039	0.068**	0.028	0.032	0.058
	(0.016)	(0.024)	(0.028)	(0.019)	(0.032)	(0.049)
N(T)	226	147	74	114	76	31
N(C)	1075	650	274	941	534	177
7) logRDshare						
ATT	-0.0006	0.001	0.003*	0.004	0.0003	0.002
	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)	(0.004)
N(T)	226	147	73	114	76	31
N(C)	1071	639	264	932	527	168

Notes: Standard errors in parentheses. N(T) and N(C) are the numbers of observations for the treated and control groups respectively. ***, ** and * denote significance at 0.01, 0.05 and 0.1 respectively.

Table 6: Effects of importing/exporting on firms' other performance

Treatment	Importing			Exporting		
Outcomes	*s=0*	*s=1*	*s=2*	*s=0*	*s=1*	*s=2*
1) logKL						
ATT	-0.030*	-0.035	-0.048	-0.005	0.006	0.034
	(0.018)	(0.025)	(0.041)	(0.013)	(0.024)	(0.035)
N(T)	196	142	63	160	95	40
N(C)	5518	3340	1492	4593	2560	1020
2) logwage						
ATT	-0.002	0.039	0.075	0.001	0.028*	-0.098
	(0.030)	(0.038)	(0.083)	(0.028)	(0.031)	(0.089)
N(T)	150	107	50	125	76	35
N(C)	4176	2482	972	3461	1961	760
3) logLP						
ATT	0.001	-0.013	-0.049	0.011	0.020	0.040
	(0.018)	(0.026)	(0.033)	(0.017)	(0.027)	(0.031)
N(T)	150	107	49	125	76	35
N(C)	4218	2522	1010	3451	1950	770
4) logRDshare						
ATT	-0.001	-0.0006	-0.003	0.001	0.001	-0.002
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
N(T)	150	107	49	125	77	35
N(C)	4215	2521	1010	3435	1942	756

Notes: Standard errors in parentheses. N(T) and N(C) are the numbers of observations for the treated and control groups respectively. ***, ** and * denote significance at 0.01, 0.05 and 0.1 respectively.

Table 7: Effects of importing from China on firms' CO_2 intensity

	s=0	*s=1*	*s=2*
PSM-DiD estimates			
ATT	-0.009	-0.028	0.024
	(0.017)	(0.027)	(0.051)
N(T)	187	132	58
N(C)	5161	3027	1119

Notes: Standard errors in parentheses. N(T) and N(C) are the numbers of observations for the treated and control groups respectively.

Appendix

Table A1: Definition of variables

Variable	Definition
OS	outsourcing dummy which equals one if a firm outsources, 0 otherwise
newOS	a dummy variable which equals one if a firm starts to outsource, 0 otherwise
EXP	export dummy which equals one if a firm exports, 0 otherwise
newEXP	a dummy variable which equals one if a firm starts to export, 0 otherwise
IMP	import dummy which equals one if a firm imports, 0 otherwise
newIMP	a dummy variable which equals one if a firm starts to import, 0 otherwise
FOR	foreign ownership dummy which equals one if the share of foreign capital to total capital is 50% or more, 0 otherwise
RD	a dummy variable which equals one if a firm has positive R&D expenditure, 0 otherwise
FDI	a dummy variable which equals one if a firm has one or more subsidiaries overseas, 0 otherwise
EXPshare	the share of import value over total sales
IMPshare	the share of export value over total sales
FORshare	the share of foreign capital over total capital
logsales	log of a firm's real total sales
co2sales	CO_2 emission intensity of a firm which is estimated by CO_2 emissions divided by real total sales
logco2sales	log of CO_2 emission intensity
pregrowth	the growth rate of CO_2 emissions intensity before the treatment
logage	log of a firm's age which is calculated as (survey year-foundation+1)
logemp	log of a firm's number of employees
logKL	log of a firm's capital intensity calculated as real tangible assets divided by the number of employees
logwage	log of a firm's average wage of the employees
logLP	log of a firm's labour productivity estimated as real total sales divided by the number of employees
logRDshare	log of (a firm's R&D intensity +1), R&D intensity is the share of R&D expenditure over sales

Table A2: Balancing test after matching I

	Unmatched	Mean		SD		t-test		V(T)/V(C)
Variable	Matched	Treated	Control		bias	t	p>t	
pregrowth	U	-0.05811	-0.00911	-17.3		-2.08	0.037	1.35
	M	-0.05811	-0.05555	-0.9	94.8	-0.08	0.939	1.32
logage	U	3.7486	3.7545	-0.9		-0.16	0.876	1.06
	M	3.771	3.8225	-7.8	-768.6	-0.77	0.442	1.24
logemp	U	5.7234	5.7416	-1.5		-0.25	0.8	0.80
	M	5.6878	5.6136	6.3	-307.1	0.57	0.567	1.11
logKL	U	2.3742	2.3587	1.6		0.27	0.79	0.92
	M	2.4439	2.3877	5.7	-262.7	0.51	0.61	0.84
logwage	U	1.5767	1.5294	11.1		1.87	0.062	0.92
	M	1.596	1.5821	3.2	70.7	0.3	0.763	0.95
logexp	U	2.6326	2.1045	14.4		2.48	0.013	1.04
	M	2.5429	2.1185	11.5	19.6	1.02	0.31	1.08
logimp	U	2.092	1.5302	17.6		3.12	0.002	1.2
	M	1.793	1.4728	10	43	0.9	0.371	1.24
FOR	U	0.04179	0.0327	4.8		0.86	0.389	.
	M	0.03425	0.02595	4.4	8.8	0.41	0.68	.
RD	U	0.58209	0.50234	16		2.74	0.006	.
	M	0.56164	0.4801	16.4	-2.2	1.39	0.164	.
FDI	U	0.0806	0.08365	-1.1		-0.19	0.85	.
	M	0.06849	0.06459	1.4	-27.7	0.13	0.894	.

Notes: Year and sector dummy variables not presented in the table but included in the balancing tests. Standardized difference is 0 and p-value of t-test is 1 for the matched sample for each of these dummies

Table A3: Balancing test after matching II

Sample	Pseudo R^2	Likelihood Ratio Chi^2	$p>Chi^2$	MeanBias	MedianBias	B	R	%Var
Unmatched	0.045	40.52	0.076	8.6	5.9	57.9	0.88	29
Matched	0.011	4.6	1	2	0	25.1	0.97	0

Notes: Rubins' B is the absolute standardized difference of the means of the linear index of the propensity score in the treated and (matched) non-treated group and Rubin's R is the ratio of treated to (matched) non-treated variances of the propensity score index. Rubin (2001) recommends that B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced.