International Institute of Social Studies

Erasmus

Working Paper
No. 688

Effectiveness of export promotion programmes

Binyam Afewerk Demena

September 2021

ISSN 0921-0210

The International Institute of Social Studies is Europe's longest-established centre of higher education and research in development studies. On 1 July 2009, it became a University Institute of the Erasmus University Rotterdam (EUR). Post-graduate teaching programmes range from six-week diploma courses to the PhD programme. Research at ISS is fundamental in the sense of laying a scientific basis for the formulation of appropriate development policies. The academic work of ISS is disseminated in the form of books, journal articles, teaching texts, monographs and working papers. The Working Paper series provides a forum for work in progress which seeks to elicit comments and generate discussion. The series includes academic research by staff, PhD participants and visiting fellows, and award-winning research papers by graduate students.

Working Papers are available in electronic format at www.iss.nl

Please address comments and/or queries for information to:

Institute of Social Studies
P.O. Box 29776
2502 LT The Hague
The Netherlands

or

E-mail: wpapers@iss.nl

Table of Contents

Abstract

Export promotion policies and programmes (EPP) are increasingly popular to enhance export performance, but the evidence from their reported estimates appear puzzling and yet vary widely. We examine 1869 estimated parameters dealing with EPP and firm-level export performance from 37 studies published up to including 2020. Our main findings are threefold. First, constructing 26 moderator variables reflecting the context in which researchers obtain their estimated parameters, we uncovered that differences across the primary studies are mainly driven by the characteristics of the data, the types of firms targeted, the set of variables controlled in the underlying estimation techniques, the adopted a four-dimensional view of export performance, and the publication characteristics. Second, controlling for publication selection bias and reducing potential endogeneity issues, the implied gains from trade-promotion polices is about 0.069, suggesting a small practical impact by the existing guidelines. Third, unlike the econometric evaluation technique, we find robust evidence that the firm-extensive and destination-extensive margins appeared to be associated with mediating factors of publication bias.

Keywords

International trade, trade policy, export promotion, firm export performance, impact evaluation, meta-analysis

JEL codes: F13, F14, L25, O55

Acknowledgements

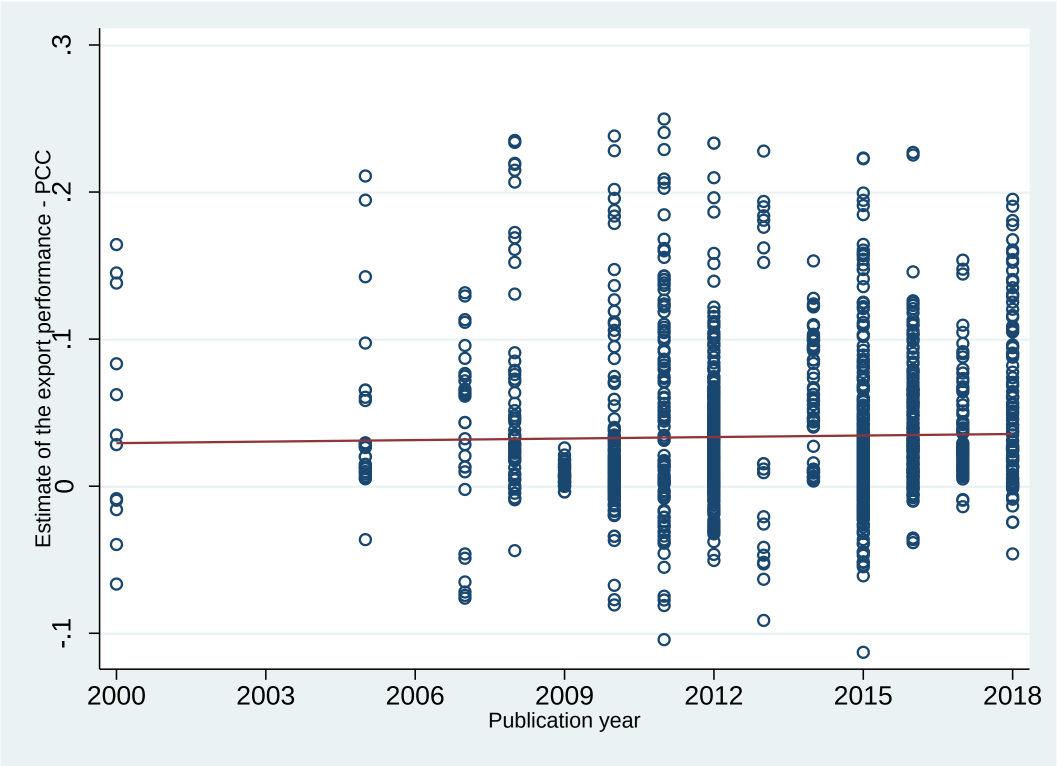
The author acknowledges the financial support of the Development Economics research group and DEC RIF 2020/4.

Effectiveness of export promotion programmes

1. Introduction

Entering export markets incurs multiple obstacles. One of the most prominent obstacles is imperfect information, that is both informational and institutional barriers (Martincus and Carballo, 2008; Cruz et al., 2018). Another major obstacle to exporting is firm-level low productivity (Kim et al., 2018; Demena et al, 2021b). According to the widely utilized trade model with heterogeneous firms developed by Meltiz (2003), low productive firms choose to become exporters if their productivity is sufficiently high enough to cover the initial costs of exporting, i.e., fixed cost (Broocks and van Biesebroeck, 2017; Kim et al., 2018). In order to ease the supply-side constraints of entering export markets, various policy measures have been utilised (Martincus and Carballo, 2010a; Cadot et al., 2015; Cruz et al., 2018; Kim et al., 2018; Munch and Schaur, 2018). Publicly funded export promotion policies and programmes (EPP) that mainly created in the late 1990s belong to this group of policy measures, (Martincus and Carballo, 2008; Cruz et al., 2018).

Figure 1
Reported EPP impacts diverge over the period 2000 – 2020 (N=1869)



Notes: The figure depicts the impact of export promotion programmes on firm-level export performance estimates reported in the individual empirical studies from 2000 and 2020 inclusive. The horizontal axis measures the year when the first drafts of the primary studies appeared in Google Scholar.

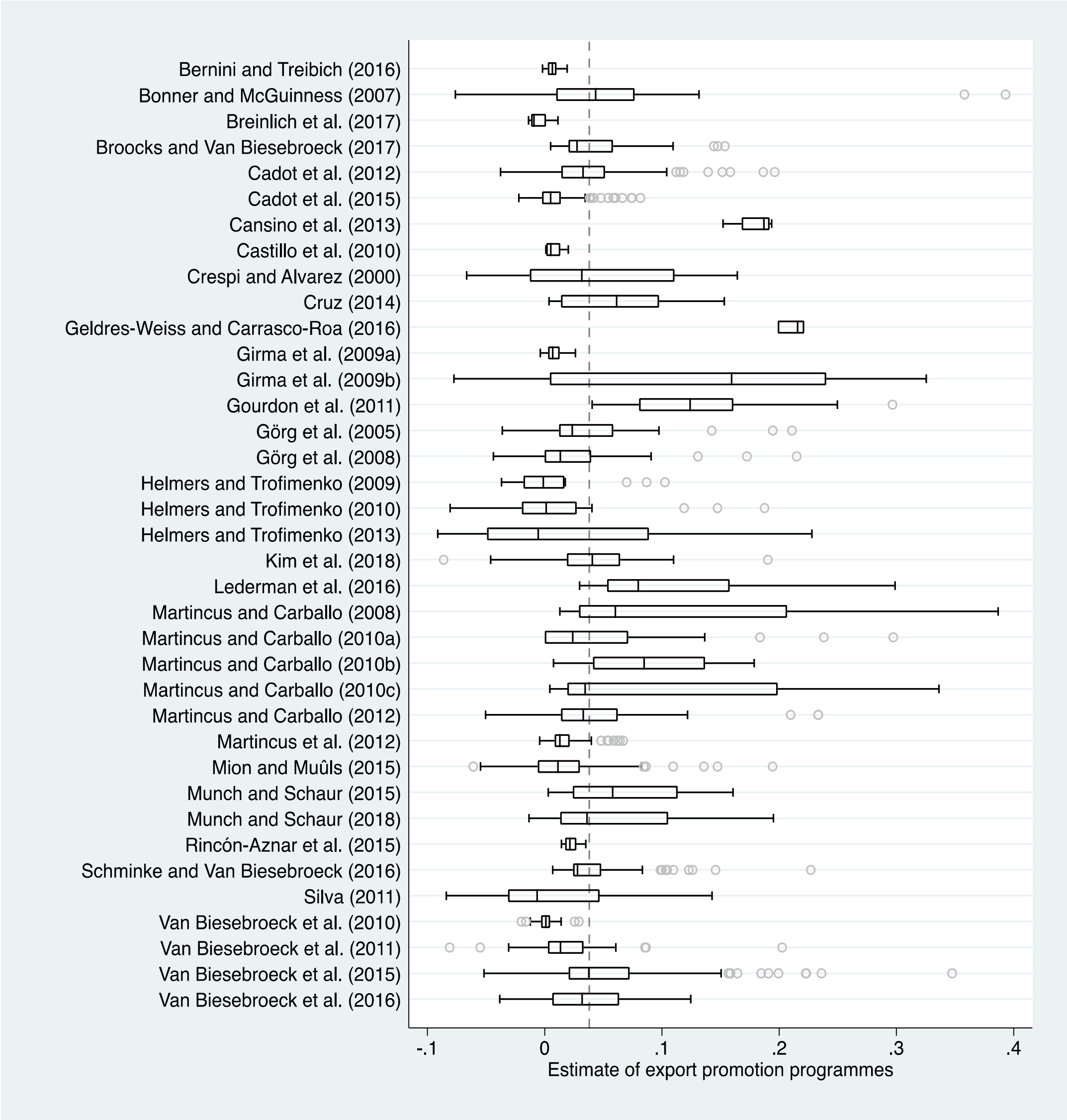
Most prominent promotion programmes help domestic firms to lower initial costs of exporting through lowering informational barriers. Specifically, through providing information on foreign markets for instance to adapt products to foreign tastes, matching firms to clients and organizing missions and trade fairs, assisting firms to find a distributor or promising channels of new distribution, establish relationships with importers, navigate foreign customs and product regulations, or export procedures among others (Cadot, 2015; van Biesebroeck et al., 2015; Broocks and van Biesebroeck, 2017; Cruz et al., 2018). They also help domestic firms to overcome trade or administrative frictions, product standards, foreign regulations or simplify customs procedures for example introducing e-customs procedures to be completed online (van Biesebroeck et al., 2016; Kim et al., 2018). Moreover, other most direct measures to help domestic firms succeed in export markets are ranging from subsidies and grants to lower tax rates for export earnings (Girma et al., 2009b). Production subsidies or grants may help firms to deal with particular market difficulties or overcome the sunk costs related to entering export markets (Görg et al., 2008; Girma et al., 2008a; Silva 2011; Cadot et al., 2015).

The hypothesis of such policies and programmes assumes that if EPP lowers informational and institutional barriers or covers initial costs of exporting, thus it can generate firms with better international experience. However, in the international trade literature, the actual evidence for firm-level export performance associated with EPP has yielded substantial disagreement both in terms of effect size magnitude and their statistical significance, and, hence, is far from definite and still lively debated (see also, van Biesebroeck et al., 2015; Geldres-Weiss and Carrasco-Roa, 2016; Broocks and van Biesebroeck, 2017; Kim et al., 2018; Munch and Schaur, 2018). We have documented the large numbers of empirical studies available in the period 2000 - 2020. This review has identified 1869 reported estimates from 37 primary empirical studies published over the last two decades carried out in 21 countries[1]. Figure 1 shows a bird's-eye survey of the development of the reported estimates overtime. There is substantial variation in the estimated parameters, in particular in the studies published in the last decade, also because the number of studies increased substantially. The reported estimates show a consistent disagreement or divergent on the value of EPP instead of converging to a consensus value.

Figure 2 gives a box plot of the 1869 estimated parameters reported in individual empirical studies. Zooming into the individual studies, reported estimates obviously vary widely both within and across the included empirical studies. However, it is seeming that positive impacts of the trade promotion policies are largely reported across the individual studies as opposed to negative ones. It is also apparent that majority of the included studies generated estimates at least 0.038, which is the overall arithmetic average of the estimated parameters represented by the short-dashed vertical line. While these estimates reflect genuine evidence of the polices driven by stimulating firms to export, it could also be due to differences in researchers' choices about econometric models, data, estimation techniques among others. In other words, publication selection bias could be one of the potential sources of the variance documented in Figure 2. In terms of the direction and significance of the reported magnitude, Figure A1 in the appendix also shows somehow contrasting findings. This growing body of international trade literature reported in slightly more than half (54%) a significantly positive effect, whereas in approximately 4% a negative and significant effect. The other 42% report both negative and positive but insignificant impact of EPPs on firm export performance.

[1] These are Argentina, Belgium, Bolivia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Denmark, Ecuador, France, Germany, Ireland, Peru, Portugal, Spain, Tunisia, UK, Uruguay and Vietnam.

Figure 2
Reported estimates vary widely both within and across the primary studies (N=1869)



Notes: The figure shows a box plot of the estimates of export promotion programmes on firm-level export performance reported in individual primary studies in the period 2000 and 2020. Following Tukey (1977), the box shows interquartile range (P25–P75), the lower limit the P25 (Q1) and the upper limit the P75 (Q3) with median represented by vertical line within the boxes. Horizontal whiskers cover the interval from (P25 - 1.5 times the interquartile range) to (P75 + 1.5 times the interquartile range) (Tukey, 1977). Any dots should show the remaining (outlying) reported estimates in each primary studies. The short dashed vertical line represents the arithmetic mean effect of the 1869 estimates reported in the primary empirical studies.

The apparent increase in variance of the reported estimates overtime (Figure 1) and the substantial heterogeneity in both within and between studies (Figure 2), provide the rational for systematically and comprehensively evaluating the diverse reported estimates of trade promotion policies. For this evaluation, we use a meta-analysis method to combine, summarize and investigate the estimated parameters. Meta-analysis is a useful statistical method to examine the inconsistent results found in reported empirical studies investigating a similar hypothesis, research question, or empirical effect (Stanley and Doucouliagos, 2012; Demena and van Bergeijk, 2017). Recent applications in international economics include Iršová and Havránek (2013), Demena (2015), Havránek et al. (2016), Afesorgbor (2017), Moons and van Bergeijk (2017), Demena and Bergeijk (2017), van Bergeijk et al. (2019), Bajzik et al. (2020) and Demena and Afesorgbor (2020). Although the current subject is very relevant to policy making to understand the outcomes of policies seeking trade promotion, there has been no meta-analysis, that we are aware of, to examine the impact of government-designed export promotion on export performance.[2] An important motivation to use the method of meta-analysis is, therefore, to investigate effectiveness of government-designed EPPs in generating trade promotion goal achievement, adopting a four-dimensional view of firm-level export performance.[3]

In finding out the underlying genuine effect of EPP, meta-analysis also allows to investigate the potential for publication bias in the reported estimates. According to Doucouliagos et al. (2005:321), publication bias is an "… often covert form of bias in empirical research arising when the selection of studies for publication is made on the basis of the statistical significance of results, and/or whether the results satisfy preconceived theoretical expectations". It is considered as the result of selecting research findings or results for their significant statistical results or whether findings satisfy prior theoretical expectations. Results of significant statistical findings can be treated more favourably by reviewers, editors or empirical researchers which may lead to over-representation of more significant effects or a larger given effect (Demena, 2017). Findings with a smaller given effect, or with statistically insignificant effects, however, tend to remain or hidden in the desk drawer and hence under-represented (Stanley, 2008). If such bias exists, it can potentially lead to distorted empirical research inferences, both in terms of policymakers' decisions and scientific conclusions. Filtering out such potential publication selection bias will therefore allow to isolate the overall 'true effect' or 'genuine effect' of EPP.

Beyond the issue of publication bias, it is essential to investigate the drivers of heterogeneity across the estimated parameters. This is evident by the within and between studies heterogeneity reported in Figure 2 and by the variance of the individual reported estimates presented in Figure 1. To do so, we construct 26 potential sources of moderator variables that reflect the context in which the empirical researchers produce their 1869 reported estimates.[4] Dubbed the analysis of the analyses, the meta-analysis also allows to trace variations in reported estimates and helps to understand how the underlying empirical studies were conducted and map the impact of the empirical framework regarding researchers' choices about the characteristics of data, estimation, specification, and econometric models onto estimated parameters (Stanley and Doucouliagos, 2012; Demena, 2017; Balima et al., 2020). In our case, the meta-analysis allows to explore and understand the reasons behind the diverse findings of the effectiveness of publicly funded-export promotion agencies reported in the empirical studies. Not only to explain the

---

[2] Under the umbrella of economic diplomacy, Moons and van Bergeijk (2017) meta-analyzed the impact of various instruments of international diplomatic relation measures, such as state visits, embassies, consulates, trade missions, investment and export promotion agencies on overall or macroeconomic flows of trade and investment, but not on firm-level EPPs considering programme recipients directly. Yet, the macroeconomic or aggregated approach cannot identify EPPs impact at firm level that actually benefited from the programmes (Cruz et al., 2018).

[3] The evaluation of the impact of EPPs is conducted through four dimensions of firm-level export performance. This refers to increasing the level of export in markets that are served by existing exporters only (the intensive margin); whether intervention draws a new firm into the export market (the firm-extensive margin); whether firms diversify their exports and reach different product or destination markets, i.e., if programme affected the destination and product extensive margins. In short, the existing studies investigated whether trade promotion polices affect the intensive margin, the firm-extensive margin, destination-extensive margin, and product-extensive margin.

[4] See section 3.3, for extensive discussion regarding the potential reasons behind the various sources of heterogeneity reported in the context of this literature.

variation under which trade promotion policies generate successful exporters, but also allows to provide guidelines to future researchers as to how this research should be designed. In turn, it informs policymakers with the underlying genuine effect and specific conditions under which such programmes may generate better export performance outcomes, and thus enabling them to make better policy decisions. It is important to note that most countries implemented EPP with a significant amount of public resources (and also recently many countries are mandated to promote EPP[5]), a systematic and comprehensive meta-regression analysis (MRA) is necessary and justifiable to understand the results of such policy measures.

The remainder of the paper is organized as follows. Section 2 discusses the data construction method and the meta-data. Section 3 introduces the methodological approach of MRA. Section 4 presents detailed results and discussion. Section 5 provides concluding remarks.

2. Data and methods

2.1 Methods, Protocols and Data Construction

In identify the relevant primary studies, coding, and analysis, the study follows the recent reporting guidelines for meta-analysis in Economics (Havránek et al., 2020). The initial stage of the extensive search started with Google Scholar web engine to identify the relevant primary empirical studies. We searched employing the following broad combination of keywords: "exporting + productivity", "export promotion + productivity", and "export promotion + firm level performance". Doing so, the search identified a huge set of literature. For instance, using the "export promotion + firm level performance" combination of keywords hits 289,000 prospective studies that were inspected based on titles, abstracts, and keywords. Further inspection of the studies was done in case of uncertainty of the initial search. We also conducted a forward search using the Google Scholar web search looking at the list of studies that cited a given study. To supplement the web engine, we also applied the exact combination of keywords as in Google Scholar using the Web of Science database. The final stage of the search strategy utilized the snowballing technique. This backward/hand search approach mainly relied on the reference list of the relevant recent studies identified via the web-based search.

Inspection of titles and abstracts and in case of uncertainty followed by introductions and conclusions resulted in 107 potential primary studies. In general, the identified studies employ the following variant of model estimating the impact of export promotions on firm-level export performance:

$$Y_{it} = \alpha + \beta P_{it} + \mu X_{it} + \varepsilon_{it} \quad (1)$$

where Y is a measure of firm-level export performance for firm i in period t, P represents a binary indicator that takes the value 1 when firm i receives trade assistance or participates in the EPP and 0 otherwise. X represents the vector of observable characteristics to control

[5] According to the 2020 World Association of Investment Promotion Agencies (WAIPA) and World Bank Group (WBG) joint global survey, currently export promotion is the most widely mandated function performed by 51% of Investment Promotion Agencies (IPA): https://waipa.org/blog/state-of-investment-promotion-agencies/ accessed 23 April 2021.

for difference in firm, such as age of a firm since establishment or size of a firm and ε is an error term. The impact of the assistance is captured by the parameter β.

For detailed review based on the full-text evaluation, we selected studies that satisfy the following criteria: English language primary empirical studies that report regression-based firm-level export performance applying either experimental (RCTs) or quasi-experimental designs. These methods are considered as best empirical design to estimate the causal impact of a policy programme or an intervention (e.g., see Duflo et al., 2007; Ravallion, 2007). From the treatment evaluation, quasi-experimental design includes difference-in-differences (DID), instrumental variables (IV), DID with matching, regression discontinuity design (RDD), propensity score matching (PSM), or any other regression design controlling for selection bias (e.g., Heckman two-step approach).

The imposition of the detailed review inclusion criteria using the full-text evaluation resulted in 38 empirical studies published until March 2020 for coding. Of these, 23 empirical studies were peer-reviewed journal articles and the other 15 were unpublished studies, evaluation reports or working papers. Of the 107 potential primary studies, the 59 studies were identified as ineligible as they provide macro perspective evidence using financial resources devoted to export promotion and network of foreign offices to measure export promotion activities. For instance, studies investigated the relationship between public expenditures on export promotion and trade outcomes (e.g., Lederman et al., 2010 for cross country; Bernard and Jensen, 2004, for United States). Other several studies examined general economic diplomacy by relying on international relations via foreign missions such as consulates, embassies, or state visits (e.g., van Bergeijk et al., 2011 and Yakop and van Bergeijk, 2011 for a group of 63 countries; Creusen and Lejour, 2012; for the Netherlands, Hayakawa et al., 2014; for Japan)[6]. In particular, the later approach uses the number or presence of foreign offices operated by export promotion agencies on aggregated export outcomes (Broocks and van Biesebroeck, 2017; Moons and van Bergeijk, 2017). Moreover, the other eight identified studies were excluded partly because they do not provide quantitative evidence or only covered descriptive statistics. Additionally, these studies were excluded partly because they were earlier versions of studies reporting identical results with studies that are already included (e.g., the working paper version of Martincus et al., 2012). Final reason for exclusion of the other two studies was that when the full text of the studies was not accessible or downloadable (e.g., Gadd et al., 2008; Hiller, 2012), though we attempted to contact the authors.

To ensure that our dataset has the highest scientific standard, two reviewers (the author and a research assistant) independently conducted the search and examination of eligibility, followed by extraction of data using a Microsoft Excel data extraction template. Extensive data coding on various dimensions of the empirical studies were conducted to avoid potential subjectivity and enhance the reliability and robustness of the findings (Demena, 2017). The two reviewers collected various aspects of the studies such as, data characteristics, evaluation features, specification, and publication characteristics, among others. Any discrepancies between the two reviewers were double checked and consensus was reached. Finally, the dataset was transferred to a Stata file for analysis.

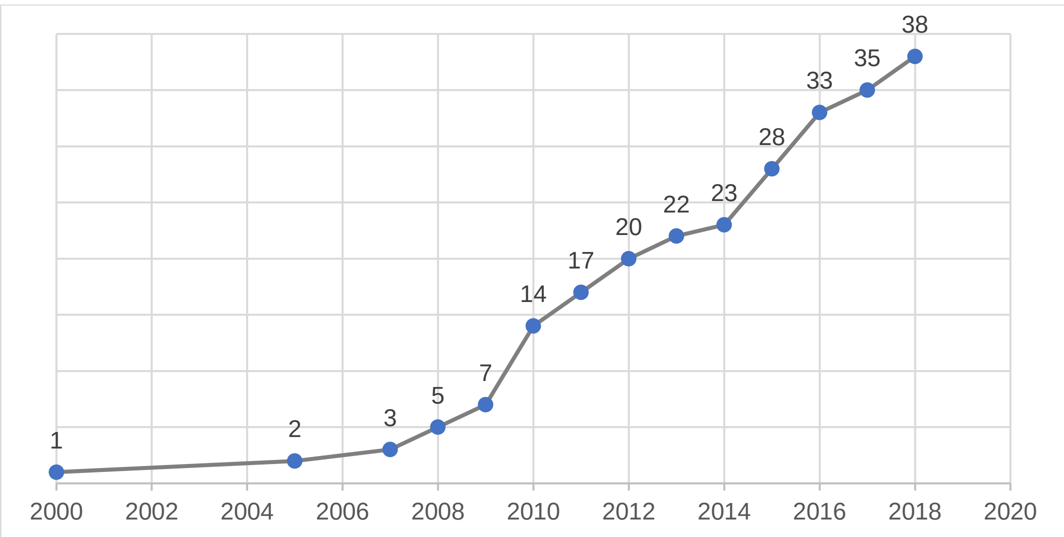
[6] For detailed review and list of such studies, see Moons and van Bergeijk (2017) or van Bergeijk and Moons (2018).

2.2 *Meta-data*

The identified 38 studies are presented in Figure 3. The Figure shows the cumulative number of studies reported in the period between 2000 and 2020 inclusive. In the review process of identifying the relevant studies, no impact evaluation that satisfy the inclusion criteria was found prior 2000. Because about two-third of the current EPPs established in the late 1990s, notably in developing countries (Cruz et al., 2018). The econometric evidence started in 2000 by Crespi and Alvarez (2000) evaluating the impact of public instruments on export performance for Chilean economy. Between 2000 and 2009, only a handful of studies (7) investigate the impact of different export promotion programmes in both developing and developed countries (China, Chile, Germany, Ireland, and Peru). A sharp increase in the number of studies was recorded in 2010, doubling the available studies from 7 to 14 studies. In fact, it was only in 2010 that the publication reached its highest point in a single year; more than three-fourth of this was contributed by developing countries. During the last decade, a wide variety of the studies has covered the firm-level trade performance from EPP. As a result, between the period 2011 and 2020, the number of studies was more than doubled, contributing an additional 24 studies. The rapid growth in the number of studies can be attributed to the increase of the impact evaluation literature on the one hand and of the growing international trade literature on the availability of firm-level data evaluating targeted EPPs on the other hand.

A large share of the observations (about 59%) evaluating government-designed interventions on firm-level export performance have been implemented in developing countries. In terms of region, majority of the observations come from Latin America, and Africa. Meanwhile, Asian countries accounted for only 2 of the 38 studies (i.e., evidence from China and Vietnam). The oldest study was published in 2000, and the most recent is in 2018. The median study was appeared in 2012, and about three-fourth of the estimates reported in the last six years suggesting the literature is recently rapidly evolving and the main question whether such export promotion policies are successful in stimulating exports is still debated.

Figure 3
Cumulative number of studies per year published in the period 2000 up to 2020



We follow the recommendation of Stanley and Doucouliagos (2012) and use all reported estimates of the primary studies (Demena, 2017 for a brief discussion of the three types of meta-datasets). In line with this, several recent meta-analyses adopted the all-set meta-data (i.e., all reported estimates) (e.g., see Moons and van Bergeijk, 2017; van Bergeijk et al., 2019; Floridi et al., 2020; Demena et al., 2021a). Another issue was missing information in the included primary studies. Instead of omitting these studies, we contacted authors of the primary studies when *N*, coefficients, t-values, or standard errors were not reported (e.g., Rincón-Aznar et al., 2015; van Biesebroeck et al., 2015). In total, we have requested 17 emails that enabled us to include 436 observations (about 22% of the extracted estimates). In contrast, there were 18 observations that were excluded as the authors unable to provide missing *N* (10 observations) or standard errors (or t-values – 8 observations). The later resulted in the exclusion of one primary study (Geldres-Weiss and Monreal-Pérez, 2018), and hence ultimately, we included 37 primary studies.

Focusing on the 37 included studies, we used the Hadi (1994) method to account for outliers filtering out effect size and their precisions simultaneously. This method has been recognised as suitable for outliers in multivariate framework and commonly applied in meta-analysis (Iršová and Havránek, 2013; Demena and Bergeijk; 2017; Floridi et al., 2020). Applying this method, 1824 estimates are available for our meta-analysis from 35 empirical studies after excluding 45 observations provided by 11 studies. We report the findings without these outliers, but for sensitivity analysis, we include findings of the full sample. In the appendix, Table A1 gives detailed information of the selected studies.

3. Empirical approach

3.1 Partial correlation coefficient

A methodological approach of a meta-analysis consists of combining all the available 1824 reported estimates and hence considered as a regression on estimates from existing and accessible regressions – analysis of the analyses. Combining estimates of the primary studies that differ in terms of the proxies for export promotion instruments and firm-level export performance requires standardization to make the reported estimates comparable across the primary studies. Following Doucouliagos (2005), Havránek et al. (2016), and Floridi et al. (2020), we employ the Partial Correlation Coefficient (PCC) as:

$$PCC_{sc} = \frac{t_{sc}}{\sqrt{t^2{}_{sc} + df_{sc}}} \qquad (2)$$

where $s=1, \ldots, 35$ represents the primary study; $c=1, \ldots, 1824$ represents the reported estimates specified in each of the regression of the primary studies; PCC_{sc} denotes the partial correlation coefficient between indicator for EPPs and the outcome on the firm-level performance. This measures the association between these two variables in terms of direction and strength holding other variables constant. t_{sc} represents the corresponding t-values and df_{sc} is the associated degree of freedom for the reported estimates in each of the regression specification of the primary studies.

3.2 Publication selection bias and genuine effect

Next, we investigate whether the PCC computed from the reported estimates subject to publication selection bias. To do so, we first use graphical inspection of the reported estimates using funnel plots. A funnel plot is a scatter diagram of the estimated effects on the horizontal axis and their precision on the vertical axis, usually the reciprocal of the standard error (Stanley and Doucouliagos 2012; Demena, 2015). However, this approach of examining publication bias, which is based on visual inspection, is subjective, and thus not convincing. Therefore, we further complement the visual inspection employing a formal statistical approach. Doing so, we use the following meta-regression model (MRM) to investigate the effect of the associated standard error of the PCC (SE_{pccse}) on the computed coefficient in Eq. (2) itself:

$$PCC_{sc} = \beta_1 + \beta_0 SE_{pccse} + u_{sc} \qquad (3)$$

where PCC_{sc} is the measure of estimated effect computed in Eq. (2) for the c^{th} regression specification estimate associated with the s^{th} study, and SE_{pccse} its standard error (accuracy of the computed coefficient, which is given by $\sqrt{(1 - pcc^2_{sc}) / df_{sc}}$). β_1 stands for the overall genuine effect of EPP on firm-level export outcome and β_0 is the publication selection bias, which is the tendency from researchers to select research findings or results that are statistically significant results using alternative estimation specifications or techniques to satisfy prior theoretical expectations, especially in the face of small samples. This MRM (Eq. (3)) shows that as sample size increases and thus the quantity of available information increases, SE_{pccse} will approach towards zero (Stanley 2005). In other words, with large sample size, PCC_{sc} will tend closer to β_1, the overall underlying size of the effect beyond publication bias. In other terms, in the absence of publication bias, the overall genuine effect will vary randomly around β_0, regardless of SE_{pccse} (Doucouliagos and Stanley, 2009). Conversely, the presence of bias can be detected if PCC_{sc} are correlated with their associated standard errors, which is SE_{pccse} (Stanley and Doucouliagos, 2010).

In a classical regression, the error term (u_{sc}) must be independent and identically distributed to obtain an efficient estimator (Demena and Afesorgbor, 2020). However, in estimating Eq. (3), the independent variable is the standard error of the dependent variable, and thus PCC_{sc} has different estimated variance (Stanley and Doucouliagos, 2012), the obvious problem of heteroscedasticity. As in our case, Stanley (2005) argued that there will be the likelihood of intra-study dependence in error terms. To reduce such dependence in error terms as well as to assign greater weight to estimates with greater precision, Eq. (3) should be estimated with a weighted least squares (WLS) approach. More specifically, Eq. (3) can be weighted by inverse variance of the estimated PCC_{sc}, which is dividing Eq. (3) by SE_{pccse} and becomes Eq. (4):

$$t_{sc} = \beta_0 + \beta_1 (1/SE_{pccsc}) + e_{sc} \qquad (4)$$

where t_{sc} (PCC_{is}/SE_{pccis}) is the *t*-value that measures the statistical significance of PCC collected from the reported primary studies. We then assessed the existence of publication bias testing β_0 = 0, which is testing the null hypothesis for the intercept in Eq.

(4) equal to 0. This is known as the funnel asymmetry test (FAT). In other terms, when β_0 is statistically significant, it indicates that the reported estimates do not vary randomly around the underlying genuine effect (β_1) and thus asymmetrical (Stanley and Doucouliagos, 2012). In this case, it might be the case that part of the estimated evidence are missing from the reported estimates in the literature, producing truncated funnel plot and hence the reference to FAT (Balima et al., 2020).

After filtering out any publication bias presence, the underlying genuine effect can be tested by the slope of Eq. (4) (β_1), which is the parameter associated with the inverse of the standard error ($(1/SE_{pccsc}$). This estimates both the direction and size of a genuine effect and carry out the precision-effect test (PET). Rejecting the null hypothesis would signal the presence of underlying genuine effect of EPP on firm export performance after accounting for potential bias. In this case, when the intercept of Eq. (3), which is β_1 has statistically significant regardless of the associated publication bias suggests the presence of overall underlying effect in the literature.

3.3 Potential factors explaining heterogeneity

To explain the various potential sources of heterogeneity among the collected estimates, we employ a multivariate meta-regression analysis (MMRA). Doing so, we augment the FAT-PET in Eq. (4) by including the moderator variables presented in the appendix, Table A2 that likely drive the heterogeneity in the reported estimates of the primary studies. More specifically, Eq. (4) becomes:

$$t_{sc} = \beta_1 + \beta_0(1/SE_{pccsc}) + \alpha_k X_{ksc}/SE_{pccsc} + e_{sc} \qquad (5)$$

where *X* denotes a vector of the moderator variables listed in the Table A2 weighted by the inverse of the variance, α_k is the associated coefficient and *K* is the specific category the moderator variable.

To explain variables used in the MMRA, Table A2 gives an overview of the primary studies characteristics that include their definitions, means and standard deviations. Following the heterogeneity reported by the primary empirical studies, the table classifies them into several categories, such as data characteristics, evaluation features, estimation methods, publication characteristics, the choice of dependent and independent variables.

Data characteristics: We incorporate the number of observations in primary studies to test for systematic variation between large and small samples. The mean logarithm of the number of observations in the primary studies is 8.984. Next, we consider the length (time span) of the dataset used in evaluating the EPP impact. Finally, we include a dummy variable taking 1 if the regression is based on a sample from developed country, 0 otherwise to assess the role of sample source. 21 out of the 38 studies tested the effectiveness of EPPs in developing countries.

Evaluation features: Various empirical approaches have been conducted for the evaluation design of trade-related policy interventions, in particular econometric evaluation of policies and programmes for export promotion. Following the debates in the impact evaluation literature and the approach presented by the reported primary studies, we report four sub-categories of potential sources of heterogeneity regarding evaluation features of the experimental and quasi experimental methods: evaluation timing follow-up, intervention policy, export promotion target, and treatment effect estimates.

Evaluation timing follow-up. Recent evidence on informal firm formalization policy shows the relevance of time for the impacts to materialize (Floridi et al., 2020). Taking this into account, we incorporate the role of timing of evaluation follow-up driven heterogeneity. We account using binary variables indicating whether the performance regression was derived through evaluation timing based on less than or equal to one year or longer than one year. In our meta-data, approximately about two-third (65%) of collected estimates stem from studies that design evaluation of the impacts building on longer than one year timing follow-up.

Export promotion target: We also incorporate whether the trade-supporting activities impact potentially vary with firms different sizes. According to Martincus et al. (2012), for instance effects of trade supporting actions are likely to be heterogenous for which smaller and medium size firms are expected to receive stronger effects as they have greater limitations in accessing relevant export information to become successful players in these markets. In contrast, Alvarez (2004) for Chilean small and medium firms did not found positive impact on firm performance. We use dummy variables to label these sources of heterogeneity if the study targets small and medium enterprises (SMEs), large firms or any firm. In general, EPPs spend more on SMEs as compared to large ones (Cruz et al., 2018).

Intervention policy: Export promotion agencies have been utilized various specific government intervention to ease the supply-side constraints to exporting (Kim et al., 2018). Most promotion programmes help domestic firms to lower variable or fixed costs of trading, such as providing information on foreign markets. For instance, to adapt products to foreign tastes, matching firms to clients, assisting firms to find a distributor, navigate foreign customs and product regulations, or export procedures among others (Cadot, 2015; Broocks and van Biesebroeck, 2017). In our meta-data, about 54% of the performance regressions use such kind of trade promotion services to local firms. Moreover, other most direct measures to help local firms succeed in export markets are export subsidies and grants. About 38% the regressions build on such large investment grants or production subsidy. Other types of government interventions have been bundled or combined both trade promotion services and export subsidies and grants or lower tax rates for export earnings.

Treatment effect estimates: Next, we included information concerning the estimation of treatment effects or techniques that may constitute another driver of heterogeneity. It is commonly agreed that the key difference in the treatment evaluation literature is the approach or estimation techniques to validate treatment effect estimates accounting for selection bias or endogeneity issues, the so-called fundamental problem of causal inference (Holland, 1986; Cadot et al., 2015). Since 98% of the collected estimates stem from a quasi-experimental design, it is important to employ appropriate estimation techniques while controlling for possible selection bias. We account the role of estimation techniques-driven heterogeneity using binary variables if the performance regression stems from an instrumental variable (IV) estimation, a multivariate linear regression (plain vanilla ordinary least square - OLS), multivariate non-linear regression model (probit, logit, or tobit), a simple DiD estimation, combining matching with DiD approach, propensity score matching or weighted regression, and a generalized methods of moments (GMM) estimation. RDD, an alternative quasi-random evaluation technique or statistical tool from the evaluation literature has not been used among the collected estimates. This was found to be infeasible in this particular literature as all firms in most countries qualify for government-designed export promotion instruments (Gourdon et al., 2011; van Biesebroeck et al., 2016). Since enrolment into export promotion programmes is not random, most papers control for selection into treatment through matching, fixed effects, or two-step (IV or Heckman) estimation methods. In our meta-dataset, majority of the studies (around 82 %) control for selection bias.

Moreover, 86% of the treatment effect estimates are average treatment effect on the treated (ATT), while few other estimates are the average treatment effect (ATE), the local average treatment effect (LATE) or other effects. We include a dummy variable indicating 1 if the effect of the treatment is ATT and 0 otherwise. The rational for this is that effect of the treatment can have different implications such as in terms of external validity, effect size magnitude, policy interest or making (van Biesebroeck et al., 2015; Kluve et al., 2019).

Firm performance outcome: The heterogeneity of the identified studies also reflected in the choice of outcome variables employed to proxy firm-level evidence from EPP. The impact on export outcomes indicated through firm-level evidence investigating the change on a firm's level of export (the intensive margin), and whether intervention draws a new firm into the export market (the firm-extensive margin). The former is easier than the latter since the firm extensive margin requires data on the universe of potential exporters (van Biesebroeck et al., 2016). In line with observing outcomes for existing exporters only, researchers also investigate whether promotion policies help firms to diversify their exports and reach different product or destination markets, i.e., if programme affected the destination and product extensive margins. We therefore incorporate the role of these four-dimensional view of export performance indicators in explaining the heterogeneity in the reported results using dummy indicators.

Specification Characteristics: The primary studies control for various variables in evaluating the firm performance regression. We included dummies for the inclusion of control variables like the age of the firms since establishment and the size of the firms. We also include whether the regression formed with firm fixed effects as well as year fixed effects. Approximately two-third of the performance regression include firm fixed effects to account for past firm export performance so as selection into export promotion programmes.

Publication characteristics: Finally, we consider the publication characteristics of the reported studies to proxy for their qualitative difference in four dimensions. First, we control whether a study is published in a peer-reviewed journal using a dummy variable equalling 1 and 0 otherwise (representing unpublished studies, book chapter, evaluation reports or working papers). Slightly less than two-third of the collected observations are published in peer-reviewed journals. The notable outlets of the publication are *Journal of International Economics* and *The World Economy* (three in each Journal). Next, we incorporate publication year of the primary studies to assess whether there is a systematic time trend in the reported findings by the primary studies. Finally, we use author citations in Google Scholar to control for the quality of the included primary studies and recursive impact factor from RePEc for journal quality or ranking where the studies are published. We choose Google Scholar in providing citation counts as it is the richest source also covers books and grey literature, hence gives a better indication of impact (Demena et al., 2021a). The RePEc provides impact factor for both peer-reviewed journal articles and working paper series in Economics. As of October 2020, this literature in total have received 1,778 citations. The most cited article in our meta-data has appeared in the *Journal of International Economics* and received 283 citations (Martincus and Carballo, 2008). The second and third most cited articles are published in the *Review of Economics and Statistics* (207 citations, Görg et al., 2008) and *Journal of Development Economics* (168 citations, Martincus and Carballo, 2010a).

3.4 Econometric concerns

In its general form, estimating Eq. (5) has problem of potential multicollinearity due to the large number of moderator variables (Table A2). In addition to potential multicollinearity concern, Demena and Bergeijk (2017) concur that including all moderator variables (which are notably binary variables) would result in reducing the degree of freedom. To address the multivariate MRA model uncertainty, Stanley and Doucouliagos (2012) recommend the general-to-specific (G-to-S) technique (also as recently outlined by the MAER-Net reporting guidelines, (Havránek et al., 2020). The procedure starts with all 26 potential moderator variables using the general specification (Eq. (5)). Next, we proceed to systematically removing insignificant variables, one at a time, until only significant variables remain to arrive at a specific/reduced model. Doing so, we exclude 11 moderator variables that are statistically insignificant at least at the 10% level.[7] In practice, robust methods in meta-analysis using the G-to-S approach are widely employed (Stanley and Doucouliagos, 2012; Abdullah et al., 2015; Demena and Bergeijk, 2017; Demena and Afesorgbor, 2020; Floridi et al., 2020). Therefore, our selection for G-to-S is consistent with existing guideline that is also being widely applied in recent MRA.

Another empirical concern is data dependence which arises when multiple reported estimates (all estimates) are gathered from the same study that would yield biased estimates of the Eq. (4) and Eq. (5). We adopted three alternative remedies. First, clustered data analysis (CDA), which is clustered ordinary least squares using study-level clustered standard errors, hence the reference to clustered data analysis. Second, we also employ a fixed effect (FE) estimation that may also address the issue of individual within-variation. These two estimators, however, can be used to control for within-study dependence correcting the standard errors. Beyond the within-study dependence, there is additional econometric concern of between-study dependence. Methodologically, this is important in our case because of multiple studies that are published by the same authors (and thus statistically likely to be dependent). We check for the existence of statistical dependency between studies using the Breusch-Pagan Lagrange multiplier (BP-LM) test, suggesting the presence of between-study dependence.[8]

To capture the between-study dependence while controlling for within-study dependence, a multi-level model or hierarchical model is considered as most appropriate. This model allows to account both the within-study and between-study dependences through the inclusion of a random individual effect for each study and thus the reference to a multilevel random effect model (Balima et al., 2020; Ugur et al., 2020). The multilevel

[7] These moderators are not only individually but also jointly equal to 0 (joint F-test F (11,1796) = 0.95 (*p-value*=0.488). In contrast, the joint test for the remaining 15 moderator variables included in the reduced model rejects the null hypothesis of a 0 joint effect, F (15, 1796) =63.69 (*p-value*=0.000).

[8] The BP-LM, a chi-squared with one degree of freedom revealed the study-level effect to be 867.41 with p<0.001. The method yields similar suggestion when outliers are included: $\chi^2_{(1)} = 979.82$, $p<0.001$.

random effect model becomes a model which is best described as mixed-effects multilevel model (MEM) when it includes the identified moderator variables controlling for both within-study and across-study variations (Demena, 2017; Balima, 2020). More specifically, the model can be described as a two-level model (both at the study level and the estimate level), hence we extend Eq. (5) as follows:

$$t_{sc} = \beta_0 + \beta_1(1/SE_{pccsc}) + \frac{\alpha_k X_{ksc}}{SE_{pccsc}} + \zeta_s + e_{sc} \qquad (6)$$

where ζ_s is the study-level random effects (random intercepts), and the others are as in Eq. (5). In this modelling, estimates, which are level 1, are clustered and nested within studies, which are level 2. Our notations follow that of Doucouliagos and Stanley (2009), Havránek and Iršová (2011), Demena and Bergeijk (2017), and Ugur et al. (2020). Following recent meta-analyses, therefore, estimates associated with EPPs that are reported by the empirical primary studies are nested within each study and the estimates are modelled to differ between studies (Floridi et al., 2020).

Robust methods in meta-analysis using the MEM model are widely employed, (Doucouliagos and Stanley, 2009; Havránek and Iršová, 2011; Demena and Bergeijk, 2017; Floridi et al., 2020; Demena and Afesorgbor, 2020). Additional feature of the multi-level model is that it allows for likelihood ratio (LR) tests to choose between multi-level estimators and other alternatives. The multi-level model allows for testing different assumptions about the variance structure through a LR test and thus we choose the MEM specification only if the LR tests indicate that a multi-level model is preferable to other specifications. Hence, we check this if the multi-level model is superior as opposed to other alternatives, and the tests in all case justify the MEM model we adopt. Apart from the advantages of the multi-level discussed above, simulation results have shown that multi-level have better tolerance to high multicollinearity is another feature that makes them preferable to other alternative estimators. (Ugur et al., 2020). To sum up, we basically refer to our baseline results using the CDA and FE estimators and hence our interpretation of the results give most weight to the MEM.

4. Findings and discussion

4.1 Weighted average effect

To derive the combined impact, we start with the computation of the simple and weighted average effect of EPP. These combined effect sizes are presented in Table 1. The simple (unweighted) shows average effect of 0.034, statistically significant at a 95% confidence interval (0.031 – 0.036). However, Copper and Hedges (1994) recommend using weighting scheme. Accordingly, to assign greater weight to more precise estimates, we provide inverse variance weighted average effect size. This procedure also gives positive and significant effect, but the overall magnitude reduced by 50% than the simple average effect. Regardless, EPP appears to have almost no practical significant positive impact on firm performance.[9]

[9] Our interpretation of the PCC effect size follows the meta-analysis guidelines by Doucouliagos (2011): a coefficient is small when it reaches to 0.07, medium if it ranges around 0.17, and large if it is at least 0.33.

Table 1
Overall reported estimates of EPP on export performance

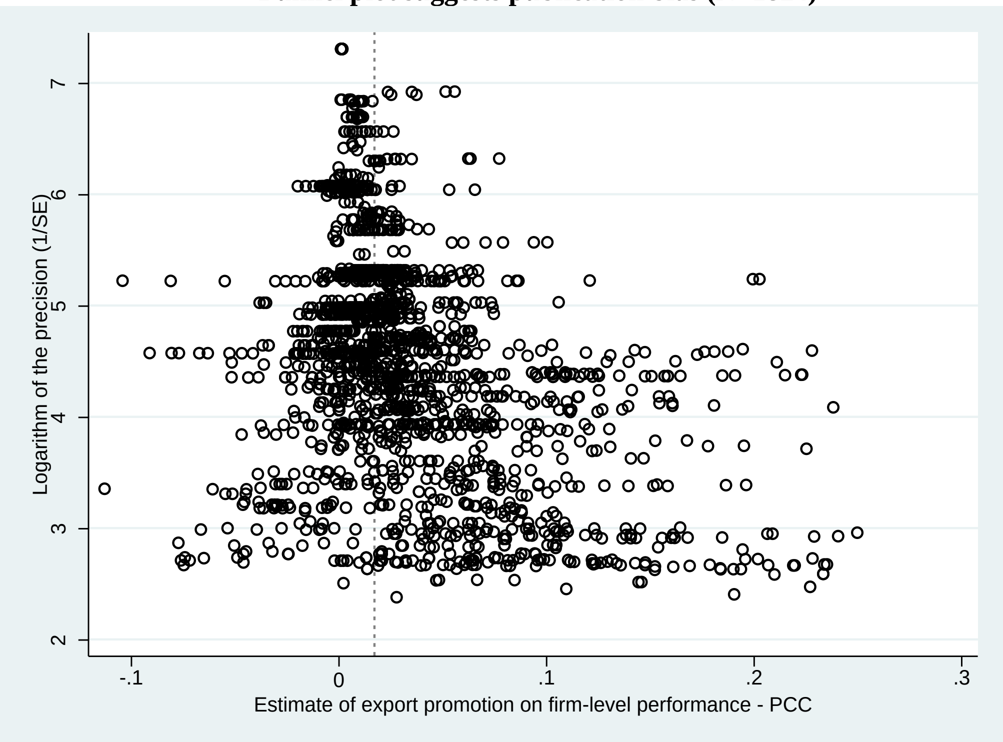
Method	Effect size	S. E.	95% confidence interval	
Simple average effect	0.034	0.001	0.031	0.036
Weighted average effect	0.017	0.001	0.016	0.018

Although this inference of the overall effect size is insightful, it is quite basic and would not be valid. As Stanley and Doucouliagos (2012) argue, we need to account for the issue of publication bias. The next step is therefore to investigate whether the overall little or non-existent practical effect is influenced by publication bias.

4.2 Funnel plots: Testing for publication bias

Figure 4 gives the common funnel plot tool to visually assess the degree of publication bias. In the absence of publication bias, a funnel plot seems to appear a symmetrical inverted funnel shaped. In this case, less precise estimates widely dispersed at the bottom of the graph, whereas most precise estimates would be closely distributed around the underlying effect at the top. In contrast, an asymmetrical funnel plot indicates the presence of publication bias, implying some estimates are more represented or discarded in favour of a particular expected sign or conclusion (type I bias). Furthermore, according to Doucouliagos et al. (2005) and Stanley (2008), when both statistically significant negative and positive findings have equal reporting preference or chosen irrespective of the direction of the effects (type II bias), the funnel plots are hallowed and excessively wide but likely to be symmetric.

Figure 4
Funnel plot suggests publication bias (N=1824)



Note: To allow better visualization of the plots between reported effect size and their precision, we present the logarithm of the inverse of the standard error. The short dashed vertical line represents the inverse variance weighted average effect of the full sample as reported in Table 1.

The funnel plot in Figure 4 displays the reported effect size (PCC) on the horizontal axis and their precisions represented by the inverse of the standard errors on the vertical axis. The funnel plot suggests a positive publication bias. The reported effect sizes create an asymmetrical funnel, i.e., the most precise estimates are close to 0.017, but there are many imprecise estimates reported larger than 0.017 as opposed to those that are smaller than 0.017. The plots are therefore somewhat skewed toward the righthand side of the diagram, implying relatively too many larger positive effect sizes are reported in the literature. This could be an indication of type I publication bias: estimates that are in favour of the most commonly cited views in the literature are more preferred and frequently reported and published. In the field of international economics, various recent meta-analyses have already reported strong publication bias towards both positive and negative estimates. For instance, Balima et al. (2020) positive bias in the literature of inflation targeting on output growth volatility; Bajzik et al., (2020) bias against small and insignificant results in estimating the Armington elasticity; van Bergeijk et al. (2019) positive and negative bias on determinants of economic sanctions; Demena (2015) and Demena and Bergeijk (2017) positive bias in the literature of foreign direct investment and productivity spillovers.

Table 2
Publication bias and underlying genuine effect test

All studies – FAT/PET				
Variables	**(1)**	**(2)**	**(3)**	**(4)**
	CDA	**FE**	**MEM**	**Wild bootstrapped**
Publication bias (FAT - constant)	1.390***	1.343***	1.330***	1.390***
	(0.358)	(0.158)	(0.403)	0.00
Genuine effect (PET - precision)	0.009***	0.010***	0.010***	0.009***
	(0.002)	(0.001)	(0.000)	0.00
Observations	*1824*	*1824*	*1824*	*1824*
Studies	*35*	*35*	*35*	*35*
LR test			*190.64*	
$p > chi^2$			*0.000*	

Notes: *** stands for 1% level of significance. Reported results are estimated from Eq. (4) specification. All estimates use the inverse variance as weights and standard errors reported in parentheses are clustered at study level. Panel 1 (CDA: clustered data analysis is estimated via the study level clustered robust standard errors; Panel 2 (FE) is fixed-effect estimation clustered at the study level; Panel 3 (MEM) is mixed-effects multilevel estimated through the restricted maximum likelihood; and Panel 4 (Wild bootstrapped) is regression bootstrapping the standard error (a non-standard cluster adjustment) reported with *p*-values. Test for between-study heterogeneity (Cochran's Q-test) is 57801.16*** on 1823 degrees of freedom with p-value less than 0.001 and I^2 statistics (variation in reported estimates attributable to heterogeneity) is 96.8%.

4.3 Publication selection bias and genuine effect

4.3.1 Publication selection bias

To test more formally for the existence and size of publication bias, we need to move beyond eye-o-metrics of the simple visual funnel plot test (Figure 4). Indeed, Doucouliagos et al. (2005) suggest that the appearance of a funnel plot can be deceiving, hence a more formal and objective statistical investigation is required as the presence of publication bias can be statically identified even if the plots more or less tend to be symmetrical. That means publication bias cannot be simply investigated with graphical inspection, and thus regression-based FAT gives a more precise formal way to assess publication bias.

Table 2 depicts results of publication bias test (FAT) using the various estimation techniques. In all the estimations, the publication bias detected through the visual test is confirmed, points to asymmetry test values. The results are consistent both in size and statistical significance across all specifications. The finding suggests substantial publication bias and highly statistically significant, and in our preferred model is 1.330 (MEM).[10] Putting into perspective, based on the 87 quantitative survey of economics research by Doucouliagos and Stanley (2011), the value of publication bias is 1.58. In contrast, the impact of unionization on worker productivity literature, for instance, has 0.65 coefficient of publication bias, little to modest selectivity bias (Doucouliagos and Laroche 2009). In this study, we note that the signal of the bias remains unchanged and robustly substantial to different econometric methods, confirming the type I bias of the visualized findings of the graphical inspection.

Table 3
Publication bias and genuine effect implied by export performance indicator

Variables	FAT/PET			
	(1)	**(2)**	**(3)**	**(4)**
	Intensive margin	Firm-extensive	Destination-extensive margin	Product-extensive margin
Publication bias	-0.099	3.791***	1.935***	0.811
(*FAT - constant*)	(0.489)	(0.996)	(0.435)	(0.553)
Genuine effect	0.018***	0.003***	0.010***	0.009***
(*PET - precision*)	(0.001)	(0.001)	(0.001)	(0.003)
Observations	*720*	*289*	*474*	*338*
Studies	*27*	*13*	*19*	*15*
LR test	*122.20*	*66.83*	*1.80*	*28.84*
$p >$ chi^2	*0.000*	*0.000*	*0.089*	*0.000*

Notes: *** stands for 1% level of significance and standard errors are reported in parentheses. Reported results are estimated from Eq. (4) specification. All estimates are obtained from mixed-effects multilevel model through the restricted maximum likelihood. The inverse variance is used as weights.

To find a more nuanced assessment of publication selection bias, Table 3 presents results for more homogeneous groups implied by firm-level export performance indicators using the mixed-effects multilevel model only, our preferred model.[11] Doing so, we limit our meta-dataset to the four-dimensional view of export performance indicators separately: the intensive margin (export intensity), the firm-extensive margin (incidence of export market entry), the destination-extensive margin (new export destination entry), and the product-extensive margin (new product export). Consistent with the whole sample, Columns (2) and (3) report positive and statistically significant findings. This suggests that the firm- and destination-extensive margins are pointing to the existence of 'severe' and 'substantial' selectivity towards positive findings, respectively. In this case, it appears that

[10] According to Doucouliagos and Stanley (2013), the size of publication selection bias found in this study is substantial: 'little to modest' if FAT is at most ±1, if it ranges between ±1 and ±2 signals 'substantial'.
[11] Several studies have shown that when the number of studies/clusters is small, standard adjustment for clustering may potentially produce biased result, and adopted the use of non-standard cluster adjustments, such as the wild bootstrap. We also follow and run this wild-bootstrap approach - a non-standard cluster adjustment recommended by Cameron et al. (2008) and results are consistent with the main findings reported in Table 3.

researchers tend to prefer export promotion enhancing impact on incidence and new destination exporting. We do not find any publication selectivity in the intensive margin and product-extensive margin groups. Therefore, our main result for the full sample in Table 2 can be driven by results of Columns (2) and (3) of Table 3.

4.3.2 Genuine effect

Going beyond the issue of publication selection bias, Table 2 also allows to test for the existence of genuine underlying effect of EPP on firm performance. The PET (i.e., the slope of estimated coefficient) gives the overall underlying average effect corrected for publication bias. This result points to a positive and highly statistically significant effect, implying that EPPs are associated with enhancing internationalization of locally oriented firms, notably after filtering out publication bias. However, the magnitude of the genuine effect is small (0.010) and reduced by about 41 per cent compared to the weighted average effect reported in Table 1. Overall, consistent with the (un)weighted mean effect reported in the literature, the underlying genuine effect EPP still tends to have little or non-existent practical impact on export performance activities.

Similarly, in Table 3 we present the genuine effects for the sub-samples implied by the export performance indicators after filtering out publication bias. Again, we continue to find comparable results of the full sample. The findings in Columns (1) – (4) suggest positive and statistically significant effects. The results are not only stable and analogous in terms of sign of the coefficients and statistical significance, but also in size of the estimates, implying our genuine effect results for the whole group are not driven by any of the sub-group analyses. Our main findings of almost no practical genuine impact of export promotion in the full sample is again consistently confirmed across the various export performance indicators.

Nonetheless, it is worth to mention that these findings are average across all empirical designs. As shown in the box plot (Figure 2), the reported estimates in the primary studies substantially vary both within and between studies. As reported under Table 2 (Cochran's Q-test), the heterogeneity across all the estimates reported in the primary studies is highly significant. The I^2 statistics of heterogeneity reports that the variation in the estimates reported due to sampling error is only 3.2%. We need a multivariate MRA as results may largely depend on other potential sources of heterogeneity across findings in the empirical studies as outlined in Section 3.3 and presented in Table A2. The next section further explores the heterogeneity behind the reported divergent empirical estimates.

4.4 Explaining heterogeneity via multivariate MRA

Table 4 reports the results of the specific MRA using the G-to-S technique as specified in Section 3.5 to address the multivariate model uncertainty. The specified model in column 1 is then estimated using our baseline CDA and FE using Eq. (5) to account for within-study dependence. Next, our main results are re-estimated via the preferred MEM multivariate model (Eq. 6) to address the drivers of heterogeneity.[12]

As discussed earlier, the test for publication selection bias with simple graphical inspection can be misleading, and in our case the skewed dispersion of the reported

[12] Our two alternative estimations (CDA and FE) are used as a baseline model for comparison and robustness checks. Results of CDA and FE are also in line with the preferred MRA model (MEM).

estimates (Figure 4) and the formal test (Tables 2 and 3), could suggest heterogeneity of the studies that systematically drive the variation in the reported coefficients, such as variation in data, estimation, or specification characteristics. After the inclusion of potential moderator variables, the presence of publication bias disappears, implying the absence of selection bias towards positive estimates reported in so far. The results agree across the three specifications, indicating the absence of robust publication selection bias to different econometric specifications. The central findings of the presence of publication bias using the visual inspection of the funnel plot and FAT-PET in the previous sections appear to be due to the difference in econometric design of the primary studies instead of systematic selection bias towards positive estimates.

We also discussed above that the overall underlying genuine effect of export promotion on export performance is somewhat weak or practically almost negligible, but statistically significant. Again, this main finding is confirmed after accounting for drivers of heterogeneity. However, the multivariate MRA in Table 4 shows that some aspects of the econometric approaches have significant influence on the reported estimates, implying that the underlying genuine effect depends on moderator variables that potentially explain the heterogeneity across the estimates. In what follows, we discuss the importance of this individual aspects of moderator variables in detail.

In reference to the data characteristics, we find that the number of time span covered by the primary studies tend to significantly affect the reported estimates and thus positively associated with the impact of export promotion on export performance. The finding suggests the relevance of longer data time span as the potential benefits of EPP may take time to materialize. The policy implication is therefore to expand the time dimension when analysing the export promotion and export performance nexus. This is consistent with other field of research, such as for formalization interventions (promoting firms to opt to operate formally) appear to take time to materialize (Floridi et al., 2020). Our results also indicate the larger number of firms included in the analysis is less likely to report positive impact on firms' export performance. This could be taken as some evidence that studies based on limited sample are more likely to be with the caveat that their specifications suffer from small sample size bias.

The trade-supporting activities impact vary depending on the types of firms targeted. On the one hand, SMEs are likely to be less responsive to export promotion policies as compared to any firms. Firms with different degree of internationalization strategy could face different obstacles, in particular SMEs have greater limitations to be successful in international markets (e.g., see Bernard and Jensen, 2004). Therefore, it could be argued that, if a significant negative impact is encountered only with SMEs, one might think that such firms have too little capacity for shouldering the costs of international involvement. Note, however, on the other hand that this is less likely to be an issue in our case because, even the impact for larger firms (as compared to any firm) remains equally negative and significant. The latter makes it more plausible that responsiveness depending on the types of firms targeted is not the entire explanation for the effectiveness of export promotion policies.[13] If exporters being larger, they will be expected to benefit more from trade-supporting polices, and thus more likely to self-select into treatment and influence

[13] Especially in far-away markets it would be too expensive for all kinds of firms, but the largest firms would be in a better position to bear costs of becoming successful players in international markets.

outcome. However, the fact that we continue to find similar result even for larger firms, it makes it more unlikely that self-selection bias is fundamental problem in here.

Table 4
Why do estimates vary? explaining the drivers of heterogeneity

Moderator variables	**(1) Specific**	**(2) CDA**	**(3) FE**	**(4) MEM**	**(5) Wild bootstrapped**	**(6) MEM**	**(7) MEM**
Bias coefficient	0.562**	0.562	-0.877	-0.094	0.562	0.144	-0.034
(constant-FAT)	(0.216)	(0.349)	(0.635)	(0.383)	(0.164)	(0.383)	(0.379)
Genuine effect	0.038***	0.038***	0.058***	0.041***	0.038***	0.045***	0.049***
(precision-PET)	(0.005)	(0.008)	(0.020)	(0.007)	(0.000)	(0.007)	(0.007)
Data characteristics							
No. of firms or obs.	-0.004***	-0.004***	-0.006***	-0.004***	-0.004**	-0.005***	-0.004***
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.020)	(0.001)	(0.001)
No. of Years	0.004**	0.004	0.014**	0.009***	0.004	0.008***	0.003
	(0.001)	(0.004)	(0.005)	(0.002)	(0.478)	(0.002)	(0.002)
Export promotion target							
SME	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.007***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Large	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***
Performance indicator							
Firm-extensive	0.011***	0.011*	0.017	0.015***	0.011	0.015***	0.015***
	(0.002)	(0.006)	(0.011)	(0.002)	(0.354)	(0.002)	(0.002)
Destination-extensive	0.004**	0.004	0.006*	0.004***	0.004	0.004***	0.003**
	(0.001)	(0.004)	(0.003)	(0.002)	(0.516)	(0.002)	(0.002)
Treatment effect estimates							
ATT	0.004**	0.004	-0.011	-0.003	0.004	-0.004	0.002
	(0.002)	(0.004)	(0.008)	(0.003)	(0.037)	(0.002)	(0.003)
OLS	-0.004***	-0.004*	-0.000	-0.002	-0.004*	-0.000	-0.002
	(0.001)	(0.002)	(0.003)	(0.001)	(0.052)	(0.002)	(0.001)
PSW	-0.012***	-0.012	-0.003	-0.003	0.012	-0.001	-0.002
	(0.003)	(0.011)	(0.014)	(0.004)	(0.522)	(0.004)	(0.004)
Non-linear	-0.009***	-0.009**	-0.003	-0.006**	-0.009*	-0.006**	-0.009***
	(0.002)	(0.004)	(0.007)	(0.002)	(0.082)	(0.002)	(0.002)
Specification characteristics							
Age	-0.002**	-0.002	-0.000	-0.000	-0.002	-0.001	-0.001
	(0.000)	(0.002)	(0.001)	(0.001)	(0.334)	(0.001)	(0.001)
Year FE	-0.007***	-0.007***	-0.007***	-0.007***	-0.007***	-0.007***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Firm FE	-0.006***	-0.006***	-0.006***	-0.007***	-0.007***	-0.007***	-0.006***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Interaction terms						-0.009***	
						(0.002)	
Developed							0.010***
							(0.002)
Publication characteristics							
Publication year	0.001***	0.001**	0.000	0.001***	0.001**	0.001***	-0.000
	(0.0002)	(0.000)	(0.001)	(0.000)	(0.022)	(0.000)	(0.000)
Published	0.011***	0.011***	0.021***	0.017***	0.011***	0.018***	0.017***
	(0.001)	(0.003)	(0.008)	(0.002)	(0.006)	(0.002)	(0.002)
Observations	1824	1824	1824	1824	1824	1824	
Studies	35	35	35	35	35	35	
LR test				*2549.53*		*77.42*	*95.46*
$p > chi^2$				*0.000*		*0.000*	*0.000*

Notes: *** stands for 1% level of significance. Reported results are estimated from Eq. (5) specification. All estimates use the inverse variance as weights and standard errors reported in parentheses are clustered at study level. Panel 1 the specific model derived from the general specification (G-to-S approach); Panel 2 (CDA: clustered data analysis, also known know as WLS) is estimated via the study level clustered robust standard errors; Panel 3 (FE) is fixed-effect estimation clustered at the study level; Panel 4 (MEM) is mixed-effects multilevel estimated through the restricted maximum likelihood; and Panel 5 (Wild bootstrapped) is regression bootstrapping the standard error (a non-standard cluster adjustment) reported with *p*-values.

Next, we find the measurement of export performance indicator in terms of the firm-extensive (0.015, *p-value*=0.000) and destination-extensive margins (0.004, *p-value*=0.000) is statistically significant and positive as compared to the intensive margin of trade. One reasons for this finding could be that the impact of export promotion programmes has stronger effects on export activities that face relatively greater hurdle or more sever informational obstacles (van Biesebroeck et al., 2016; Broocks and van Biesebroeck., 2017). The findings accordingly suggest that for established exporters, impacts are likely to be larger along the extensive margin, when firms opting to break into an entirely new country market than along the intensive margin. This pattern is consistent with the literature from Latin American countries showing that trade promotion is more successful in assisting firms diversify their exports penetrating new markets primarily overcoming the most information incompleteness (Martincus and Carballo, 2008; Martincus and Carballo, 2012; Schminke and van Biesebroeck, 2016). Other studies outside Latin America, evaluation of production subsidy in China (Girma et al., 2009a) and Germany (Girma et al., 2009b) find significant effects at expanding export volumes - the intensive margin of trade, but not appear to compel non-exporters to enter foreign markets for the first time. In contrast, our evidence is in line with the most recent addition to the literature (Broocks and van Biesebroeck, 2017) that trade promotion polices are successful along the firm-extensive margin than on the intensive one. Overall, the severity of information incompleteness on export activities to sell new products abroad for the first time or to enter new export markets tend to be more important than in ramping up export volumes been already trading and/or to countries that are already among their destination markets.

In classical quasi-experimental approaches, the so-called fundamental evaluation problem of causal inference specified by Holland (1986) needs to account for selection bias. As proposed by Heckman et al. (1997), the standard approach is combining matching with DiD approach, which has not been commonly employed in export promotion evaluation, as validated by only one in four of the estimates reported in the primary studies.[14] Instead, in more than two-third of the estimates reported, researchers include year- and firm-fixed effects to give the DiD interpretation comparing programme participants and non-participants. Selection bias could therefore potentially affect the estimated effects of trade-promotion programmes. While there is no complete fix, dealing with unobservable firm-fixed effects and time-varying effects can successfully reduce selection bias in quasi-experimental impact evaluation. This pattern is confirmed in our MRA in that specifications employing firm- and year-fixed effects lead to a significantly negative effect. The implied coefficient size, however, is equally a small effect by the interpretation of Doucouliagos (2011), but in general the MRA evidence compels to account for confounding factors for programme assignment to be random. Accordingly, it can be assumed that unobservable firm characteristics and time-varying effects influencing grant assignment and exporting can be eliminated if not potentially reduced. It is then only necessary to rely on the selection-on-observables assumption to reduce the main identification problem that participation assignment is non-random.

Lastly, concerning publication characteristics, our result suggests a strong positive association between reported estimates and two moderators of publication status. First,

[14] However, as outlined by Görg et al. (2008) and recently by Cadot et al. (2015), even this combination of matching and DiD estimations might still leave potential problem when there are unobserved time-varying effects.

reported estimates from peer-reviewed journals are more likely to document positive and significant impact of treatment assignment on export performance than unpublished studies. Second, we find that the publication year of the study affects reported estimates, in that recent studies tend to report higher performance estimates. A potential higher quality effect is, therefore, appear to report higher programme assignment benefit.

Although the G-to-S approach in column 1 includes other variables reflecting potential drivers of heterogeneity (treatment estimates with OLS, ATT and PSW among others), the statistical variation in these variables is insignificant to generate any useful information. These variables were jointly statistically insignificant as noted earlier, but also individually insignificant to explain the heterogeneity in the estimates reported (Column 4).

4.5 Further investigation: publication selection bias and genuine effect

Publication selection bias: The inclusion of potential drivers of heterogeneity vanishes the publication selection bias noted earlier. Instead, we find that two publication characteristics strongly positively associated with reported estimates (Table 4). We then interpreted this result as a potential impact of quality, i.e., higher quality studies appear to generate significantly positive impact. However, publication selection pressure can also affect this causality nexus. If journal editors and reviewers for accepting publication or authors prior interest for publication submission is to follow significantly positive estimates, then this may lead to overrepresentation of large or more significant effects, and under representation of smaller or statistically insignificant effects. Similarly, if researchers estimating various econometrics model prefer large or more significant effects, they may also tend to cite recent studies that provide such large estimates. In each case, publication selection pressure or bias influences future peer-reviewed studies and more recent studies to be associated with larger positive evidence. Therefore, in such unclear causality, we ask the question: could it be that studies that are peer-reviewed and more recent studies mediate the identified publication bias. The idea is therefore to test for the driving force of publication bias investigating whether the slope in Eq. 3 is larger for peer-reviewed and newly published studies.

Moreover, we also included the time dimension of the evaluation, which is not related to publication status, but rather data characteristics. We have earlier noted that longer evaluation time span tends to generate more successful programme assignment impact, hence effects appear to take time to materialize. At the same time, the longer the evaluation time period, the difficult the task to distil the treatment effect as other factors could also impact the outcome.[15] The results are reported in Table 5 and the publication characteristics do not appear to be causing publication selection. However, results are mixed for the length of the time span of the dataset. We obtain evidence that the data characteristics in terms of the time span covered in the evaluation of programme assignment does not appear to mediate publication bias (Column 4, when separately

[15] A first concern is that informational (positive) spillovers from programme participants to non-participants has been empirically tested and likely to attenuate any positive effects of export promotion, though potentially take some time to kick in. A second concern is the negative spillover effects of firms receiving treatment on other firms in the market. The treatment can induce competition and increases labour mobility: firms receiving assistance can steal markets, crowd out other firms and even poach labour (hiring the most productive workers from firms not receiving support). Therefore, the programme treatment creates a negative impact on other competing firms, implying the overall impact (in whatever sense) may not necessarily a positive and being far from the underlying true value. Last, programme treatment could give participant firms some protection to facilitate or inhibit export performance.

included). In contrast, the latter tend to be driving force of publication bias when jointly included with the two publications variables (Column 5).[16] The results are therefore reliable across the various specifications for the two publication characteristics, but not for the time span covered by the evaluation. The latter indicates some weak evidence of mediating publication selection.

Table 5
Potential mediating factors of publication bias

Variables	(1) Published (P)	(2) Publication year (Y)	(3) P + Y	(4) No. of Years (NY)	(5) P + Y + NY
Constant	0.006	-0.007	-0.009	-0.028*	-0.066*
	(0.008)	(0.043)	(0.041)	(0.014)	(0.038)
SE	1.387***	1.429	1.340	0.603	0.817
	(0.364)	(1.336)	(1.270)	(0.735)	(1.266)
SE * Published	0.326		0.325		0.225
	(0.514)		(0.513)		(0.442)
SE * Pub. Year		0.009	0.003		-0.021
		(0.094)	(0.086)		(0.069)
SE * No. of Years				0.556	0.537*
				(0.393)	(0.330)
Observations	*1824*	*1824*	*1824*	1824	1824
Studies	*35*	*35*	*35*	*35*	*35*
LR test	*135.22*	*99.21*	*98.19*	*181.19*	*90.68*
$p > \text{chi}^2$	*0.000*	*0.000*	*0.089*	*0.000*	*0.000*

Notes: ***, **, * stand for 1%, 5%, and 10% level, respectively. Standard errors are reported in parentheses. Reported results are estimated from Eq. (3) specification with mixed-effects multilevel model through the restricted maximum likelihood. All columns use inverse variance weights. The moderator variables interacted with SE are also respectively included but the results are not reported.

Genuine effect: In comparison to the initial size of genuine effect, the inclusion of observable drivers of heterogeneity improves the true underlying value (PET) by about three times. However, there are many potential genuine heterogeneity effects than measure of the underlying effect like the case related to a single PET. Instead of reliant on selection of baseline studies in attempt to create synthetic studies that would employ a given approach to estimate the programme assignment impact on trade performance, we follow recent meta-analyses to derive the 'best practice' genuine effect from the multivariate MRA conditional on the identified moderator variables used to capture the heterogeneity (Demena and van Bergeijk, 2017; Floridi et al., 2020). The approach is labelled as best practice as it potentially alleviates omitted variable bias and endogeneity problems on top of accounting for publication selection bias.

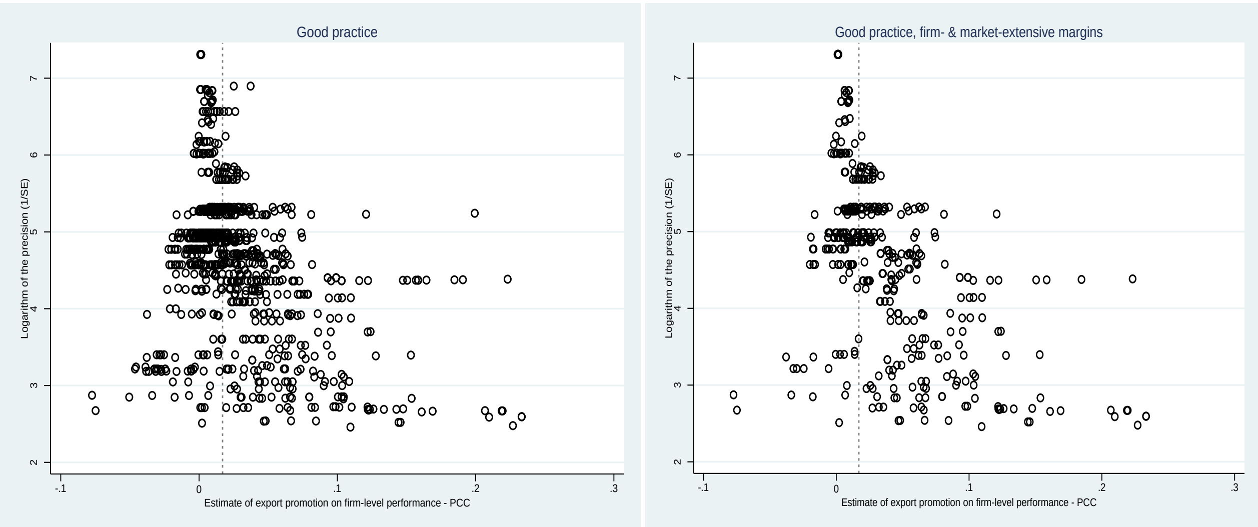
We specify the best practice approach conditional on the moderator variables that are most frequently used by the primary studies included in our analysis. Doing so, we include estimates reported with the number of years of data used in the evaluation, the number of observations from which the primary estimates derive, the target of the programme policy is for SMEs or large firms, the outcome indicator being firm-extensive and destination-

[16] Jointly they are not significant: the joint test of these variables interacted with the standard error reveals that the null hypothesis of a 0 effect cannot be rejected (*p-value* of 0.3752).

extensive margins than for the intensive one, controlling time-invariant unobservable factors and time-varying effects with firm- and year-fixed effects. We also included the two controls for the quality of the studies: the publication year of the study and publication in peer-reviewed journals.[17] The procedure yields the predicted genuine underlying effect conditional on the identified heterogeneity is 0.069, being statistically significant at the 99 per cent confidence level. This implies that the corrected correlation coefficient derived from the best practice approach is considerably larger than any of the effect noted earlier, which testifies the statistical power of the exercise in controlling for the drivers of heterogeneity. In this simple illustration of accounting the genuine effect, the implied gains from trade-promotion polices enhanced by about 68 per cent than the single PET of the multivariate MRA. However, this implied genuine underlying effect is just reached the threshold recommend by Doucouliagos (2011) for interpretation as a small impact.

We have noted earlier in the discussion of factors mediating publication bias, the number of years of data used in the evaluation somehow appears to cause publication bias. The next important issue is therefore whether the best practice choices regarding research design and methods could be the driving force of publication bias. To be specific, we investigate whether the positive correlation between the computed estimates (PCC_{sc}) and the corresponding standard errors (SE_{pccse}) in Eq. 3 is larger in estimates represented by good research design and methods.

Figure 5
Funnel plots show more asymmetrical than Figure 3 (N=943 (left); (N=409 (right)



We start with visual inspection. Funnel plots associated with estimates reported in our earlier best practice research design are reported in Figure 5. Following Havránek and Iršová (2011) and Demena and Bergeijk (2017)[18], we use reported estimates with firm-fixed effects and control for time-varying unobservable effects in the performance

[17] We do so; first most economists would argue that estimates in a peer-reviewed studies are of higher quality than studies that have not undergone peer-review procedures (Stanley and Doucouliagos, 2012); second, almost two-third (59%) of the estimates reported are constructed from studies published in a peer-reviewed journal; third the positive impacts of most recent studies may also indicate quality (Floridi et al., 2020)
[18] Since it is not feasible to apply the full specification of bets practice as a small fraction of the estimates would remain for analysis

regression (the left panel in Figure 5). Next, we extend such aspects of best practice by adding the two significantly positive export performance indicators: the firm- and destination-extensive margins (the right panel in Figure 5). The result of this exercise shows that both funnel plots somehow mimic the pattern of the funnel plot in Figure 4 (when all estimates by any method were considered). But also, the new funnel plots appear to be more asymmetrical (specifically the right funnel plot) – it turns out to seem that some negative estimates are discarded in the performance regression. Although the new funnel plots are thinner than Figure 4 because of the substantial reduction in the estimates considered, the shape and location are less comparable with (more asymmetrical in) the right panel diagram of Figure 5. Therefore, it appears that the reported estimates of the firm- and destination-extensive margins are more likely to mediate publication bias.

Following Stanley et al. (2008); Havránek and Iršová (2011); and Demena and Bergeijk (2017), we test more formally if some aspects of research design have a (stronger) association with publication bias. As illustrated earlier, we do so by interacting the main drivers of heterogeneity that define 'best practice' with the corresponding standard error. If the aspects of research design (best practice) are mediating factors of the publication bias, their interaction with the slope of Eq. 3 will be significant (Demena and Bergeijk, 2017; Bajzik et al., 2020). Adding these interaction variables to our full model in Eq. (6), the results are presented in Table 6 and all in all three findings standout: First, regarding performance regression research design, merely one out of six of these interactions is statistically significant (Columns 1 and 2), implying the best practice approach does not appear to mediate publication selection bias. Second, when we extend such aspects of best practice with the two export performance indicators, we find clear evidence of mediating publication bias (Columns 3 to 6). Indeed, this confirms the selection bias noted in Table 3: severe and substantial bias (respectively for firm-extensive and destination-extensive margins). Third, including the two controls for quality of the studies to the full model of Eq. 6 (Columns 5 – 6), none of these two aspects are statistically significant, once more confirming the results of Table 5 that they do not appear to be causing publication selection bias. Therefore, the overall findings is that consistent with Table 3, the two export performance indicators (firm-extensive and destination-extensive margins) appeared to be associated with mediating factors of publication bias and the results are stable and comparable across specifications.

Table 6
Further investigation for potential mediating factors of publication bias

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.061***	0.061***	0.032*	0.032*	0.032*	0.032*
	(0.014)	(0.014)	(0.017)	(0.017)	(0.017)	(0.017)
SE	0.732	0.765	-1.538	-1.476	-0.814	-1.048
	(1.584)	(1.601)	(1.635)	(1.648)	(2.015)	(2.045)
SE * No. of Years	-0.015	-0.032	0.229	0.218	0.150	0.153
	(0.585)	(0.594)	(0.525)	(0.531)	(0.523)	(0.533)
SE * No. of Obs.	-0.318	-0.307	-0.033	-0.028	-0.043	-0.036
	(0.215)	(0.216)	(0.240)	(0.241)	(0.240)	(0.241)
SE * Year EF	0.325	0.391	0.321	0.383	0.283	0.343
	(0.455)	(0.458)	(0.436)	(0.439)	(0.439)	(0.443)
SE * Firm EF	1.594***	1.538***	1.440***	1.386***	1.433***	1.393***
	(0.548)	(0.549)	(0.518)	(0.519)	(0.515)	(0.519)
SE * SMEs	0.585		0.586		0.615*	
	(0.374)		(0.369)		(0.372)	
SE * Large		-0.049		-0.063		-0.061
		(0.430)		(0.427)		(0.427)
SE * Firm-extensive			2.075***	2.096***	2.011***	2.036***
			(0.516)	(0.517)	(0.521)	(0.524)
SE * Dest.-extensive			1.564***	1.553***	1.567***	1.548***
			(0.383)	(0.384)	(0.384)	(0.385)
SE * Published					-0.012	0.006
					(0.085)	(0.086)
SE * Pub. Year					-0.607	-0.535
					(0.710)	(0.724)
Observations	*1824*	*1824*	*1824*	*1824*	*1824*	*1824*
Studies	*35*	*35*	*35*	*35*	*35*	35
LR test	*102.11*	*103.49*	*92.27*	*91.98*	*76.79*	*79.54*
p > chi^2	*0.000*	*0.000*	*0.089*	*0.000*	*0.000*	*0.000*

Notes: ***, **, * stand for 1%, 5%, and 10% level, respectively. Standard errors are reported in parentheses. Reported results are estimated from Eq. (6) specification with mixed-effects multilevel model through the restricted maximum likelihood. All columns use inverse variance weights. The moderator variables interacted with SE are also respectively included but the estimates are not reported. Columns 1 and 2 present for Data and Estimation characteristics; Columns 3 and 4 add the firm and market extensive margins; and finally, Columns 5 and 6 include the two controls for quality as in Table 5.

4.6 Robustness

We provide a set of robustness checks to test for sensitivity of the reported results. First, we compare our MEM results with the basic CDA and FE (Tables 2 and 4). We find that the identified effects tend to be comparable and stable in magnitude and statistical significance.[19] Second, we also provide results of bootstrapping the standard error (a non-standard cluster adjustment) to account for a potential error-in-variance bias (Columns 4

[19] The main exception is that the statistical significance disappears for the length of data time used and destination-extensive margin for CDA – Column 2 and Firm-extensive margin and the non-linear treatment effect estimate for FE – Column 4).

and 5 in Tables 2 and 4, respectively).[20] We continue to find similar results with the exception that the length of the time span and the two export performance indicators lost statistical significance.[21] Third, we control for two additional dummy variables equalling 1 for estimates derived from interacted coefficients, or for developed country, respectively, and 0 otherwise.[22] Again, we continue to find similar results of the full sample (Columns 6 and 7 in Table 4). We also run our regressions with the inclusion of outliers. Results are reported in Table 7 and once more, corroborating our main findings, the reported estimates are not only comparable and stable in terms of statistical significance and sign, but also in magnitude of the coefficients.[23]

Table 7
Publication bias and underlying genuine effect test: robustness checks

Panel 1 - Excluding interaction terms				
Variables	**(1)**	**(2)**	**(3)**	**(4)**
	CDA	**FE**	**MEM**	**Wild bootstrapped**
Publication bias (FAT - constant)	1.916***	1.904***	1.904***	1.916***
	(0.328)	(0.168)	(0.403)	0.00
Genuine effect (PET - precision)	0.009***	0.010***	0.009***	0.009***
	(0.001)	(0.001)	(0.000)	0.00
Observations	*1365*	*1365*	*1365*	*1365*
Studies	*35*	*35*	*35*	*35*
LR test			*190.64*	
$p >$ chi[2]			*0.000*	
Panel 2 - Including outlier estimates				
Variables	**(1)**	**(2)**	**(3)**	**(4)**
	CDA	**FE**	**MEM**	**Wild bootstrapped**
Publication bias (FAT - constant)	1.560***	1.500***	1.480***	1.560***
	(0.436)	(0.154)	(0.427)	0.00
Genuine effect (PET - precision)	0.009***	0.010***	0.010***	0.009***
	(0.002)	(0.001)	(0.000)	(0.000)
Observations	*1869*	*1869*	*1869*	*1869*
Studies	*35*	*35*	*35*	*35*
LR test			*193.57*	
$p >$ chi[2]			*0.000*	

Notes: See Table 2.

We have also provided further attempt to address model uncertainty in the multivariate MRA. To address this issue, we have initially adopted G-to-S approach. Others have

[20] Yet, the number of primary studies included in our analysis are not limited. For instance, Cameron (2008) suggests bootstrapping when a low number of clusters, i.e., less than 20 groups are used.
[21] It should be noted that as indicated earlier for these estimators (CDA, FE, and bootstrapping standard error), it is the within-study heterogeneity only that matters as between-study is assumed to be 0 (Stanley and Doucouliagos, 2017 and 2015) and in our case there is significant statistical dependency between studies (see Note under Table 2).
[22] In so far, we follow Havránek and Iršová (2011) and Floridi et al. (2020) and use sample means of the interacted variables. To check whether assessing the reported coefficients at sample means of the interacted variables results in different estimates, we follow Demena and van Bergeijk (2017) adding an explanatory variable accounting for estimates from interacted terms.
[23] Specifically, the results for the genuine underlying effects mimic the corresponding findings. Whereas the size of the magnitude for the publication selection bias is higher for estimates that include outliers (Panel 2) and exclude interacted variables (Panel 1) compared to all studies, but this difference is not statistically significant.

adopted the weighted-average least squares (WALS) or Bayesian Model Averaging (BMA) (e.g., Iršová and Havránek, 2013; Ugur et al., 2020). We have additionally run all our regressions using the WALS that suggested to provide similar estimations with the BMA but at a much lower cost in terms of computation time (Magnus et al., 2010; De Luca and Magnus, 2011). Furthermore, De Luca and Magnus (2011) argue that the WALS combines the BMA and frequents approaches and relies on transparent ignorance in the selection of the focus variable(s). This alternative approach to model uncertainty does not change our main results both qualitatively and quantitatively.[24]

5. Conclusion

Despite the policy rational promoting internationalization of domestic firms, findings in the reported empirical estimates are contrasting and thus not conclusive. We conduct a MRA of the large numbers of empirical studies dealing with EPP and firm-level export performance. The analysis has identified 1869 estimated parameters from 37 studies dealing with 21 countries published over the last two decades.

Our results suggest that on average EPP have very weak practical impact on export performances. Staring with publication bias, our initial results suggest authors' preference towards positive estimates: estimates that are in favour of the most commonly cited views in the literature are more preferred and frequently reported and published. However, the heterogeneity in terms of the reported findings motivate our meta-analysis the need for multivariate MRA beyond the FAT-PET bivariate analysis. In this procedure, consistent with results from the earlier approach, the overall average effect of EPP impact on export performance is weak. In contrast, the initial bias towards positive estimates disappears, indicating the importance of controlling for heterogeneities that drive the results before any underlying conclusion is drawn.

The MRA, however, shows that individual aspects of empirical design have a strong influence on the reported estimates of export performance. We uncovered differences across the primary studies are mainly driven by the characteristics of the data, the types of firms targeted, the set of variables controlled in the underlying estimation techniques of the studies, the adopted a four-dimensional view of export performance, and the publication characteristics. Specifically, we find that the number of time span covered by the primary studies tend to significantly affect the reported estimates and thus suggests the policy relevance of longer data timespan to bear more fruits from the potential benefits of EPP. We also find that the firm-extensive and destination-extensive margins are the most positively affected dimensions of export performance. From policy implication, this refers to the severity of information incompleteness on export activities or the necessity of subsidies and grants to sell new products abroad for the first time or to enter new export destination markets tend to be more important than exports to existing product-destination markets – the intensive margin effect. The overall findings of this are consistent with Broocks and van Biesebroeck (2017) along the firm-extensive margin and Martincus and Carballo (2010b) for destination-extensive margin as opposed to the intensive margin of trade. We further obtain that part of the difference in the EPP estimates reported explained by the characteristics of the publication outlets. In particular, the quality of the reported

[24] The results are not reported for space purpose. In our view, it remains an issue for future research whether the G-to-S version, the WALS or the Bayesian model averaging should only be used.

estimates captured by the publication status and the publication year of the study are positively and significantly associated with study results - higher quality effects appear to report higher programme assignment benefit. We then controlled for both publication quality of the study and publication selection bias to explore whether studies that appear in peer-reviewed journals and recently published are the driving force of publication selection pressure. Doing so, we obtain that publication quality does not appear to be causing publication selection bias.

Good practice methodology in estimating the impact of EPPs yields better estimate of the underlying genuine impact. After controlling for publication selection bias and reducing potential endogeneity issues for instance unobserved heterogeneity or omitted variable bias, the implied gains from trade-promotion polices is about 0.069. This suggests that the implied genuine underlying effect is just reached the small impact threshold recommend by Doucouliagos (2011). The overall positive but small effects corroborate with the existing findings notably after controlling for possible endogeneity (van Biesebroeck et al., 2015; Munch and Schaur, 2018). Finally, we further explore that the best practice approach does not appear to mediate publication selection pressure. Unlike the research regression design (evaluation technique), we find robust evidence that the identified two-dimensional view of export performance appeared to be associated with mediating factors of publication bias. In our robustness checks, results are consistent and comparable to the use of alternative estimation methods and addressing model uncertainty.

References

Abdullah, A., Doucouliagos, H. and Manning, E. (2015) Does education reduce income inequality? A meta-regression analysis. *Journal of Economic Surveys*, *29*(2), pp.301-316.

Afesorgbor, S.K. (2017) Revisiting the effect of regional integration on African trade: evidence from meta-analysis and gravity model. *The Journal of International Trade & Economic Development*, *26*(2), pp.133-153.

Bajzik, J., Havránek, T., Iršová, Z. and Schwarz, J. (2020) Estimating the Armington elasticity: The importance of study design and publication bias. *Journal of International Economics*, *127*, p.103383.

Balima, H.W., Kilama, E.G. and Tapsoba, R. (2020) Inflation targeting: Genuine effects or publication selection bias? *European Economic Review*, *128*, p.103520.

Bernard, A.B. and Jensen, J.B. (2004) Why some firms export. *Review of economics and Statistics*, *86*(2), pp.561-569.

Cameron, A.C., Gelbach, J.B. and Miller, D.L. (2008) Bootstrap-based improvements for inference with clustered errors. *The* Review *of Economics and Statistics*, *90*(3), pp.414-427.

Copper, H.M. and L.V. Hedges (Eds) (1994) *Handbook of research synthesis.* New York: Russell Sage.

Creusen, H. and Lejour, A. (2013) Market entry and economic diplomacy. *Applied Economics Letters*, *20*(5), pp.504-507.

Cruz, M., Lederman, D. and Zoratto, L. (2018) The anatomy and the impact of export promotion agencies. In van Bergeijk, P.A.G. and Moons, S.J. (eds), *Research Handbook on Economic Diplomacy*. Cheltenham: Edward Elgar.

De Luca, G. and Magnus, J.R. (2011) Bayesian model averaging and weighted-average least squares: Equivariance, stability, and numerical issues. *The Stata Journal*, *11*(4), pp.518-544.

Demena, B.A. (2015) Publication bias in FDI spillovers in developing countries: a meta-regression analysis. *Applied Economics Letters*, *22*(14), pp.1170-1174.

Demena, B.A. (2017) Essays on Intra-Industry Spillovers from FDI in Developing Countries: A Firm-Level Analysis with a Focus on Sub-Saharan Africa. PhD diss. Erasmus University, The Hague.

Demena, B.A. and van Bergeijk, P.A.G. (2017) A meta-analysis of FDI and productivity spillovers in developing countries. *Journal of Economic Surveys*, *31*(2), pp.546-571.

Demena, B.A. and Afesorgbor, S.K. (2020) The effect of FDI on environmental emissions: Evidence from a meta-analysis. *Energy Policy*, *138*.

Demena, B. A., van Bergeijk, P.A.G, Reta, A., Jativa, G.B. and Kimararungu, P. (2021a) Publication bias of economic sanction research: A Meta-analysis of the impact of trade-linkage, duration and prior relations on sanction success, Edward Elgar (forthcoming).

Demena, B.A, Msami, J., Mmari, D., and van Bergeijk, P.A.G. (2021b) Productivity premia and firm heterogeneity in Eastern Africa. *ISS Working Paper No. 680*.

Doucouliagos, C. (2005) Publication bias in the economic freedom and economic growth literature. *Journal of Economic Surveys, 19*(3), pp.367-387.

Doucouliagos, H., Laroche, P. and Stanley, T.D. (2005) Publication bias in union-productivity research? *Relations Industrielles/Industrial Relations*, *60*(2), pp.320-347.

Doucouliagos, H. and Stanley, T.D. (2009) Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, *47*(2), pp.406-428.

Doucouliagos, H. and Laroche, P. (2009) Unions and Profits: A Meta-Regression Analysis. *Industrial Relations: A Journal of Economy and Society*, *48*(1), pp.146-184.

Doucouliagos, H. (2011) How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics (No. 2011_5). Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Doucouliagos, C. and Stanley, T.D. (2013) Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys*, *27*(2), pp.316-339.

Duflo, E., Glennerster, R. and Kremer, M. (2007) Using randomization in development economics research: A toolkit. *Handbook of development economics*, *4*, pp.3895-3962.

Floridi, A., Demena, B.A. and Wagner, N. (2020) Shedding light on the shadows of informality: A meta-analysis of formalization interventions targeted at informal firms. *Labour Economics*, *67*, p.101925.

Gadd, H., Hansson, G. and Månsson, J. (2008) *Evaluating the impact of firm subsidy using a multilevel propensity score approach* (No. 2009: 3). Linnaeus University, Centre for Labour Market Policy Research (CAFO), School of Business and Economics.

Geldres-Weiss, V.V. and Monreal-Pérez, J. (2018) The Effect of Export Promotion Programs on Chilean Firms' Export Activity: A Longitudinal Study on Trade Shows and Trade Missions. *Journal of Promotion Management*, *24*(5), pp.660-674.

Hadi, A.S. (1994) A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(2), pp.393-396.

Havránek, T. and Iršová, Z. (2011) Estimating vertical spillovers from FDI: Why results vary and what the true effect is. *Journal of International Economics*, *85*(2), pp.234-244.

Havránek, T., Horvath, R. and Zeynalov, A. (2016) Natural resources and economic growth: A meta-analysis. *World Development*, *88*, pp.134-151.

Havránek, T., Stanley, T.D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W.R., Rost, K. and Van Aert, R.C.M. (2020) Reporting guidelines for meta-analysis in economics. *Journal of Economic Surveys*, *34*(3), pp.469-475.

Hayakawa, K., Lee, H.H. and Park, D. (2014) Do export promotion agencies increase exports? *The Developing Economies*, *52*(3), pp.241-261.

Heckman, J.J., Ichimura, H. and Todd, P.E. (1997) Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, *64*(4), pp.605-654.

Hiller, S. (2012) Does export promotion work in Denmark? Evidence from a matching approach. Working Paper. Available at http://idbdocs. iadb. org/wsdocs/getdocument. aspx.

Iršová, Z. and Havránek, T. (2013) Determinants of horizontal spillovers from FDI: Evidence from a large meta-analysis. *World Development*, *42*, pp.1-15.

Kluve, J., Puerto, S., Robalino, D., Romero, J.M., Rother, F., Stöterau, J., Weidenkaff, F. and Witte, M. (2019) Do youth employment programs improve labor market outcomes? A quantitative review. *World Development*, *114*, pp.237-253.

Lederman, D., Olarreaga, M. and Payton, L. (2010) Export promotion agencies: Do they work? *Journal of development economics*, *91*(2), pp.257-265.

Magnus, J.R., Powell, O. and Prüfer, P. (2010) A comparison of two model averaging techniques with an application to growth empirics. Journal of econometrics, 154(2), pp.139-153.

Melitz, M. (2003) The impact of trade on aggregate industry productivity and intra-industry reallocations. *Econometrica*, *71*(6), pp.1695-1725.

Moons, S.J. and van Bergeijk, P.A. (2017) Does economic diplomacy work? A meta-analysis of its impact on trade and investment. *The World Economy*, *40*(2), pp.336-368.

Ravallion, M. (2007) Evaluating anti-poverty programs. *Handbook of development economics*, *4*, pp.3787-3846.

Stanley, T.D. (2005) Beyond publication bias. *Journal of economic surveys*, *19*(3), pp.309-345.

Stanley, T.D. (2008) Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and statistics, 70* (1), pp.103-127.

Stanley, T.D. and H. Doucouliagos (2012) *Meta-Regression Analysis in Economics and Business*. Oxford: Routledge.

Stanley, T.D., Doucouliagos, C. and Jarrell, S.B. (2008) Meta-regression analysis as the socioeconomics of economics research. *The Journal of Socio-Economics*, *37*(1), pp.276-292.

Stanley, T.D. and Doucouliagos, H. (2015) Neither fixed nor random: weighted least squares meta-analysis. *Statistics in medicine*, *34*(13), pp.2116-2127.

Stanley, T.D. and Doucouliagos, H. (2017) Neither fixed nor random: weighted least squares meta-regression. *Research synthesis methods*, *8*(1), pp.19-42.

Tukey, J.W. (1977) Exploratory Data Analysis. Reading, MA: Addison—Wesley.

Ugur, M., Churchill, S.A. and Luong, H.M. (2020) What do we know about R&D spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power. *Research Policy*, *49*(1), p.103866.

van Bergeijk, P.A.G., Yakop, M. and de Groot, H.L. (2011) The economic effectiveness of diplomatic representation: An economic analysis of its contribution to bilateral trade. *The Hague Journal of Diplomacy*, *6*(1-2), pp.101-120.

van Bergeijk, P.A.G. and Moons, S.J. eds. (2018) *Research Handbook on Economic Diplomacy: Bilateral Relations in a Context of Geopolitical* Change. Cheltenham: Edward Elgar.

van Bergeijk, P.A.G., Demena, B.A., Reta, A., Jativa, G.B. and Kimararungu, P. (2019) Could the literature on the economic determinants of sanctions be biased? *Peace Economics, Peace Science and Public Policy*, *25*(4).

WAIPA and WBG (2020) State of Investment Promotion Agencies, Evidence from WAIPA-WBG's Joint Global Survey, https://waipa.org/blog/state-of-investment-promotion-agencies/ accessed 23 April 2021.

Yakop, M. and van Bergeijk, P.A.G. (2011) Economic diplomacy, trade and developing countries. *Cambridge Journal of Regions, Economy and Society*, *4*(2), pp.253-267.

Appendices

Figure A1
Sign and significance of EPP impact on firm-level performance (N=1869)

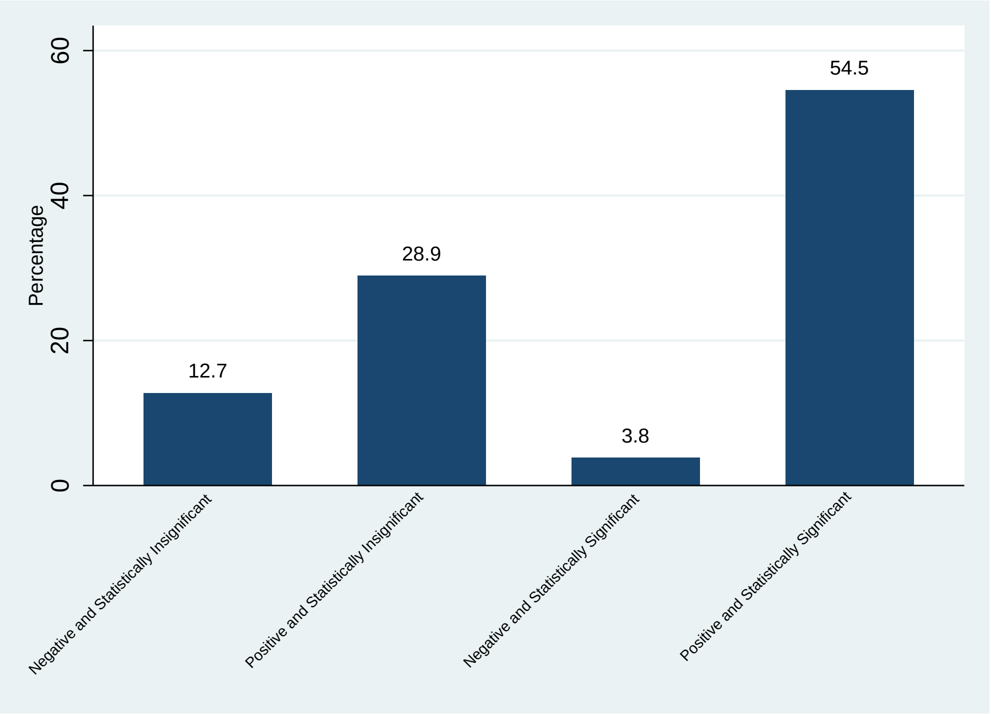


Table A1
Studies included in the meta-analysis: Overview of the evidence base

Study (year)	Pub type	Country	Data start	Data end	No of est.	Mean effect size	St. Dev.	Range	
Bernini and Treibich (2016)	PR	France	1997	2007	16	0.007	0.006	-0.002	0.019
Bonner and McGuinness (2007)	PR	Ireland	1995	1999	30	0.037	0.061	-0.076	0.132
Breinlich et al. (2017)	WP	UK	2013	2014	4	-0.005	0.011	-0.014	0.011
Broocks and van Biesebroeck (2017)	PR	Belgium	2006	2010	70	0.042	0.034	0.005	0.154
Cadot et al. (2012)	WP	Tunisia	2000	2010	192	0.036	0.035	-0.038	0.196
Cadot et al. (2015)	PR	Tunisia	2000	2010	216	0.008	0.017	-0.022	0.082
Castillo et al. (2010)	WP	Argentina	1999	2008	32	0.007	0.006	0.001	0.020
Cansino et al. (2013)	PR	Spain	2007	2010	4	0.180	0.019	0.152	0.194
Crespi and Alvarez (2000)	PR	Chile	1992	1996	12	0.043	0.076	-0.067	0.164
Cruz (2014)	WP	Brazil	2005	2010	64	0.062	0.041	0.004	0.153
Girma et al. (2009a)	PR	China	1999	2005	44	0.008	0.007	-0.004	0.026
Görg et al. (2005)	WP	Ireland	1983	1998	30	0.041	0.055	-0.036	0.211
Görg et al. (2008)	PR	Ireland	1983	1998	24	0.034	0.061	-0.044	0.215
Gourdon et al. (2011)	BC	Tunisia	2004	2008	38	0.126	0.054	0.040	0.250
Helmers and Trofimenko (2009)	WP	Colombia	1981	1991	16	0.009	0.043	-0.037	0.103
Helmers and Trofimenko (2010)	WP	Colombia	1981	1991	16	0.016	0.076	-0.081	0.187
Helmers and Trofimenko (2013)	PR	Colombia	1981	1991	16	0.025	0.102	-0.091	0.228
Kim et al. (2018)	PR	Vietnam	2014	2016	23	0.044	0.049	-0.046	0.190
Lederman et al. (2016)	PR	Argentina	2006	2010	10	0.080	0.055	0.030	0.225
Martincus and Carballo (2008)	PR	Peru	2001	2005	36	0.077	0.071	0.013	0.235
Martincus and Carballo (2010a)	PR	Chile	2002	2006	32	0.045	0.060	0.000	0.238
Martincus and Carballo (2010b)	PR	Uruguay	2000	2007	4	0.089	0.071	0.008	0.179
Martincus and Carballo (2010c)	PR	Colombia	2003	2006	17	0.060	0.073	0.004	0.228
Martincus et al. (2012)	PR	Argentina	2002	2006	153	0.017	0.013	-0.004	0.067
Martincus and Carballo (2012)	PR	Costa Rica	2001	2006	93	0.040	0.051	-0.050	0.233
Mion and Muûls (2015)	ER	UK	2008	2012	110	0.014	0.044	-0.113	0.194
Munch and Schaur (2015)	WP	Denmark	1999	2012	24	0.068	0.052	0.003	0.161
Munch and Schaur (2018)	PR	Denmark	1999	2012	94	0.058	0.056	-0.013	0.195
Rincón-Aznar et al. (2015)	ER	UK	2006	2012	12	0.023	0.006	0.014	0.035
Schminke and van Biesebroeck (2016)	WP	Belgium	2006	2010	70	0.044	0.039	0.007	0.227
Silva (2011)	WP	Portugal	1996	2003	37	0.011	0.058	-0.077	0.143
van Biesebroeck et al. (2010)	WP	Canada	1999	2006	66	0.002	0.008	-0.020	0.029
van Biesebroeck et al. (2011)	WP	Canada	1999	2006	39	0.016	0.048	-0.104	0.203
van Biesebroeck et al. (2015)	PR	Canada	1999	2006	84	0.053	0.060	-0.052	0.223
van Biesebroeck et al. (2016)	PR	Belgium and Peru	2006	2011	96	0.038	0.038	-0.038	0.125

Note: PR - peer-reviewed articles, WP - working papers, ER - evaluation report, BC – book chapter.

Table A2
Description and summary statistics of variables

Moderator Variables	Description	Mean	Std. Dev.
	Outcome characteristics		
PCC	Partial correlation coefficient	0.034	0.049
PCC_SE	Standard error of PCC	0.019	0.020
	Data characteristics		
No. Years	The logarithm of the number of years of data used	1.808	0.549
No. of firms	The logarithm of the number of firms or observations	9.060	2.185
Developed	=1 if data come from a developed country (data from a developing country as reference category)	0.414	0.493
Evaluation features			
	Evaluation timing follow-up		
One	=1 if evaluation follow-up is one year after export promotion treatment (less than one year as reference category)	0.152	0.359
More_Years	=1 if evaluation follow-up is two or more years after export promotion treatment	0.654	0.476
	Export promotion target		
SMEs	=1 if export promotion programme targets SMEs (policies targeting any firm as reference category)	0.119	0.324
Large	=1 if export promotion programme targets large firms	0.094	0.292
	Intervention policy		
Service	=1 if intervention or trade-promotion is providing various services (mixed or a policy not listed here as reference category)	0.543	0.498
Subsidy	=1 if intervention or trade-promotion is providing subsidy or grants	0.379	0.485
	Treatment effect estimates		
Quasi	=1 if evaluation design is quasi-experimental approach (randomized experiment as reference category)	0.980	0.121
OLS	=1 if estimation method is multivariate linear regression (IV, Heckman or other method not listed here as reference category)	0.085	0.279
PSW	=1 if estimation method is propensity score weighted regression	0.235	0.424
DiD	=1 if estimation method is simple DiD	0.084	0.277
mDiD	=1 if estimation method is DiD and matching combined	0.260	0.439
Non-linear	=1 if estimation method is multivariate non-linear regression – probit, logit, tobit	0.103	0.303
ATT	=1 if treatment effect estimate is average treatment effect on treated (average treatment effect – ATE, local average treatment effect – LATE, other method not listed here as reference category)	0.861	0.346
	Firm performance indicator		
Incidence	=1 if export performance indicator is incidence of export – the firm-extensive margin (change in export intensity – the intensive margin as reference category)	0.158	0.365
Destination	=1 if export performance indicator is new destination market – destination-extensive margin	0.262	0.439
Product	=1 if export performance indicator is new product – product-extensive margin	0.185	0.389
	Specification characteristics		
Age	=1 if specification controls for age of the firm since operation or establishment	0.522	0.499
Size	=1 if specification controls for size of the firm	0.663	0.473
Year-FE	=1 if specification controls for year fixed effects	0.611	0.488
Firm-FE	=1 if specification controls for firm fixed effects	0.650	0.477
Interaction terms	=1 if coefficient results from interaction variables	0.252	0.434
	Publication characteristics		
Publication Year	Publication year or age of the study (base, 2000)	13.080	3.064
Reviewed	=1 if published in a peer-reviewed journal	0.589	0.492
Citations	Logarithm of citations in Google Scholar per the age of the study, as of October 2020	1.855	1.145
Impact	Recursive journal impact factor from RePEc	0.513	0.722

Table A3
List of primary studies used in meta-analysis

- Bernini, M. and Treibich, T. (2016) Killing a second bird with one stone? Promoting firm capital growth and exports through tax policy. *Industrial and Corporate Change*, *25*(5), pp.829-845. *
- Bonner, K. and McGuinness, S. (2007) Assessing the impact of marketing assistance on the export performance of Northern Ireland SMEs. *International Review of Applied Economics*, *21*(3), pp.361-379. *
- Breinlich, H., Donaldson, D., Nolen, P.J. and Wright, G.C. (2017) Information, perceptions and exporting evidence from a randomized controlled trial. *
- Broocks, A. and Van Biesebroeck, J. (2017) The impact of export promotion on export market entry. *Journal of International Economics*, *107*, pp.19-33. *
- Cadot, O., Fernandes, A.M., Gourdon, J. and Mattoo, A. (2012) *Are the Benefits of Export Support Durable? Evidence from Tunisia*, Policy Research WP (No. 2012-30). *
- Cadot, O., Fernandes, A., Gourdon, J. and Mattoo, A. (2015) Are the Benefits of Export Support Durable? Evidence from Tunisia. *Journal of International Economics*, *97*(2), 310-324. *
- Cameron, A.C., Gelbach, J.B. and Miller, D.L. (2008) Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), pp.414-427.
- Cansino, J.M., Lopez-Melendo, J., Pablo-Romero, M.D.P. and Sánchez-Braza, A. (2013) An economic evaluation of public programs for internationalization: The case of the Diagnostic program in Spain. *Evaluation and Program Planning*, *41*, pp.38-46. *
- Castillo, V., Maffioli, A., Monsalvo, A.P., Rojo, S. and Stucchi, R. (2010) Can SME Policies Improve Firm Performance? Evidence from an Impact Evaluation in Argentina. IDB-OVE Working Paper No. 06/10. *
- Crespi, G. and Alvarez, R. (2000) Exporter performance and promotion instruments: Chilean empirical evidence. *Estudios de economía*, *27*(2), pp.225-241. *
- Cruz, M. (2014) Do export promotion agencies promote new exporters? World Bank Policy Research Working Paper No. 7004. *
- Geldres-Weiss, V.V. and Carrasco-Roa, J.A. (2016) Impact evaluation of national export promotion programs on export firms using contrast groups. *International Journal of Export Marketing*, *1*(1), pp.77-95. *
- Girma, S., Gong, Y., Görg, H. and Yu, Z. (2009a) Can production subsidies explain China's export performance? Evidence from firm-level data. *Scandinavian Journal of Economics*, *111*(4), pp.863-891. *
- Girma, S., Görg, H. and Wagner, J. (2009b) Subsidies and Exports in Germany. First Evidence from Enterprise Panel Data. *Applied Economics Quarterly*, *55*(3), p.179. *
- Görg, H., Henry, M. and Strobl, E. (2005) Grant support and exporting activity. Working paper. *
- Görg, H., Henry, M. and Strobl, E. (2008) Grant support and exporting activity. *The review of economics and statistics*, *90*(1), pp.168-174. *
- Gourdon, J., Marchat, J.M., Sharma, S.I.D.D.H.A.R.T.H. and Vishwanath, T. (2011) Can Matching Grants Promote Exports? Evidence from Tunisia’s FAMEX II Programme. *Where to Spend the Next Million*, *81*. *
- Helmers, C. and Trofimenko, N. (2009) *Export subsidies in a heterogeneous firms framework* (No. 1476). Kiel Working Paper, No. 1476. *
- Helmers, C. and Trofimenko, N. (2010) Export subsidies in a heterogeneous firms framework: Evidence from Colombia. *Centre for the Study of African Economies Working Paper Series*, (2010-26). *
- Helmers, C. and Trofimenko, N. (2013) The use and abuse of export subsidies: evidence from Colombia. *The World Economy*, *36*(4), pp.465-486. *
- Kim, Y.R., Todo, Y., Shimamoto, D. and Matous, P. (2018) Are seminars on export promotion effective? Evidence from a randomised controlled trial. *The World Economy*, *41*(11), pp.2954-2982. *
- Lederman, D., Olarreaga, M. and Zavala, L. (2016) Export promotion and firm entry into and survival in export markets. *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, *37*(2), pp.142-158. *
- Martincus, C.V. and Carballo, J. (2008) Is export promotion effective in developing countries? Firm-level evidence on the intensive and the extensive margins of exports. *Journal of International Economics*, *76*(1), pp.89-106. *
- Martincus, C.V. and Carballo, J. (2010a) Beyond the average effects: The distributional impacts of export promotion programs in developing countries. *Journal of Development Economics*, *92*(2), pp.201-214. *

- Martincus, C.V. and Carballo, J. (2010b) Entering new country and product markets: does export promotion help? *Review of World Economics*, *146*(3), pp.437-467. *
- Martincus, C.V. and Carballo, J. (2010c) Export promotion: bundled services work better. *The World Economy*, *33*(12), pp.1718-1756. *
- Martincus, C.V., Carballo, J. and Garcia, P.M. (2012) Public programmes to promote firms' exports in developing countries: Are there heterogeneous effects by size categories? *Applied Economics*, *44*(4), pp.471-491. *
- Martincus, C.V. and Carballo, J. (2012) Export promotion activities in developing countries: What kind of trade do they promote? *The Journal of International Trade & Economic Development*, *21*(4), pp.539-578. *
- Mion, P.G. and Muûls, M. (2015) The impact of UKTI trade services on the value of goods exported by supported firms final report. *
- Munch, J. and Schaur, G. (2015) The effect of export promotion on firm-level performance. *
- Munch, J. and Schaur, G. (2018) The effect of export promotion on firm-level performance. *American Economic Journal: Economic Policy*, *10*(1), pp.357-87. *
- Rincón-Aznar, A., Riley, R. and Rosso, A. (2015) Evaluating the impact of UKTI trade services on the performance of supported firms. *Report to the UKTI, the National Institute for Economic and Social Research*. *
- Schminke, A. and J. Van Biesebroeck (2016) Evaluation of export promotion policies in Belgium', Mimeo, KU Leuven. *
- Silva, A. (2011) The role of subsidies for exports: Evidence for Portuguese manufacturing firms. *GEE Papers*, *35*, p.19. *
- van Biesebroeck, J., Yu, E. and Chen, S. (2010) The impact of trade promotion services on Canadian exporter performance. K.U. Leuven, Centre for Economic Studies, Naamsestraat 69, 3000 Leuven, Belgium; and Office of the Chief Economist, Foreign Affairs and International Trade Canada. *
- van Biesebroeck, J., Konings, J. and Volpe Martincus, C. (2011). Did export promotion help firms weather the crisis? Center for Economic Studies Discussions Paper Series (DPS) 11.21. *
- van Biesebroeck, J., Konings, J. and Volpe Martincus, C. (2016) Did export promotion help firms weather the crisis?. *Economic Policy*, *31*(88), pp.653-702. *
- van Biesebroeck, J., Yu, E. and Chen, S. (2015) The impact of trade promotion services on Canadian exporter performance. *Canadian Journal of Economics/Revue canadienne d'économique*, *48*(4), pp.1481-1512. *