Comparing Apples to Apples:
Estimating Consistent Partial Effects of Preferential Economic Integration Agreements

Peter H. Egger* and Filip Tarlea**

Abstract

The dominant paradigm of the estimation of causal partial effects of preferential economic integration agreements (PEIAs; e.g., customs unions or free-trade areas) on trade costs and trade flows is to rely on selection on observables, with propensity-score matching being the leading example. Conditional on some metric (score) of observable joint determinants PEIAs and trade flows, the causal partial effect of PEIAs on trade is obtained from a simple mean comparison of trade flows between members and non-members. A key prerequisite for this approach to obtain consistent estimates is that the score is balanced: similarity of country pairs in the score (the propensity of PEIA membership) means similarity in each and everyone of the observables. A violation of this assumption may lead to biased estimates of the effects. We employ a remedy of this bias through entropy balancing, demonstrate that there is an upward bias of PEIA effects on trade flows from lack of balancing, and quantify the bias for partial as well as general equilibrium effects.

Keywords: Preferential economic integration agreements; Causal effects; Propensity score estimation; Entropy balancing; Weighting regression; Balancing property; Gravity models

JEL codes: F10; F12; F17

---

*ETH Zurich, LEE G101, Leonhardstrasse 21, 8092 Zurich, Switzerland. Co-affiliations: CEPR, CESifo, GEP, and WIFO.

**ETH Zurich, LEE G101, Leonhardstrasse 21, 8092 Zurich, Switzerland.

Acknowldgements: To be added.

1 Introduction

Obtaining valid estimates of the partial (or direct) effects of the membership in preferential economic integration agreements (PEIAs) on bilateral trade flows is the primary object of interest of a cottage literature in international economics (see, e.g., Ghosh and Yamarik, 2004; Carrére, 2006; Baier and Bergstrand, 2007, 2009; Egger, Egger, and Greenaway, 2008; Chang and Lee, 2011), and using consistent estimates in quantitative models is vital to obtain reasonable estimates of general equilibrium (or total) economic responses (see Egger and Larch, 201X; Egger, Larch, Staun, and Winkelmann, 201X; Caliendo and Parro, 2015). The econometric problem with this task is that PEIAs are meant to stimulate trade,[1] and, according to economic theory, concluding PEIAs has greater benefits for natural trading partners than otherwise (see Frankel, Stein, and Wei, 1996; Baier and Bergstrand, 2004; Egger and Larch, 2008). Hence, PEIA membership is not randomly assigned to country pairs, which is confirmed by a glance on the frequency of such agreements across types of countries and country-pairs in terms of observable characteristics capturing country size, per-capita income, geography, and remoteness. An influential paper by Baier and Bergstrand (2004) illustrated that the fundamental drivers of trade flows alone explain a lion's share in the variation of binary preferential trade-agreement (PTA) indicators as one form of PEIAs. Egger and Wamser (2013) demonstrate that this is the case also for other forms of PEIAs such as bilateral investment treaties (BITs) or double-tax treaties (DTTs).

The theoretical arguments put forward in earlier work suggest that it will be hard if not impossible to find fundamentals which directly determine

[1] On a broader scheme, PEIAs do not only include preferential trade agreements, but even preferential investment agreements and double-taxation treaties explicitly aim at stimulating trade beyond investment.

PEIAs while influencing trade flows exclusively through PEIA membership. Econometrically speaking, this means that it will be virtually impossible to find identifying instruments for PEIAs for which exclusion restrictions are met in trade-flow regressions, as would be required for instrumental-variables regression. Consequently, the leading assumption in empirical work geared towards estimating PEIA effects is one of the so-called *selection on observables*. According to this framework, it should be possible – guided by economic theory as in Baier and Bergstrand (2004) – to (i) identify all *joint determinants* of PEIA membership and trade flows, and (ii) to condition in some way on them so that the remainder (conditional) variation in PEIA membership and trade flows reveals the causal effect of the former on the latter.

While earlier work used a log-linear regression approach for identification conditional on observables (see Aitken, 1973; Soloaga and Winters, 2001), more recent work resorted to nonparametric estimation techniques (see, e.g., Egger, Egger, and Greenaway, 2008; Baier and Bergstrand, 2009). The latter – with the most prominent example in related applied work being propensity-core matching (see Rosenbaum and Rubin, 1983) – relies on the idea of obtaining a compact metric which captures the joint fundamentals behind PEIA membership and trade flows, and which permits determining similar country pairs which more or less solely differ in terms of PEIA membership for identification of the treatment effect. A prerequisite for this approach is that similarity in terms of the compact, scalar-valued score metric (the propensity of PEIA membership) is not an artifact which could flow from largely different individual observable fundamentals whose differences between PEIA members and non-members are eliminated through aggregation into the score. If that were the case, one would compare PEIA-member *ap-*

ples to -non-member *oranges*. Econometrically, this problem is referred to as a lack of *balancing* of the observables, whereby members and non-members of PEIAs with similar-valued propensity scores of being a PEIA member would have very different moments in the distribution of at least some of the observables the score is based on. Lack of balancing may lead to a bias in the estimates of partial PEIA effects on outcomes such as bilateral trade flows.

The goal of this paper is enforce the balancing of observables by a relatively modern method, entropy balancing (see XXX), illustrate that the usually-employed observables lack balancing, and compare causal PEIA-effect estimates on trade flows (partial effects) and welfare (general-equilibrium effects) among leading methods with the proposed estimates. In a large panel of 1,801,930 observations for all years 1961-2008 and (at least) three types of PEIAs (PTA-, BIT-, DTT-membership, and all combinations thereof – distinguishing between PTAs of different depth in some of the analysis) the paper demonstrates that the lack of balancing of the covariates in a customary nonparametric selection-on-observables approach leads to upward-biased PEIA effects. For instance, the partial impact of a membership in an average PTA alone is estimated to be almost 7 percentage points lower with enforced covariate balancing than without it. The partial effect of a membership in an average BIT alone is estimated to be almost 15 percentage points lower with enforced covariate balancing than without it, and the bias in the estimated partial impact of a membership in an average DTT alone is estimated at a similar magnitude. We illustrate that the quantitative importance of proper conditioning on the covariates in nonparametric selection-on-observables approaches relative to not doing so is of a similar magnitude as the difference between simple (biased) ordinary-least-squares estimates and simple (and also biased) selection-on-observables estimates of partial PEIA treatment effects as relied upon in earlier work. Hence, conventional approaches towards estimating causal PEIA effects tend to overestimate the effects of PEIAs to a nontrivial extent.

The remainder of the paper is organized as follows. The subsequent section briefly portrays nonparametric selection-on-observables estimates of PEIA treatment effects as weighting estimators and distinguishes between covariate-balancing-enforcing and -not-enforcing approaches. Section 3 introduces the specification of the vector of observables and the underlying data considered. Section 4 summarizes estimates of the comparison (propensity) score and illustrates the degree of lack of balancing of the covariates. Moreover, this section demonstrates the difference between the scores of a covariate-balancing-enforcing approach and an -not-enforcing one. This section also provides a comparison of the estimates of partial effects of PEIA treatment effects on bilateral exports. The last section concludes with a brief summary.

2 Causal partial PEIA-effects estimation as weighting regression

The customary conditioning-on-observables approaches towards estimating causal partial PEIA effects can all be portrayed as variants of weighting regressions (see Wooldridge, 2007; Huber, 2014). With this in mind, the simple linear conditioning approach in the form of ordinary least squares of log bilateral exports on one or more PEIA indicator variables and a linear function of observable control variables conforms to an approach with identical weights for each observation. Also matching on the propensity score (of PEIA membership) can be represented as a weighting regression. However, neither linear regression nor matching on the propensity score ensure that the distributions of *all* the joint determinants (the observables) are the same between PEIA members and nonmembers. But only then the two groups would be fully comparable, and we could speak of a quasi-randomization of PEIA membership. The reason is that the linear index with OLS or the propensity score with matching may take on similar values when the individual covariates are quite different in a few or many dimensions of the observables.[2] However, there are weighting approaches which are capable of ensuring comparability in a defined set of moments of the distributions of the observables. One such weighting approach is entropy balancing, which is based on optimally-chosen weights as a function of the distributions of observables for the treated and the untreated (see Hainmüller, 2012; Imai and Ratkovic, 2014; Zubizarreta, 2015). In what follows, we will briefly describe this approach in comparison to inverse-probability-weighting regression.

2.1 Notation

Let us use $P_{ijs}^{\theta 0}$ to denote the propensity score of exporter i and importer j to be members of a PEIA of type θ rather than being a member in no PEIA whatsoever at time s. Denoting the binary indicator for specific PEIA memberships by $T_{ijs}^{\theta 0}$ and the specific realization of θ for ijs by Θ_{ijs}. $T_{ijs}^{\theta 0}$ is unity in case that i and j have a PEIA of type θ at time s and zero else, and denoting the vector of observables determining membership for observations in state θ or 0 by $H_{ijs}^{\theta 0}$, the propensity score is defined as the conditional probability of having treatment θ relative to 0 on the joint determinants of

[2]In empirical work, it is sometimes tested whether the individual averages (means, first moments of the distribution) of the observables are the same between the treated or not. However, even that is not sufficient, as also higher moments of the covariate distributions ought to be the same between the treated and the control observations (see Huber, 2011).

outcome and $T_{ijs}^{\theta 0}$, $H_{ijs}^{\theta 0}$:

$$\widehat{P}_{ijs}^{\theta 0} = P(T^{\theta 0} = 1 | H_{ijs}^{\theta 0}). \tag{1}$$

In this paper we consider at least three types of binary PEIA indicators, so that there there are $2^3 = 8$ possible combinations of agreement types, one being no PEIA agreement of any kind in place which will serve as the general *control* or comparison state in this paper.[3] The remaining seven combinations are

$$\Theta_{ijs} = \theta \in \begin{cases} PTA & \text{if } PTA_{ijs} = 1, BIT_{ijs} = 0, DTT_{ijs} = 0 \\ BIT & \text{if } PTA_{ijs} = 0, BIT_{ijs} = 1, DTT_{ijs} = 0 \\ DTT & \text{if } PTA_{ijs} = 0, BIT_{ijs} = 0, DTT_{ijs} = 1 \\ PTA\&BIT & \text{if } PTA_{ijs} = 1, BIT_{ijs} = 1, DTT_{ijs} = 0 \\ PTA\&DTT & \text{if } PTA_{ijs} = 1, BIT_{ijs} = 0, DTT_{ijs} = 1 \\ BIT\&DTT & \text{if } PTA_{ijs} = 0, BIT_{ijs} = 1, DTT_{ijs} = 1 \\ PTA\&BIT\&DTT & \text{if } PTA_{ijs} = 1, BIT_{ijs} = 1, DTT_{ijs} = 1 \end{cases} \tag{2}$$

each of which we will refer to as one form (or status) of *treatment*.[4]

We use N^θ and N^0 for the number of treatment-θ and control observations and $N^{\theta 0} = N^\theta + N^0$, and we refer to the sets of observations corresponding

[3]In theory, any other state than complete PEIA nonmembership could serve as a comparison. However, for the sake of simplicity, and given the extent amount of control units in that state, we choose it as a natural reference point in this paper. It should be borne in mind that what we do in the comparison state is to switch off all PEIAs at the same time.

[4]In general, with M treatment types there are 2^M possible combinations. In any case, we always compare $2^M - 1$ states to the all-zero state for the sake of brevity. Clearly, it would be possible to compute – depending on treatment-group size – up to $2^{2M} - 2^M$ average treatment effects of the treated and up to $(2^{2M-1} - 2^{M-1})$ average treatment effects.

to these numbers by $\mathscr{N}^{\theta}$, $\mathscr{N}^{0}$, and $\mathscr{N}^{\theta 0} = \mathscr{N}^{\theta} \bigcup \mathscr{N}^{0}$, respectively.

2.2 Assumptions behind consistent estimates of partial treatment effects of the treated

The goal of selection-on-observables approaches – upon choice of untreated units (here, indicated by super-script 0) being the single reference group – is to estimate the *average treatment effect* of PEIA membership from a comparison of outcome Y^{θ} of the units $\{ijs\} \in \mathscr{N}^{\theta}$ with observable characteristics $H_{ijs}^{\theta 0}$ to outcome Y^{0} of the units $\{ijs\} \in \mathscr{N}^{0}$. In order to not misattribute the average difference in Y^{θ} and Y^{0} to differences in $H^{\theta 0}$, the vector of propensity-scores $P^{\theta 0}$ – which is a compact vector representation of the matrix $H^{\theta 0}$ – is used for weighting in some way, depending on the required similarity between treated and control units in terms of $P^{\theta 0}$ specified by the researcher. With matching, the similarity of treated and control units in terms of $P^{\theta 0}$ is specified by way of k-nearest neighbor matching, radius matching, or kernel matching. The matching-function type determines the nature of the weights based on the propensity scores. However, this approach only leads to consistent estimates of the average treatment effect under the following assumptions.

Balancing of the observables $H_{ijs}^{\theta 0}$ with regard to $\widehat{P}_{ijs}^{\theta 0}$:

The first key condition is the aforementioned balancing of the covariates. Informally, balancing makes sure that the propensity score is a meaningful metric of comparison. Notice that this is the case only if, for units $\{ijs\}$ and $\{i'j's'\}$, $P_{ijs}^{\theta 0}$ and $P_{i'j's'}^{\theta 0}$ means pairwise similarity for all columns in $H_{ijs}^{\theta 0}$ and $H_{i'j's'}^{\theta 0}$, respectively. Otherwise, similarity of $P_{ijs}^{\theta 0}$ and $P_{i'j's'}^{\theta 0}$ would be an artifact, and estimating the average treatment effect from comparison groups of treated and untreated units with similar propensity score will eventually not be consistent.

However, the assumption about balancing of the covariates is testable regarding the first as well as higher moments (see Huber, 2011), and remedies against a lack of balancing are available. One such remedy entropy-balancing weighting regression (see Hainmüller, 2012; Imai and Ratkovic, 2014; Zubizarreta, 2015) covariate balancing can be enforced for several moments.

Unconfoundedness or conditional mean independence of treatment:

The second key condition is (weak) unconfoundedness. It means that, for the same unit $\{ijs\}$ and conditional on the observable determinants of its treatment status, $H_{ijs}^{\theta 0}$, the hypothetical outcomes Y_{ijs}^{θ} and Y_{ijs}^{0} for that unit are independent of the treatment θ. Formally, using $Y_{ijs}^{\theta 0}$ for all units with either treatment θ or control units (i.e., an element of the vector $Y^{\theta 0} = (Y^{\theta\prime}, Y^{0\prime})'$):

$$Y_{ijs}^{\theta 0} \perp T_{ijs} | H_{ijs}^{\theta 0}. \tag{3}$$

The latter means that $H_{ijs}^{\theta 0}$ needs to include all joint determinants of outcome $Y_{ijs}^{\theta 0}$ and treatment T_{ijs}^{θ} (and, hence, $P_{ijs}^{\theta 0}$).

Consistency of the functional form of $\widehat{P}_{ijs}^{\theta 0}$:

An inconsistency of the propensity-score estimates could flow from an erroneous assumption about the functional form of the distribution for the mapping of $H_{ijs}^{\theta 0}$ into $P_{ijs}^{\theta 0}$. Then, maximum-likelihood estimates of the scores $P_{ijs}^{\theta 0}$ would be biased and inconsistent.[5] However, this would not necessarily be a problem for comparison estimators as long as the ranking of units $\{ijs\}$

[5] Leading estimators and functional forms in applied work are probit (normality) and logit. In estimating the propensity of PEIA membership, probit is used in most applications (e.g., see Baier and Bergstrand, 2004, 2009; Egger, Egger, and Greenaway, 2008; Egger and Wamser, 2013). In principal, $P_{ijs}^{\theta 0}$ could be estimated by any parametric or nonparametric consistent estimator.

would be consistent, and the degree of similarity if two comparison units would be only marginally affected (i.e., similarity in $\widehat{P}_{ijs}^{\theta 0}$ would still mean similarity of all columns in $H_{ijs}^{\theta 0}$).

2.3 Treatment-effect estimation through weighting regression

In this subsection, we present two alternative types of weighting regression for a framework of selection on observables, each of which involves a specific first stage to determine the weights and an outcome which corresponds to weighted least squares. In each case, the second stage is run on a subset of the data where either $\Theta_{ijs} = \theta$ or $\Theta_{ijs} = 0$, namely $\mathscr{N}^{\theta 0}$. For convenience, let us also introduce a subvector of the joint determinants of PEIA membership and bilateral trade, H_{ijs}, which we refer to as Z_{ijs}. We introduce this subvector in order to be able to indicate that one may (and we do) condition on some (or even all) of the covariates in H_{ijs} *after* conditioning on the propensity score. For instance, such a procedure is suggested by Blundell and Costa Dias (2009) to reduce the bias from a lack of covariate balancing.

2.3.1 Inverse-probability weighting (IPW) regression

With inverse probability weighting, the first stage of the approach is concerned with estimating the response (or PEIA-membership) probabilities, $P_{ijs}^{\theta 0}$. Since response probabilities are typically estimated parametrically by a maximum-likelihood estimator for nonlinear probability models, one concern may arise with respect to the repeated occurrence of the index tuple ij over s and an associated cluster structure of the variance-covariance matrix of the disturbances.[6] We address this problem by generally estimating $P_{ijs}^{\theta 0}$ year by year (see also Wooldridge, 1995, for a recommendation along those lines) and for each treatment θ separately (ensuring that all propensities including the one for zero treatment add up properly to unity). One prominent example of a first-stage model along those lines is the probit model, which we employ here. The (inverse) propensities obtained in this first stage are the weights used in the second stage of the IPW regression framework. Formally, the (conditional) propensity score is obtained by conditioning on H_{ijs}^{θ} from eq. (??).

In the second stage, we may condition on the covariates, Z_{ijs}, exerting an impact on bilateral exports, Y_{ijs}, or not. If we do, we obtain parameters from two weighting expressions, namely for the treated as

$$\min_{\alpha^\theta,\beta^\theta} \sum_{ijs \in \mathscr{N}^\theta} \frac{(X_{ijs} - \alpha^\theta - Z_{ijs}\beta^\theta)^2}{\widehat{P}_{ijs}^{\theta 0}}, \tag{4}$$

and for the controls as

$$\min_{\alpha^0,\beta^0} \sum_{ijs \in \mathscr{N}^0} \frac{(X_{ijs} - \alpha^0 - Z_{ijs}\beta^0)^2}{1 - \widehat{P}_{ijs}^{\theta 0}}. \tag{5}$$

Using the notation $\overline{Z}^{\theta 0}$ and $\overline{Z}^{\theta}$ to denote row vectors containing the average values of Z_{ijs} in the subsets of the observations in $\mathscr{N}^{\theta 0}$ and $\mathscr{N}^{\theta}$, respectively, the average treatment effect (ATE) and the average treatment effect of the actually treated (ATT) of a type-θ PEIA membership in comparison to no treatment at all with inverse probability weighting are then defined as

$$\widehat{ATE}_{ipwra}^{\theta 0} = (\widehat{\alpha}^\theta - \widehat{\alpha}^0) + \overline{Z}^{\theta 0}(\widehat{\beta}^\theta - \widehat{\beta}^0) \tag{6}$$

[6] Note that the country-pair dimension indexed by ij accounts for the lion's share of the variance in bilateral exports, even in a long panel of data as the one used here.

and

$$\widehat{ATT}^{\theta 0}_{ipwra} = (\widehat{\alpha}^{\theta} - \widehat{\alpha}^{0}) + \overline{Z}^{\theta}(\widehat{\beta}^{\theta} - \widehat{\beta}^{0}), \tag{7}$$

respectively.

One major advantage of this framework is its simplicity. However, a fundamental drawback is that it assumes covariate balancing in the columns of H_{ijs}, at least in higher moments than the first one, which in reality is often rejected by the data. As argued before, a lack of covariate balancing may lead to a bias of the second-stage weighting-regression estimates and, hence of $\widehat{ATE}^{\theta 0}_{ipwra}$ and $\widehat{ATT}^{\theta 0}_{ipwra}$.

2.3.2 Covariate-balance-enforcing (CBE) weighting regression

The second approach to PEIA-treatment-effect estimation by weighting regression differs from the one in the previous subsection only with respect to the first stage. In contrast to a propensity-score model, the weights here are obtained as follows (see Hainmüller, 2012; Hainmüller and Xu, 2013).

Define an ex-ante unknown weight for unit $\{ijs\} \in \mathscr{N}^0$, e_{ijs}, a base weight, q_{ijs}, and a distance metric between the two as

$$f(e_{ijs}) = e_{ijs} log(e_{ijs}/q_{ijs}). \tag{8}$$

Then, the weights e_{ijs} are chosen so as to minimize the loss function

$$\min_{e_{ijs}} F(e) = \sum_{\{ijs \in \mathscr{N}^0\}} f(e_{ijs}) \tag{9}$$

subject to the set of balance constraints

$$\sum_{\{ijs \in \mathscr{N}^0\}} e_{ijs} c_{r,ijs}(H_{ijs}) = m_r^\theta \tag{10}$$

where $c_{r,ijs}(H_{ijs})$ is the moment function for the covariates H_{ijs} among the control observations $\{ijs\} \in \mathscr{N}^0$ up to moment r and the r-th moment of the (base-unweighted) treated observations $\{ijs\} \in \mathscr{N}^\theta$, m_r^θ, and subject to the normalization constraints

$$e_{ijs} \geq 0, \text{ and } \sum_{\{ijs \in \mathscr{N}^0\}} e_{ijs} = 1. \tag{11}$$

Let us denote the solution for e_{ijs} by this procedure by $\widehat{e}_{ijs}^{\theta 0}$. Using this estimate, we may formulate the entropy-balancing counterparts to equations (**??**) and (**??**) for the treated and control observations as

$$\min_{\alpha^\theta, \beta^\theta} \sum_{ijs \in \mathscr{N}^\theta} \frac{(X_{ijs} - \alpha^\theta - Z_{ijs}\beta^\theta)^2}{\widehat{e}_{ijs}^{\theta 0}} \tag{12}$$

and

$$\min_{\alpha^0, \beta^0} \sum_{ijs \in \mathscr{N}^0} \frac{(X_{ijs} - \alpha^0 - Z_{ijs}\beta^0)^2}{\widehat{e}_{ijs}^{\theta 0}} \tag{13}$$

respectively, and the corresponding treatment effects are defined as

$$ATE_{balance}^{\theta 0} = ATT_{balance}^{\theta 0} = \widehat{\alpha}^\theta - \widehat{\alpha}^0. \tag{14}$$

There are two notable and desirable differences between eq. (**??**) and eqs. (**??**) and (**??**), apart from their being based on different parameter estimates. First, any difference in the targeted moments of H_{ijs} (and Z_{ijs})

is eliminated by the minimization in eq. (**??**) subject to the aforementioned constraints. Hence, differences in the these moments of H_{ijs} between the treated and control units do not influence average treatment effects. Second, for the same reason, the average treatment effect (ATE) is identical to the average treatment effect of the treated (ATT) which is not automatically the case with inverse probability weighting.

3 Estimating partial PEIA effects with covariate-balancing-enforcing versus non-enforcing methods

This section is organized in two subsections. The first subsection provides an overview of the panel data we use in the empirical analysis of PEIA effects on trade flows in this paper. In the second subsection we summarize the empirical findings based on covariate-balancing-enforcing versus non-enforcing methods. The latter will present estimates of both partial (or direct) effects on trade flows which do not account for adjustments of prices in general equilibrium and total (or general-equilibrium) effects on welfare which do account of such adjustments.

3.1 Data

We cover annual data of bilateral trade flows and their determinants over the time interval of 1961-2008. The trade data are collected from the United Nations' (UN) Comtrade Database, and we restrict our interest to positive trade flows.[7] The main regressor of interest to this paper is a country pair's membership status in one of the preferential economic integration agreements – PTA, BIT, DTT, or combinations thereof – which we may refer to as the treatment indicator, referred to as $T_{ijs}^{\theta 0}$ above. Information on PTA membership is collected from World Trade Organization's (WTO) website.[8] The variable PTA_{ijs} is unity, if trade between countries i and j is covered by a PTA in year s. Information on DTT and BIT membership is collected from the website of the United Nations Conference on Trade and Development (UNCTAD).[9] BIT_{ijs} and DTT_{ijs} are unity if investment between countries i and j is covered by a BIT or a DTT, respectively, in year s.

We use other covariates based on variables contained in the World Bank's World Development Indicators (WDI) and the Centre d'Études Prospectives et d'Informations Internationales' (CEPII) gravity data-set. We summarize the acronyms, provide a brief description, and report the source of all variables except the fixed country-time effects in Table . We suppress subscripts in the table but would like to not that the outcome, the treatment indicators, and the 1st-stage-only covariates all vary in the three dimensions i, j, and s, whereas the geographical, cultural, and historical variables included in both stages vary only in dimensions i and j but not s.

Among the 1st-stage only covariates, we have four regressors in the spirit of Baier and Bergstrand (2004) and Egger and Larch (2008). These measure total economic size between countries i and j in year s in terms of their log total real GDP ($RGDP_{ijs}$), the dissimilarity in economic size between

[7] Notice that the vector of observables, H_{ijt}, includes exporter-time and importer-time fixed effects, and so does the vector Z_{ijs}. Conditional on these fixed effects, it turns out that there is no significant bias associated with sample selection (see Wooldridge, 1995, for the consistency of fixed-effects estimates in the case of specific forms of sample selection.)

[8] see XXX

[9] see XXX

countries i and j in year s in terms of the absolute difference of their real GDP ($DRGDP_{ijs}$), the difference in capital-labor ratios of countries i and j in year s approximated by the absolute difference in log real per-capita income (DKL_{ijs}), and the difference in capital-labor ratios of countries i and j together in year s and the rest of the world (in our case, 165 countries) approximated by the absolute difference in log real per-capita income ($DROWKL_{ijs}$). The results in Baier and Bergstrand (2004) suggest that such measures successfully co-determine the propensity of signing at least a PTA, and the findings in Egger and Wamser (2013) suggest that similar conclusions apply for BITs and DTTs. Such regressors are historically known to determine the volume of trade and, in particular, of intra-industry trade (see Helpman, 1987). Hence, the result in Baier and Bergstrand (2004) and others suggest that PEIAs are concluded primarily among *natural trading partners*, i.e., countries which would trade and direct invest a lot with each other in the absence of political barriers.

The other covariates, which are included in the first- as well as the second-stage models are also standard in empirical research, and they capture geographical, cultural, historical, and political factors determining bilateral trade as well as PEIA membership. All of those covariates are time-invariant. The geographical factors include log bilateral distance ($DIST_{ij}$) and common land border ($BORDER_{ij}$). The cultural variables include common official language ($LANG1_{ij}$) and common ethnological language when spoken by a sufficiently large base ($LANG1_{ij}$). The historical variables are four indicators which capture colonial relationships of some form ($COLONY1_{ij}$, $COLONY2_{ij}$, $COLONY3_{ij}$, $COLONY4_{ij}$). Finally, we measure some special political relationship between entities i and j, if they did or currently do belong to the same country ($MCTRY_{ij}$). Again, for of these sets of variables it is documented that they successfully co-explain bilateral trade flows as well as PEIA membership.[10]

Table **??** provides desctriptive statistics for all variables involved, including the dependent variable, log bilateral exports (Y_{ijs}). We suppress a detailed discussion of those statistics but would like to single out the following observations. First, the overall data-set which is used for estimation contains 434,895 observations (made up of 158 exporters, 160 importers, and 50 years). Second, in the data the most frequent PEIAs are having a PTA or a DTT alone (PTA, DTT). This state prevails for 10% of the observations each. The least frequent states are having a PTA combined with a BIT or a DTT ($PTA\&BIT$, $PTA\&DTT$) which prevails for about 2% of the observations each. In any case, Table **??** reports on the absolute number of the treatment states behind the percentages in Table **??** and suggests that all states occur frequently enough so that we should be able to estimate treatment effects – if there are any – with sufficient precision given the large number of observations.

$--$ Table **??** about here $--$

$--$ Table **??** about here $--$

$--$ Table **??** about here $--$

While Tables **??** and **??** reflect the base case of treatment configurations which we consider in this paper. However, recent research by Dür and Elsnig (2014) and XXXX suggests that it could be useful to distinguish among PTAs, as these PTAs contain a host of different provisions. The customary approach to approach this issue in empirical and quantitative work appears

[10]EXPLAIN WHY NO 3RD COUNTRY EFFECTS INCLUDED.

to be to distinguish among important types of PTAs. In this paper, we distinguish between four of them along the following lines: ($PTA1$); ($PTA2$); ($PTA3$); and ($PTA4$). Of course, doing so leads to 19 treatment states (except for the null state without any PEIA) relative to the 7 states (except for the null state without any PEIA) in Tables **??** and **??**. For this more fine-grained definition of PEIAs we generate Tables **??** and **??** as counterparts to Tables **??** and **??**. These tables suggest that PTA-state $PTAX$ is most frequent (when computing the sum of all states involving $PTAX$) and PTA-state $PTAX$ is least frequent (when computing the sum of all states involving $PTAX$). We will use both the coarse and the fine-grained set of treatment definitions in the subsequent analysis.

$--$ Table **??** about here $--$

$--$ Table **??** about here $--$

XXX PUT TABLES XXX

3.2 Results and discussion

Notice that with 50 years of data and annual probit estimates, we would have to report an enormous number of parameter estimates for the first-stage models. We suppress those in the main text but relegate the respective presentation to the accompanying online appendix. In any case, those parameter estimates are of limited interest for two reasons: first, the probit models – as first-stage models in general – do not have a structural interpretation and, second, only the signs of the parameter estimates are interpretable due to the nonlinear model structure. In what follows, we will focus on partial and general-equilibrium PEIA-effect estimates and, as a prerequisite, on balancing tests.

3.2.1 Covariate balancing tests

Before turning to effect estimates, let us focus on covariate balancing. Notice that with 13 main covariates, 50 years of data and 7 or 19 treatment states there are $13 \cdot 50 \cdot 7 = 4,550$ and $13 \cdot 50 \cdot 19 = 12,350$ tests for, say, the first and the second moment. It would not be convenient to present the associated results by way of tables, but we present figures of kernel density plots of p-values of mean and variance comparison tests to assess the differences of the first and second moments of the covariates between the treated and the controls. We generally report three kernel density plots on p-values for all covariates together per moment: one for the simple OLS model, one for the customary IPW model, and one for the EB model each.

The density plots suggest that the mode of the distributions for both the first and the second moment is obviously highest for the OLS model, and it is also obviously much higher for the IPW model than for the EB model. In fact, the mode of the p-value for the equality-of-means tests between the treated and the control units amounts to 0.XXX for the EB model, while it is 0.XXX for the IPW model and 0.XXX for OLS. The mode of the p-value for the equality-of-variances tests between the treated and the control units amounts to 0.XXX for the EB model, and it is 0.XXX for the IPW model and 0.XXX for OLS. Hence, the first two moments differ starkly between the treated and the controls before any weighting and even when weighting by the propensity score. Hence, we have a situation where one fundamental assumption for not only IPW regression but also propensity-score matching to obtain consistent partial effects is starkly violated.

XXX CONTINUE AFTER HAVING RESUITS. EV. ALSO INCLZUDE PLOT FOR 3RD MOMENT XXX

3.2.2 Partial effects of PEIAs on bilateral exports

The tables and text in this sub-sub-section summarize partial effects of PEIA treatments on bilateral exports estimated with the coarse-grained (7 treatment states; Table **??**) and the fine-grained (19 treatment states; Table **??**) differentiation of PEIAs.

Tables **??** and **??** follow the same principal organization. Apart from the column referring to the treatment at stake, the columns labelled OLS, IPW, and CBE pertain to average treatment effects associated with the respective estimators. The numbers in italics reported below the ATE parameters are country-pair-block-bootstrapped standard errors (see Fitzenberger, 199X). The columns Obs. and Treated refer to numbers of all and respective treated observations which estimates are based on, and Treated % expresses the Treated in percent of Obs. Clearly, the number of control observations with zero PEIA treatment is always Obs. -Treated=311,974. Since the only difference between Tables **??** and **??** is that PTA treatment in Table **??** is split up in four categories in Table **??**, only those ATE estimates differ between the tables, where any PTA treatment is involved.

It is apparent from the comparison of columns OLS and CBE on the one hand and of columns IPW and CBE on the other hand – especially, when recalling the substantial lack of covariate balancedness behind propensity scores – that inverse probability weighting is quite problematic in the data. The ranking of the magnitude of the partial effects is quite similar between CBE and OLS on average in both tables across the ATEs estimated, while it is quite different with IPW. The CBE estimates tend to be non-trivially smaller than the OLS estimates in absolute value, while no clear-cut pattern emerges for IPW relative to OLS. Of all the ATEs four of the signs in Tables **??** and **??** differ between IPW and OLS, only one of the signs differs between

CBE and OLS, and five of the signs differ between IPW and CBE.

Let us consider the magnitude of the differences between the partial ATE estimates across all lines in Tables **??** and **??** together. Let us compute the simple (unweighted) average of the absolute differences between the ATEs under OLS versus IPW (which is 1.5249 for the two tables and all ATEs together), OLS versus CBE (which is 0.3504 for the two tables and all ATEs together), and IPW versus CBE (which is 1.5542 for the two tables and all ATEs together). These numbers clearly indicate that the bias induced by lack of balancing under IPW is actually larger than the one of OLS, and the applied economist would have done better to ignore any self-selection into PEIAs rather than doing IPW or matching.

The biggest biases of OLS relative to CBE in absolute terms emerge for the combination of PTA with BIT and DTT (PTA&BIT&DTT), amounting to 0.6953 in Table **??** and the deepest PTA form with BIT and DTT (PTA4&BIT&DTT), amounting to 0.9197 in Table **??**. The biases of IPW versus CBE for these treatments in the two tables amount to 1.3354 and 8.0764, respectively, and they are not even the biggest ones of the IPW estimator across different treatments. Most of the biases of OLS relative to CBE amount to substantially less than 0.5.

−− Table **??** about here −−

−− Table **??** about here −−

3.2.3 Welfare effects of PEIAs

While empirical economists tended to stop at reporting partial effects of PEIAs less than a decade ago, it is now customary to quantify such effects when taking into account general-equilibrium repercussions (see, e.g., Egger and Larch, 2011?; Egger, Larch, Staub, and Winkelmann, 2013; Caliendo and Parro, 2015).

−− Table **??** about here −−

−− Table **??** about here −−

4 Conclusions

- New covariate balance enforcing technique for non-parametric estimation of trade determinants.

- The new technique generates identical to the first moment and very similar to the second and third moment output variables conditional on covariates.

- Failure to reach covariate balance generates on average an upward bias in estimated coefficients of PEIAs.

- Despite a significant difference in magnitude of the second-stage coefficients, probit estimation is far from satisfying CIA.

- Any combination of two or three PEIAs has a stronger effect on exports than any individual PEIA has.

- Our results are robust to accounting for different levels of depth that a PTA between two countries has. In general, one PTA with at least one provision generates more trade than a relatively shallow PTA, but also than a PTA foreseeing that all tariffs should be reduced to zero, even when that includes provisions.

- In the general equilibrium (GE) context, we find that PEIA-related trade costs as estimated using the newly-introduced covariate-balance enforcing estimator lead to smaller welfare gains than those estimated with the least-squares method.

5 Table Appendix

Table 1: Determinants of PEIAs and trade (description and source)

Variables	Description	Source
	Outcome variable	
Y	Log bilateral exports	UN
	Binary treatment variables	
PTA	Preferential trade agreement only	WTO
BIT	Bilateral investment treaty only	UNCTAD
DTT	Double-taxation treaty only	UNCTAD
PTA&BIT	Combination of the above	
PTA&DTT	Combination of the above	
BIT&DTT	Combination of the above	
PTA&BIT&DTT	Combination of the above	
	1st-stage-only covariates	
RGDP	The sum of two countries' log real GDPs	WDI
DRGDP	The absolute difference of two countries' log real GDPs	WDI
DKL	The absolute difference in the two countries' log real per-capita incomes	WDI
DROWKL	The average absolute difference in log per-capita incomes of two countries with the rest of the world	WDI
	1st- and 2nd-stage covariates	
DIST	The log distance between two countries' economic centers	CEPII
BORDER	Binary common country border	CEPII
LANG1	Binary for common official primary language	CEPII
LANG2	Binary for common language if spoken by at least 9% of the population	CEPII
COLONY1	Binary for colonial relationship (ever)	CEPII
COLONY2	Binary common colonizer post 1945	CEPII
COLONY3	Binary for pair currently in colonial relationship	CEPII
COLONY4	Binary for pair in colonial relationship post 1945	CEPII
SMCTRY	Binary for entities that were or are part of the same country	CEPII

Table 2: Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Outcome variable					
Y (log exports)	434,895	1.35	3.36	-6.91	12.78
1st-stage covariates only					
RGDP	434,895	48.96	2.80	37.36	59.45
DRGDP	434,895	2.83	2.02	0	10.97
DKL	434,895	0.03	0.07	0	0.64
DROWKL	434,895	1.42	0.62	0	3.85
1st- and 2nd-stage covariates					
DIST	434,895	8.57	0.87	4.09	9.89
BORDER	434,895	0.03	0.17	0	1
LANG1	434,895	0.18	0.39	0	1
LANG2	434,895	0.19	0.39	0	1
COLONY1	434,895	0.03	0.16	0	1
COLONY2	434,895	0.09	0.29	0	1
COLONY3	434,895	0.00	0.02	0	1
COLONY4	434,895	0.02	0.13	0	1
SMCTRY	434,895	0.02	0.13	0	1

One-year-lagged complemetary treatments are included as regressors in the first stage, weighted by distance in the following categories: [0km, 500km], (500km, 1000km], (1000km, 2000km], (2000km, 5000km], (5000km, ∞ km].

Table 3: Presence of PEIAs: Coarse-grained PEIA Definition

PEIA Types	Observations	Percentage
Null-state	311,974	71.74
PTA	34,870	8.02
BIT	17,734	4.08
DTT	32,756	7.53
PTA&BIT	6,225	1.4
PTA&DTT	5,480	1.26
BIT&DTT	16,822	3.87
PTA&BIT&DTT	9,034	2.08
Total	434,895	100.00

Table 4: PTA Depth

Depth	Total	%
1. Shallow (PTA1)	20,707	60.75
2. At least one provision (PTA2)	5283	15.50
3. Full FTA (PTA3)	4444	13.04
4. Full FTA + at least one provision (PTA4)	3649	10.71
	34,083[a]	

[a]This should add up to 34,870 as per the number of observations with a PTA in Table ??, however, the DESTA database covers slightly less PTAs then those recorded by WTO

Table 5: Presence of PEIAs: Fine-grained PEIA Definition

PEIA Types	Observations	Percentage
Null-state	311974	71.88
PTA1	20707	4.77
PTA2	5283	1.22
PTA3	4444	1.02
PTA4	3649	0.84
BIT	17734	4.09
DTT	32756	7.55
PTA1&BIT	1932	0.45
PTA2&BIT	603	0.14
PTA3&BIT	1213	0.28
PTA4&BIT	2456	0.57
PTA1&DTT	1954	0.45
PTA2&DTT	1248	0.29
PTA3&DTT	770	0.18
PTA4&DTT	1464	0.34
BIT&DTT	16822	3.88
PTA1&BIT&DTT	2291	0.53
PTA2&BIT&DTT	982	0.23
PTA3&BIT&DTT	894	0.21
PTA4&BIT&DTT	4842	1.12
Total	434018[a]	100

[a]This should add up to 434,895 as per the number of observations in Table ??, however, the DESTA database covers slightly less PTAs then those recorded by WTO

Table 6: ATE of PEIAs on Log Exports: Coarse-grained PEIA Definition

	OLS	**IPW**	**CBE**	**Obs.**	**Treated**	**Treated %**
PTA	0.8046	1.2617	0.5256	346,844	34,870	5.21
SE	*0.0008*	*0.0576*	*0.0009*			
BIT	0.6776	1.0251	0.4720	329,708	17,734	2.65
SE	*0.0007*	*0.1918*	*0.0008*			
DTT	0.7758	0.1878	0.6374	344,730	32,756	4.89
SE	*0.0006*	*0.0963*	*0.0006*			
PTA&BIT	1.5943	-0.8580	1.1565	318,199	6,225	0.92
SE	*0.0012*	*0.0826*	*0.0017*			
PTA &DTT	1.6625	2.9734	1.3157	317,454	5,480	0.82
SE	*0.0011*	*0.1159*	*0.0012*			
BIT&DTT	1.2782	-1.9000	0.9659	328,796	16,822	2.51
SE	*0.0008*	*0.3452*	*0.0009*			
PTA&BIT&DTT	1.9674	-0.0633	1.2721	321,008	9,034	1.35
SE	*0.0009*	*0.1587*	*0.0017*			

OLS: coefficient estimates of one stage least-squares regression of dependent variable *log(EXPORTS)* on covariates shown in Table **??**. IPW: second-stage coefficient estimates of inverse probit-based weights-weighted regression. CBE: second-stage coefficient estimates of covariate-balance-enforcing weights-weighted regression. Exporter*year fixed effects and importer*year fixed effects included in both stages. Standard errors are obtained through bootstrapping, by drawing 100 times with replacement.

Table 7: ATE of PEIAs on Log Exports: Fine-grained PEIA Definition

	OLS	**IPW**	**CBE**	**N**	**Treated**	**Treated %**
PTA1	0.5279	0.1882	0.3295	318,729	6757	2.12
SE	*0.0313*	*0.1295*	*0.0124*			
PTA2	1.0713	0.7526	0.7643	312,533	563	0.18
SE	*0.0116*	*0.4322*	*0.0224*			
PTA3	0.7318	1.8870	0.0156	313,744	1757	0.56
SE	*0.0570*	*0.3458*	*0.0197*			
PTA4	0.9346	2.3113	0.2794	337,760	25771	7.63
SE	*0.0190*	*0.2377*	*0.0126*			
BIT	0.6776	1.0251	0.4720	329,708	17738	5.38
SE	*0.0025*	*0.4004*	*0.0026*			
DTT	0.7758	0.1878	0.6374	344,730	32749	9.50
SE	*0.0013*	*0.4169*	*0.0018*			
PTA1&BIT	1.4508	2.1255	1.3789	312,434	469	0.15
SE	*0.0685*	*0.3224*	*0.0281*			
PTA2&BIT	1.8255	1.2578	1.5240	312,174	187	0.06
SE	*0.0746*	*0.5943*	*0.0420*			
PTA3&BIT	1.2698	4.2840	0.6233	312,140	156	0.05
SE	*0.0413*	*0.6026*	*0.0297*			
PTA4&BIT	1.3934	1.0592	0.7687	317,373	5395	1.70
SE	*0.0060*	*0.4932*	*0.0073*			
PTA1&DTT	1.6660	1.5496	1.3065	312,514	531	0.17
SE	*0.0506*	*0.3095*	*0.0444*			
PTA2&DTT	1.8550	6.0443	1.6676	312,134	156	0.05
SE	*0.0116*	*0.3650*	*0.0200*			
PTA3&DTT	1.0806	0.1346	0.9561	312,254	281	0.09
SE	*0.0090*	*0.3535*	*0.0084*			
PTA4&DTT	1.2118	3.1360	-5.1567	316,474	4494	1.42
SE	*0.0250*	*0.3855*	*0.6884*			
BIT&DTT	1.2782	-1.9000	0.9659	328,796	16834	5.12
SE	*0.0025*	*0.9989*	*0.0030*			
PTA1&BIT&DTT	1.9219	0.8238	1.5053	312,358	375	0.12
SE	*0.1121*	*0.7419*	*0.0979*			
PTA2&BIT&DTT	2.1628	0.1289	1.6443	312,157	187	0.06
SE	*0.0179*	*0.3473*	*0.0303*			
PTA3&BIT&DTT	1.9800	1.9170	1.4761	312,397	437	0.14
SE	*0.0486*	*0.4338*	*0.0407*			
PTA4&BIT&DTT	1.8342	1.4772	1.0828	320,018	8032	2.51
SE	*0.0072*	*2.3977*	*0.0121*			

OLS: coefficient estimates of one stage least-squares regression of dependent variable *log(EXPORTS)* on covariates shown in Table **??**. IPW: second-stage coefficient estimates of inverse probit-based weights-weighted regression. CBE: second-stage coefficient estimates of covariate-balance-enforcing weights-weighted regression. Exporter*year fixed effects and importer*year fixed effects included in both first and second stage. Standard errors are obtained through bootstrapping, by drawing 10 times with replacement. PTA1 is a dummy indicating a country pair with PTA depth level of 1, PTA2 - two, PTA3 - three, and PTA4 - four.

Table 8: Welfare Effects From All Country Pairs Having a PEIA Relative to Status Quo (Coarse PEIA Definition)

	% **Welfare Change**		
	OLS	IPW	CBE
PTA	5.45	21.51	2.18
BIT	2.09	6.89	1.13
DTT	7.60	0.98	4.88
PTA&BIT	44.20	0.10	17.38
PTA&DTT	44.02	386.62	20.75
BIT&DTT	25.87	0.19	11.34
PTA&BIT&DTT	142.69	0.26	44.26

Table 9: Welfare Effects From All Country Pairs Having a PEIA Relative to Status Quo (Fine-grained PEIA Definition)

	% **Welfare Change**		
PEIAs	OLS	IPW	CBE
PTA1	1.0141	0.6495	0.7520
PTA2	2.4036	1.3477	0.8156
PTA3	1.3477	16.4702	0.5770
PTA4	3.8031	75.0455	0.8156
BIT	2.1048	6.8221	1.1605
DTT	7.0127	0.9861	4.5293
PTA1&BIT	4.0050	18.7968	3.3563
PTA2&BIT	4.4642	1.4470	2.4046
PTA3&BIT	5.5533	649.5653	1.2667
PTA4&BIT	17.1303	8.7078	4.3552
PTA1&DTT	14.8228	11.4417	6.4192
PTA2&DTT	11.3812	1439.7520	8.2098
PTA3&DTT	1.8160	0.5937	1.4115
PTA4&DTT	5.7759	133.2927	0.4273
BIT&DTT	24.6472	0.2712	10.7307
PTA1&BIT&DTT	26.3835	2.6418	11.8842
PTA2&BIT&DTT	33.7238	0.6542	13.7982
PTA3&BIT&DTT	22.7528	20.4637	9.1668
PTA4&BIT&DTT	62.2179	34.6331	16.2719

Table 10: P-Values of Balancing Tests

		7 Treatments		**19 Treatments**	
		Median	Mean	Median	Mean
Means' test comparison	UNW	0.0000	0.0186	0.0000	0.0875
	IPW	0.0000	0.0456	0.0002	0.1297
	CBE	0.9968	0.9870	0.9977	0.9844
Variance test comparison	UNW	0.0000	0.0907	0.0000	0.0819
	IPW	0.0013	0.1405	0.0021	0.1458
	CBE	0.2664	0.2664	0.0923	0.2529

Table 11: Propensity of Coarse-definition PEIAs

		1961		**1970**		**1980**		**1990**		**2000**		**2010**	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
BIT	*Control*	.	.	0.0036	0.0114	0.0076	0.0206	0.0151	0.0366	0.0377	0.0709	0.0465	0.0803
	Treated	.	.	0.0226	0.0163	0.0506	0.0380	0.1029	0.0674	0.2076	0.1583	0.2385	0.1643
DTT	*Control*	0.0087	0.0314	0.0196	0.0528	0.0294	0.0653	0.0359	0.0744	0.0454	0.0839	0.0515	0.0873
	Treated	0.0697	0.0980	0.1556	0.1610	0.1915	0.1667	0.2197	0.1749	0.2515	0.1912	0.2456	0.1934
PTA	*Control*	0.0031	0.0162	0.0139	0.0398	0.0237	0.0530	0.0392	0.0659	0.0815	0.0930	0.0860	0.0938
	Treated	0.0655	0.1027	0.1589	0.1945	0.1449	0.1576	0.1637	0.1826	0.2566	0.2075	0.2677	0.2174
BIT&DTT	*Control*	.	.	0.0013	0.0093	0.0043	0.0225	0.0106	0.0432	0.0294	0.0804	0.0442	0.1006
	Treated	.	.	0.0123	0.0129	0.0618	0.0926	0.1803	0.1854	0.3647	0.2472	0.4095	0.2500
PTA&BIT	*Control*	.	.	.	.	0.0011	0.0082	0.0040	0.0213	0.0218	0.0535	0.0269	0.0593
	Treated	.	.	.	.	0.0003	0.0004	0.2345	0.2527	0.2430	0.2289	0.2301	0.2279
PTA&DTT	*Control*	0.0003	0.0060	0.0016	0.0166	0.0037	0.0214	0.0054	0.0279	0.0079	0.0365	0.0179	0.0568
	Treated	0.0704	0.0986	0.2332	0.2469	0.3002	0.2786	0.1279	0.1390	0.1686	0.1710	0.2344	0.2061
BIT&DTT	*Control*	.	.	.	.	0.0104	0.0433	0.0094	0.0398	0.0064	0.0309	0.0073	0.0331
	Treated	.	.	.	.	0.0018	0.0000	0.4987	0.2850	0.3226	0.2865	0.2302	0.2347

Table 12: Propensity of Fine-grained-definition PEIAs

		1961		1970		1980		1990		2000		2010	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
PTA1	*Control*	0.0121	0.0159	0.0155	0.0185	0.0173	0.0198	0.0175	0.0175	0.0176	0.0190	0.0202	0.0214
	Treated	0.0622	0.0364	0.0697	0.0417	0.0543	0.0404	0.0362	0.0362	0.0314	0.0291	0.0358	0.0325
PTA2	*Control*	0.0037	0.0162	0.0039	0.0167	0.0037	0.0167	0.0036	0.0036	0.0029	0.0154	0.0028	0.0158
	Treated	0.0869	0.0767	0.0738	0.0684	0.0617	0.0729	0.0731	0.0731	0.0457	0.0535	0.0477	0.0488
PTA3	*Control*	0.0110	0.0222	0.0139	0.0262	0.0160	0.0290	0.0176	0.0176	0.0207	0.0373	0.0221	0.0394
	Treated	.	.	0.0498	0.0306	0.0732	0.0618	0.0672	0.0672	0.0862	0.0733	0.0925	0.0788
PTA4	*Control*	0.0124	0.0238	0.0143	0.0263	0.0145	0.0261	0.0150	0.0150	0.0150	0.0250	0.0143	0.0259
	Treated	0.0926	0.0755	0.0428	0.0562	0.0694	0.0780	0.0595	0.0595	0.0690	0.0667	0.0525	0.0659
BIT	*Control*	.	.	0.0144	0.0330	0.0186	0.0384	0.0198	0.0198	0.0199	0.0403	0.0238	0.0451
	Treated	.	.	0.0668	0.0396	0.0867	0.0481	0.1031	0.1031	0.0981	0.0876	0.1136	0.0935
DTT	*Control*	0.0174	0.0444	0.0278	0.0592	0.0346	0.0660	0.0361	0.0361	0.0358	0.0648	0.0423	0.0712
	Treated	0.0999	0.1326	0.1703	0.1703	0.1938	0.1658	0.2019	0.2019	0.1949	0.1554	0.2037	0.1657
BIT&DTT	*Control*	.	.	0.0142	0.0463	0.0180	0.0525	0.0181	0.0181	0.0150	0.0445	0.0179	0.0490
	Treated	.	.	0.2032	0.1563	0.1414	0.1368	0.1767	0.1767	0.1636	0.1391	0.1656	0.1404
PTA1&BIT	*Control*	.	.	.	.	0.0009	0.0027	0.0009	0.0009	0.0007	0.0022	0.0010	0.0026
	Treated	.	.	.	.	0.0029	0.0027	0.0049	0.0049	0.0068	0.0112	0.0090	0.0128
PTA2&BIT	*Control*	.	.	.	.	0.0003	0.0016	0.0003	0.0003	0.0002	0.0011	0.0002	0.0012
	Treated	.	.	.	.	0.0002	.	0.0028	0.0028	0.0045	0.0054	0.0075	0.0065
PTA3&BIT	*Control*	.	.	.	.	0.0015	0.0049	0.0019	0.0019	0.0023	0.0083	0.0029	0.0096
	Treated	.	.	.	.	0.0214	0.0161	0.0385	0.0385	0.0351	0.0362	0.0364	0.0384
PTA4&BIT	*Control*	.	.	.	.	0.0039	0.0132	0.0038	0.0038	0.0033	0.0098	0.0034	0.0106
	Treated	.	.	.	.	0.0169	0.0158	0.1733	0.1733	0.0715	0.1037	0.0628	0.0936
PTA1&DTT	*Control*	0.0007	0.0060	0.0011	0.0080	0.0014	0.0068	0.0014	0.0014	0.0011	0.0057	0.0015	0.0073
	Treated	0.0213	.	0.0738	0.0927	0.0639	0.0954	0.0200	0.0200	0.0201	0.0245	0.0192	0.0266
PTA2&DTT	*Control*	0.0005	0.0055	0.0008	0.0070	0.0008	0.0046	0.0008	0.0008	0.0005	0.0048	0.0006	0.0043
	Treated	.	.	0.0172	.	0.0181	0.0193	0.0295	0.0295	0.0229	0.0309	0.0132	0.0177
PTA3&DTT	*Control*	0.0010	0.0083	0.0015	0.0109	0.0016	0.0105	0.0018	0.0018	0.0020	0.0108	0.0024	0.0108
	Treated	.	.	.	.	0.1051	0.1122	0.1186	0.1186	0.1132	0.1539	0.0977	0.1565
PTA4&DTT	*Control*	0.0014	0.0142	0.0024	0.0181	0.0028	0.0146	0.0026	0.0026	0.0019	0.0128	0.0022	0.0124
	Treated	0.1932	0.1458	0.1943	0.2040	0.2558	0.2127	0.1728	0.1728	0.0718	0.0898	0.0632	0.1016
PTA1&BIT&DTT	*Control*	.	.	.	.	0.0008	0.0035	0.0008	0.0008	0.0007	0.0040	0.0008	0.0039
	Treated	.	.	.	.	0.0183	0.0208	0.0333	0.0333	0.0208	0.0384	0.0228	0.0392
PTA2&BIT&DTT	*Control*	.	.	.	.	0.0008	0.0044	0.0007	0.0007	0.0004	0.0036	0.0004	0.0026
	Treated	.	.	.	.	0.0594	0.0815	0.0285	0.0285	0.0188	0.0330	0.0202	0.0355
PTA3&BIT&DTT	*Control*	.	.	.	.	0.0008	0.0052	0.0009	0.0009	0.0009	0.0066	0.0013	0.0063
	Treated	.	.	.	.	0.0026	0.0012	0.0599	0.0599	0.0514	0.0667	0.0493	0.0772
PTA4&BIT&DTT	*Control*	.	.	.	.	0.0052	0.0291	0.0045	0.0045	0.0027	0.0194	0.0029	0.0192
	Treated	.	.	.	.	0.2570	.	0.4764	0.4764	0.3229	0.2511	0.2159	0.2363

6 References

Aitken, N. D., 1973. The effect of EEC and EFTA on European trade: A temporal cross-section analysis. American Economic Review 63, 881–892.

Baltagi, B.H., Egger, P.H., Pfaffermayr, M., 2008. Estimating regional trade agreement effects on FDI in an interdependent world. Journal of Econometrics 145.1, 194–208.

Baghdadi, L., Martinez-Zarzoso, I. and Zitouna, H., 2013. Are RTA agreements with environmental provisions reducing emissions?. Journal of International Economics 90.2, 378–390.

Baier, S. L., Bergstrand, J. H., 2004. Economic determinants of free trade agreements. Journal of International Economics 64.1, 29–63.

Baier, S. L., Bergstrand, J. H., 2007. Do free trade agreements actually increase members' international trade?. Journal of International Economics 71.1, 72–95.

Baier, S. L., Bergstrand, J. H., 2009. Estimating the effects of free trade agreements on international trade flows using matching econometrics. Journal of International Economics 77.1, 63–76.

Blundell, R., Costa Dias, M., 2009. Alternative approaches to evaluation in empirical microeconomics. Journal of Human Resources 44(3), 565-640.

Céline Carrére, Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model, European Economic Review 50.2, 223–247.

Chang, P.L., Lee, M.J., 2011. The WTO trade effect. Journal of International Economics 85.1, 53–71.

Egger, H., Egger, P.H., Greenaway, D., 2008. The trade structure effects of endogenous regional trade agreements. Journal of international Economics, 74.2, 278–298.

Egger, P.H., Larch, M., Pfaffermayr, M., Winner, H. (2006). The Impact of Endogenous Tax Treaties on Foreign Direct Investment: Theory and Evidence. Canadian Journal of Economics 39.3, 901–931.

Egger, P.H., Larch, M., Staub, K.E., Winkelmann R., 2011. The trade effects of endogenous preferential trade agreements. American Economic Journal: Economic Policy 3(XX), 113–143.

Egger, P.H. and Merlo, V. (2007), The Impact of Bilateral Investment Treaties on FDI Dynamics. World Economy, 30, 1536–1549.

Egger, P.H., Pfaffermayr, M., 2004. The Impact of Bilateral Investment Treaties on Foreign Direct Investment. Journal of Comparative Economics, 32.4, 788–804.

Egger, P.H., Wamser, G., 2013. Effects of the endogenous scope of preferentialism on international goods trade. The BE Journal of Economic Analysis & Policy 13.2, 709–731.

Frankel, J.A., Stein, E., Wei, S., 1996. Regional Trading Arrangement: Natural or Super-Natural? American Economic Review, 86.2

Fugazza, M., Nicita, A., 2013. The direct and relative effects of preferential market access. Journal of International Economics 89.2, 357–368.

Ghosh S., Yamarik S., 2004. Are Regional Trading Arrangements Trade Creating?: An Aplication of Extreme Bounds Analysis. Journal of International Economics 63.2, 369–395.

Hainmüller, J., 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political Analysis 20(1), 25-46.

Hainmüller, J., Xu, Y., 2013. *ebalance*: a Stata package for entropy balancing Journal of Statistical Software 54(7), XXX-XXX.

Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. Journal of International Economics 81.1, 1–14.

Helpman, E., 1987. Imperfect competition and international trade: Evidence from fourteen industrial countries. Journal of the Japanese and International Economies 1(1), 62-81.

Huber, M., 2011. Testing for covariate balance using nonparametric quantile regression and resampling methods. Journal of Applied Statistics 38(12), 2881-2899.

Huber, M. 2014. Identifying causal mechanisms (primarily) based on inverse probability weighting. Journal of Applied Econometrics 29(6), 920-943.

Hur, J., Park, C., 2012. Do free trade agreements increase economic growth of the member countries?. World development 40.7, 1283–1294.

Imai, K. and Ratkovic, M., 2014. Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76.1, 243–263.

Rose, A., 2005. Which international institutions promote international trade?. Review of International Economics 13.4, 682–698.

Rose, A. K., Van Wincoop, E., 2001. National money as a barrier to international trade: The real case for currency union. The American Economic Review 91.2, 386–390.

Roy, J., 2010. Do Customs Union Members Engage in More Bilateral Trade than Free-Trade Aggreement Members? Review of International Economics 18.4, 663–681.

Soloaga, I., Wintersb, L.A., 2001. Regionalism in the nineties: What effect on trade?. The North American Journal of Economics and Finance 12.1, 1–29.

Wooldridge, J.M., 1995. Selection corrections for panel data models under conditional mean independence assumptions, Journal of Econometrics, 68.1, 115–132.

Wooldridge, J. M., 2007. Inverse probability weighted estimation for general missing data problems. Journal of Econometrics, 141.2, 1281–1301.

Zubizarreta, J., 2015. Stable Weights that Balance Covariates for Estimation with Incomplete Outcome Data. Journal of the American Statistical Association 110.511,: 910–922.