



Estimating the effects of free trade agreements on international trade flows using matching econometrics

Scott L. Baier^{a,b}, Jeffrey H. Bergstrand^{c,d,*}

^a The John E. Walter Department of Economics, Clemson University, Clemson, SC 29634 USA

^b Federal Reserve Bank of Atlanta, Atlanta, GA 30309 USA

^c Department of Finance, Mendoza College of Business, and Kellogg Institute for International Studies, University of Notre Dame, Notre Dame, IN 46556 USA

^d CESifo, Munich, Germany

ARTICLE INFO

Article history:

Received 5 December 2006

Received in revised form 16 July 2008

Accepted 9 September 2008

Keywords:

International trade flows

Free trade agreements

Matching econometrics

Treatment effects

JEL classification:

F10

F13

F15

ABSTRACT

This paper provides the first cross-section estimates of long-run treatment effects of free trade agreements on members' bilateral international trade flows using (nonparametric) matching econometrics. Our nonparametric cross-section estimates of *ex post* long-run treatment effects are much more stable across years and have more economically plausible values than corresponding OLS cross-section estimates from typical gravity equations. We provide plausible estimates of the long-run effects of membership in the original European Economic Community (EEC) and the Central American Common Market (CACM) between 1960 and 2000 and the estimates confirm anecdotal reports of these agreements' effectiveness.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

For the past half century, quantitative evaluation of the effect of a trade policy change on the bilateral international trade of a pair of countries has been addressed traditionally in two ways, one *ex ante* and one *ex post*. *Ex ante* quantitative analysis of the effects of a policy change – such as formation of a free trade agreement (FTA) – on bilateral trade flows has been conducted using computable general equilibrium (CGE) models of trade, such as GTAP and the Michigan Model.¹ While pros and cons of these techniques have been discussed extensively, CGE models remain a standard tool to evaluate quantitatively *ex ante* trade effects of FTAs.

For decades, most *ex post* analyses of the (partial) effects of FTAs on trade flows have been conducted using “gravity equations,” a log-linear ordinary least squares (OLS) regression specification that typically is interpreted theoretically as the reduced-form from a formal general equilibrium model, cf., Anderson (1979), Bergstrand (1985), Anderson and van Wincoop (2003, 2004), Baier and Bergstrand (2001, 2002, 2007a,b), Eaton and Kortum (2002), Evenett and Keller (2002), Feenstra (2004), and Bergstrand et al. (2007). Up until a few years ago, empirical researchers typically employed cross-sectional data for a particular year or for multiple years and used the coefficient estimates associated with a dummy variable representing the presence or absence of an FTA to estimate the “average (partial) treatment effect” (ATE) of an FTA on members' bilateral trade, cf., Linnemann (1966), Aitken (1973), Sapir (1981, 2001), Baldwin (1994), Frankel (1997), and Schott (2005). However, such dummy variables' coefficient estimates often display extreme instability across years, and in many cases seemingly successful economic integration agreements – such as the European Union (formerly, European Economic Community) – have *negative* estimated treatment effects, cf., Frankel (1997). The “fragility” of these estimates has been documented using extreme-bounds analysis, cf., Ghosh and Yamarik (2004).²

* Corresponding author. Department of Finance, Mendoza College of Business, and Kellogg Institute for International Studies, University of Notre Dame, Notre Dame, IN 46556 USA. Tel.: +1 574 631 6761; fax: +1 574 631 5255.

E-mail address: Bergstrand.1@nd.edu (J.H. Bergstrand).

¹ In reality, there are numerous types of international economic integration agreements: free trade agreements, customs unions, common markets, and economic unions. For ease of reference, we use the term “free trade agreements” generically, as most economic integration agreements are FTAs or customs unions. The European Economic Community and Central American Common Market were customs unions initially.

² Widely varying estimates have also been found in empirical analyses of the effects of currency unions on trade, cf., Alesina et al. (2002, Table 8).

Yet, a fundamental question that remains to be addressed as the world continues to pursue economic integration agreements is simply this: What is the *ex post* long-run effect of a particular economic integration agreement between a pair of countries on the level of trade between those members? This is the central question addressed here.³

This paper is the first to evaluate *cross-sectionally* the (*ex post*) long-run treatment effects of FTAs on trade flow volumes using nonparametric estimation.⁴ Several reasons exist to employ a matching estimator as a nonparametric benchmark for the empirical analysis of FTA treatment effects. First, while the log-linear gravity equation has worked well for decades to explain fundamental determinants of bilateral trade flows (such as GDPs and bilateral distance), the effects of FTAs on trade flows may be related to the levels of trade flows and other covariates (i.e., nonlinearities may exist), cf., Brada and Mendez (1985), Frankel (1997), Baier and Bergstrand (2002), Anderson and van Wincoop (2003), Santos Silva and Tenreyro (2006), Bergstrand et al. (2007), and Henderson and Millimet (2008). Matching provides an alternative approach to log-linear gravity equations to estimate treatment effects without knowing a precise functional relationship. Since it is not yet clear that there is any *well-accepted* methodology for estimating *ex post* the effects of an FTA on trade between a country pair in a given year, it is important to consider alternative methodologies, which may or may not confirm previous estimates using gravity equations, such as those in Baier and Bergstrand (2007a), Baier et al. (2007), and Baier et al. (2008). Second, matching estimators allow ready estimation of average treatment effects when FTA treatments are subject to self-selection and the relationships between FTA treatments and other trade-flow covariates are nonlinear. As Baier and Bergstrand (2004a) show, selection into FTAs is not random. In fact, country pairs that select into FTAs tend to share similar economic characteristics that gravity equations use to explain their trade flows. The combination of non-random selection into FTAs and omitted non-linearities can bias OLS estimates of FTA treatment effects, cf., Persson (2001) for similar concerns in estimating the effects on trade of currency unions. Matching econometrics provide a simple method to form treatment and control groups by selecting on observable covariates and comparing observations drawn from the “same” distribution. For each country pair with an FTA, we construct a control group of country pairs with nearly identical economic characteristics but having no FTA; the average difference in trade flows between the pairs with FTAs and their control groups provides an estimate of “treatment.” Third, if the relationships between the FTA treatments and other trade-flow determinants are nonlinear, then the OLS average treatment effect (ATE) on a randomly selected pair will differ from the average treatment effect on the *treated* (ATT), and the latter may be of relevance for the policymaker.

We focus on two sets of potential contributions, one methodological and the other empirical. Methodologically, matching econometrics have been employed in economics predominantly in the labor economics literature, in particular in the evaluation of either job training or benefits programs in large cross-sections of individuals.

The key consideration for matching econometrics is the formulation of a credible counterfactual, cf., Diamond (2006a,b). First, international trade provides an excellent context (outside of labor economics) for evaluating treatment effects because the theoretical foundations for the gravity equation in international trade offer a convincing method for constructing a credible counterfactual. Well-established theoretical foundations for the gravity equation provide an excellent framework for selecting “control groups” (i.e., matched pairs without FTAs). Second, we are able to match treatments with their nearest neighbor controls and ensure that there are no statistically significant observable differences between the treatment and control groups for *all* conditioning characteristics common to the two groups. Successful identification of the control group requires acknowledging empirically the multilateral (price) resistance terms influencing trade identified in Anderson and van Wincoop (2003). By employing a simple Taylor-series expansion of the theoretically motivated Anderson–van Wincoop multilateral resistance terms, we are able to construct theoretically appropriate bilateral and multilateral “trade cost” covariates for selection. The theory and structure of the gravity model of trade is very useful to help determine the “selection on observables.”⁵

Empirically, we provide the first cross-sectional nonparametric matching estimates of long-run FTA treatment effects on levels of the volume of trade. First, we find across many settings and years that the matching estimates of treatment effects are much more stable and economically plausible than average treatment effect (ATE) estimates using typical cross-section OLS (or OLS with country fixed effects) gravity equations. The three main findings are: our ATT estimates indicate that FTAs increased members' trade by an economically and statistically significant amount in each of the nine years of our sample; the average treatment effects on the treated (ATTs) are lower than the ATEs for a randomly selected country pair; and the ATTs have less variance across years than ATEs calculated using matching techniques or using OLS. Second, we use matching econometrics to assess the effects of two historically prominent economic integration agreements on trade in a given year; specific treaties help to ensure an important assumption in matching estimation is met. Using nine cross-sections of annual trade flows from 1960 to 2000 (every 5 years), we find interesting results for two (still active) regional economic integration agreements that were formed between 1957 and 1960. First, the average treatment effect for a typical pair of members (ATT) of the *original* European Economic Community (EEC) was positive and economically significant from 1960 to 1970; by 1970, the EEC's Treaty of Rome had more than *doubled* on average the six members' bilateral trade. However, the enlargements of the Community in the early 1970s, the EC–EFTA free trade agreements, and the economic liberalization of Central and Eastern European countries and their subsequent FTAs with EC and EFTA members in the early 1990s reduced the original six members' bilateral trade significantly (i.e., trade diversion).⁶ Second, we find that the treatment effect for a typical pair of countries in the original Central American Common Market (CACM) was positive and rose monotonically until 1970. However, in the early 1970s and early 1980s, the treatment effect was considerably lower, rebounding in the 1990s. These estimates support anecdotal evidence that the CACM flourished in the 1960s, but as some member countries faced considerable political instability and cross-border armed conflicts beginning in 1969 and then again in 1979, CACM became less effective between 1970 and 1990, reinvigorated by the nations' presidents in the 1990s.

³ Evidence for long-run effects of other trade policies and their importance has been addressed in the literature, cf., Trefler (1993). Moreover, Anderson and van Wincoop (2003) and Bergstrand et al. (2007) remind us that there are both long-run partial- and general-equilibrium effects. In this paper, we focus only on the long-run partial (or “treatment”) effects. However, the partial effects estimated here could be combined with a nonlinear system of structural equations to generate general-equilibrium comparative statics, as done in the two papers just noted.

⁴ We know of one other study that has used matching estimation for FTA effects on trade flows, cf., Egger et al. (2008). However, that study focused on the effects of FTAs on intra-industry trade-share indexes, rather than trade volumes, and focused on *short-run* changes requiring a panel to do difference-in-differences estimation. As a tangent, that study did provide panel estimates of short-run (2-year) new FTA effects on trade flows using difference-in-differences matching. In international trade, a few nonparametric studies have addressed the long-run effects of *currency unions* on trade flow volumes using propensity-score estimates in cross-section, cf., Persson (2001), Kenen (2002), and Chintrakarn (2008), but not long-run effects of FTAs.

⁵ Note that instrumental variables (IV) is not appropriate for addressing the bias raised by selection on observables. IV can address potentially the issue of selection bias on unobservables, cf., Baier and Bergstrand (2002). However, two points are worth noting. First, many have argued that the selection bias on observables may well dominate that on unobservables. Second, it is difficult to find instruments for FTAs that are not also correlated with trade flows, making IV techniques' usefulness limited, cf., Baier and Bergstrand (2007a), discussed more later.

⁶ The EEC began to be called the EC (European Community) in the 1970s. EFTA denotes the European Free Trade Association, which was formed in 1960.

The structure of the paper is as follows. Section 2 motivates the matching estimator as an alternative to gravity equations in the empirical trade literature to estimate *ex post* the effects of FTAs on members' trade flows. Section 3 addresses the matching estimator approach. Section 4 presents the empirical findings for average treatment effects across all FTAs and a sensitivity analysis. Section 5 presents the estimated average treatment effects for two particular treatments: EEC and CACM. Noting some caveats, Section 6 concludes that FTAs have had considerably larger effects on members' trade (*ex post*) than previous gravity models have suggested and that the coefficient estimates have economically plausible values and are much less fragile.

2. Motivation for nonparametric estimation of FTA effects on trade

Conventional wisdom leads one to argue that the trade between a pair of countries should increase when two countries enter into an FTA. First, trade could increase due to (gross) trade “creation” as the pair benefits from lower trade costs between them. Second, trade between a pair could result from the “diversion” of imports from (previously) lower cost suppliers. For both reasons, trade (on the extensive and intensive margins) between the pair should increase. As Table 1 verifies, country pairs with FTAs trade more on average than those without FTAs. However, such statistics do not imply that FTAs necessarily *cause* more trade.

2.1. Traditional (atheoretical) gravity equations

Ex post econometric analysis of the effects of FTAs on (bilateral) trade has been dominated over the past 45 years by the gravity equation. A typical specification estimated by OLS is:

$$\ln TF_{ijt} = \beta_0 + \beta_1 (\ln GDP_{it} GDP_{jt}) + \beta_2 (\ln DIST_{ij}) + \beta_3 (ADJ_{ij}) + \beta_4 (LANG_{ij}) + \beta_5 (FTA_{ijt}) + \varepsilon_{ijt} \quad (1)$$

where TF_{ijt} denotes the sum of the values of the nominal bilateral trade flows between countries i and j in year t , GDP_{it} (GDP_{jt}) denotes the nominal gross domestic product in country i (j) in year t , $DIST_{ij}$ denotes the bilateral distance between the economic centers of countries i and j , ADJ_{ij} is a dummy variable assuming the value 1 if both countries are adjacent (i.e., share a land border) and 0 otherwise, $LANG_{ij}$ is a dummy variable assuming the value 1 if both countries share a common language and 0 otherwise, FTA_{ijt} is a dummy variable assuming the value 1 if both countries are members of a free trade agreement (or deeper economic integration agreement) in t and 0 otherwise, and ε_{ijt} is a normally-distributed error term.⁷

Formulations in Frankel (1997, Table 4.2) provide representative and highly cited *ex post* estimates of average treatment effects (ATEs) of several FTAs for several years.⁸ As is common, the effects of such agreements – say, the EEC – on trade vary considerably across years, even though the coefficients of the other RHS variables are fairly stable. One even finds *negative* effects of the EEC on trade – other things constant – in some years. The instability of these estimates across years – and even some negative values – raised concerns in Frankel (1997), but were not addressed. Moreover, Ghosh and Yamarik (2004) conducted an extreme-bounds analysis of the estimated effects of several FTAs on trade using cross-section data in typical gravity equations and found the estimates to be quite “fragile.”

⁷ Some studies have shown recently that zero trade flows can be accounted for by estimating a multiplicative version of the gravity equation using a Poisson Quasi Maximum Likelihood estimator, cf., Santos Silva and Tenreiro (2006). However, this issue is beyond the scope of our paper, but a useful direction for future research.

⁸ OLS estimates of the coefficient for the FTA dummy variable provide an average treatment effect (ATE) on a randomly selected country pair. By contrast, policymakers may be interested in the average treatment effect on a country pair with an FTA – that is, the treated (ATT). Generally, the ATE is an average of the ATT and the average treatment effect on an untreated pair (ATU).

Table 1

Sum of bilateral trade flows in countries with FTAs and without FTAs (in logs)

	Log sum of trade (with an FTA)	Log sum of trade (without an FTA)	Difference in logs
1960	11.33	9.21	2.11
1965	11.55	9.32	2.23
1970	12.21	7.68	4.53
1975	12.55	9.30	3.25
1980	13.22	9.94	3.28
1985	14.33	10.57	3.76
1990	14.14	10.43	3.71
1995	14.92	10.41	4.51
2000	14.57	10.50	4.07

Nominal bilateral trade flows are from the International Monetary Fund's *Direction of Trade Statistics* for 96 countries, which are listed in Baier and Bergstrand (2007a, Data Appendix). There are potentially $96 \times 95/2 = 4560$ bilateral trade sums (e.g., aggregate trade flow from France to Germany plus that from Germany to France); however, zero trade flows and missing values are excluded. For later analysis for real trade flows, nominal flows are divided first by the exporter's GDP deflator.

Table 2 provides some representative ATE estimates of FTAs for several years for 96 countries using OLS estimation of Eq. (1). We note first that in some years (e.g., 1970) an FTA had a large positive effect on trade. However, in other years (e.g., 1980, 1990, 1995) an FTA had a *negative* effect on trade (though statistically insignificant); these results are quite similar to those in Frankel (1997). Frankel (1997, p. 62) showed that, for some years, membership in the European Community *reduced* trade. Second, the ATEs are very unstable across years.⁹ Although we would expect different agreements to have different economic effects, even allowing for differing economic impacts, one might not expect the average effect of an FTA to vary from an *increase* in trade of 274% in 1970 to a *decrease* of 10% in 1980. Such empirical results seem implausible.

2.2. Methodological problems

Several methodological reasons can potentially explain the fragility of these ATE estimates. First, the FTA dummy is likely representing other factors – unobserved by the researcher – that influence trade but are omitted. These omitted variables may bias the coefficient estimates up or down. Considerable work has been done over the past four decades to address conventional omitted variables bias in gravity equations. Rose (2004), for example, provides one of the most extensive treatments of omitted variables bias by including a range of “trade cost” proxies.

Perhaps the most influential recent theoretical and empirical paper addressing omitted variables bias in the gravity equation is Anderson and van Wincoop (2003). These authors demonstrate that traditional gravity Eq. (1) is mis-specified and coefficient estimates of RHS variables are likely biased owing to omission of nonlinear “multi-lateral (price) resistance” terms of countries i and j in each year (or including inappropriately atheoretical “remoteness” indexes). Anderson and van Wincoop show that estimation of unbiased coefficients ($\alpha_0, \dots, \alpha_4$) requires minimizing the sum-of-squared residuals of (with time-subscript t now omitted):

$$\ln [TF_{ij} / (GDP_i GDP_j)] = \alpha_0 + \alpha_1 \ln DIST_{ij} + \alpha_2 ADJ_{ij} + \alpha_3 LANG_{ij} + \alpha_4 FTA_{ij} - \ln P_i^{1-\sigma} - \ln P_j^{1-\sigma} + \varepsilon_{ij} \quad (2)$$

subject to the N nonlinear market-equilibrium conditions:

$$P_i^{1-\sigma} = \sum_{k=1}^N P_k^{\sigma-1} (GDP_k / GDP^T)^{\alpha_1 \ln DIST_{ki} + \alpha_2 ADJ_{ki} + \alpha_3 LANG_{ki} + \alpha_4 FTA_{ki}} \quad i = 1, \dots, N \quad (3)$$

⁹ As evidence, the coefficient of variation of the coefficient estimates for FTA across the nine years was six times that of the next highest coefficient estimate's coefficient of variation.

Table 2
Typical gravity equation coefficient estimates

Variable	(1) 1960	(2) 1965	(3) 1970	(4) 1975	(5) 1980	(6) 1985	(7) 1990	(8) 1995	(9) 2000
$\ln(GDP_i GDP_j)$	0.75 (43.53)	0.76 (53.07)	0.81 (57.18)	0.84 (63.64)	0.90 (71.01)	0.87 (73.25)	0.90 (79.62)	0.95 (93.88)	0.99 (100.11)
$\ln DIST_{ij}$	-0.56 (-1.33)	-0.65 (-14.40)	-0.65 (-14.10)	-0.75 (-16.29)	-0.88 (-21.19)	-0.88 (-22.48)	-0.98 (-24.24)	-1.01 (-26.15)	-1.08 (-27.39)
AD_{ij}	0.31 (1.64)	0.06 (0.37)	0.10 (0.54)	0.33 (1.93)	0.45 (2.77)	0.43 (2.80)	0.40 (2.40)	0.63 (3.99)	0.84 (5.17)
$LANG_{ij}$	0.03 (0.25)	0.24 (2.33)	0.36 (3.29)	0.48 (4.57)	0.47 (4.72)	0.39 (3.98)	0.68 (6.57)	0.76 (7.62)	0.63 (6.35)
FTA_{ij}	0.43 (2.02)	0.80 (4.24)	1.32 (6.08)	0.42 (2.47)	-0.11 (-0.62)	0.32 (1.98)	-0.04 (-0.27)	-0.04 (-0.27)	0.15 (1.37)
Constant	-9.66 (-15.28)	-9.45 (-17.45)	-11.58 (-20.89)	-12.17 (-23.04)	-13.40 (-26.11)	-12.66 (-26.76)	-13.49 (-28.83)	-14.97 (-35.52)	-15.87 (-37.12)
RMSE	1.1504	1.1089	1.2715	1.3218	1.3342	1.3227	1.4315	1.4029	1.4717
R^2	0.6649	0.6985	0.6969	0.7038	0.7250	0.7171	0.7308	0.7684	0.7741
No. observ.	1059	1325	1570	1947	2189	2433	2802	3073	3342

t-statistics are in parentheses. The dependent variable is natural log of the sum of nominal bilateral trade flows between *i* and *j*. The gravity equation was also estimated using the log of the bilateral trade flow from *i* to *j*; the results are not materially different. The data are the same as used in Baier and Bergstrand (2007a). The binary variable for an FTA (which also includes customs unions, common markets, and economic unions) was constructed by the authors, is described in Section 5 and Table 3 (a listing of all the agreements used) of Baier and Bergstrand (2007a), and is available at <http://www.nd.edu/~jbergstr> and <http://people.clemson.edu/~sbaier>. See Section 4.1, Data Description for sources for other variables.

where, in the model's context, $\alpha_0 = -\ln GDP^T$, where GDP^T is world GDP (a constant). One insight from Eqs. (2) and (3) is that the potential effect of FTA_{ij} on TF_{ij} is nonlinear. This obviously requires a custom nonlinear least squares (NLS) program to estimate unbiased coefficients.

Estimation of this system of nonlinear equations using NLS for the non-zero trade flows among the 96 countries in our study is beyond the scope of this paper. However, as Anderson and van Wincoop (2003) and Feenstra (2004) note, unbiased estimates of α_0 through α_4 can be obtained also using country fixed effects for $P_i^{1-\sigma}$ and $P_j^{1-\sigma}$. Table 3 reports the results of estimating Eq. (2) using fixed effects to account for the multilateral resistance terms $P_i^{1-\sigma}$ and $P_j^{1-\sigma}$. A comparison of Table 2 with Table 3 indicates that omission of the multilateral resistance terms in the traditional gravity specification tends to bias the FTA coefficient estimates. However, considerable instability of the ATEs for FTAs from year to year remains, several estimates are statistically insignificant, and – perhaps most importantly – some negative estimates seem implausible. For example, the effect of an FTA in 1970 was to more than double members' trade. However, by 1980, an FTA on average reduced trade by 54%. In fact, the results imply that by 2000 an FTA had no economically or statistically significant effect on members' trade! Thus, the issues raised in Anderson and van Wincoop (2003) are important, but do not fully address our concerns. Proper estimation of the gravity equation does require recognizing theoretically-motivated multilateral price terms. However, the Anderson–van Wincoop procedure cannot fully ensure unbiased estimation of the effect of an FTA on two members' trade for reasons detailed below. Still, these recent advances in the theoretical foundations of the gravity equation will be quite important for our matching estimation technique's "selection on observables."

Even after accounting for multilateral price terms, (log-) linear regressions may not provide accurate estimates of the average treatment effect of an FTA if (nonlinear) interactions of FTA_{ij} and typical gravity-equation trade-flow determinants are significant and selection into FTAs is non-random (i.e., FTA_{ij} is systematically correlated with these standard gravity-equation covariates). First, there is now significant evidence that

FTAs are not determined randomly. For instance, Baier and Bergstrand (2004a) provide empirical evidence that pairs of countries that are larger in economic (GDP) size, more similar in GDPs, closer in distance, and more remote from other countries tend to have FTAs and they provide a theoretical rationale for each relationship. Of course such variables are typical gravity equation regressors. Indeed, Table 4 reveals that the means of traditional gravity equation covariates differ significantly between country pairs with FTAs (i.e., treated) versus those without FTAs (i.e., untreated) for 1990. Moreover, Figs. 1 and 2 show quite dramatically that for two typical gravity equation covariates – (log) bilateral distance and (the sum of the logs of) GDPs – the distributions of treated pairs differ substantively from those of untreated pairs for all nine years of our sample. In Fig. 1, the kernel density function of (log) bilateral distances for non-FTA country pairs (dashed line) is centered considerably to the right of that for FTA country pairs (solid line); country pairs with FTAs tend to be closer. In Fig. 2, the kernel density function of GDPs of pairs with FTAs (solid line) is centered to the right of that for pairs without FTAs (dashed line); country pairs with FTAs tend to be economically larger. Similar results hold for other covariates (not shown, for brevity).

Second, it is possible that the effects of an FTA on trade are dependent on levels of typical gravity-equation covariates. For instance, Brada and Mendiz (1985) and Frankel (1997) found economically and statistically significant interaction effects between FTA dummies and log bilateral distance in gravity equations. Baier and Bergstrand (2002) provide empirical evidence that interaction terms of FTA_{ij} with log GDPs and FTA_{ij} with log populations are economically and statistically significant in gravity equations. Moreover, in the parallel gravity-equation literature on the effects of currency unions on trade flows, Rose (2000) and Persson (2001) provide evidence of significant nonlinear terms interacting currency union dummies with typical gravity-equation covariates, such as log GDPs and log distance. As noted in Persson (2001), if selection into FTAs is non-random and if nonlinearities exist between the FTA variable and other covariates in the gravity equation, then parametric estimates of treatment effects will be biased and the average

Table 3
Theoretically-motivated gravity equation coefficient estimates (using country fixed effects)

Variable	(1) 1960	(2) 1965	(3) 1970	(4) 1975	(5) 1980	(6) 1985	(7) 1990	(8) 1995	(9) 2000
$\ln(GDP_i GDP_j)$	0.32 (1.68)	0.79 (11.55)	1.00 (18.17)	0.97 (24.72)	0.98 (71.01)	0.89 (26.75)	0.90 (79.62)	1.06 (38.09)	1.16 (39.47)
$\ln DIST_{ij}$	-0.70 (-13.47)	-0.79 (-17.25)	-0.74 (-16.15)	-0.88 (-18.53)	-1.06 (-24.40)	-1.05 (-25.39)	-0.98 (-24.24)	-1.20 (-29.00)	-1.35 (-32.26)
AD_{ij}	0.28 (1.68)	-0.05 (-0.31)	0.12 (0.77)	0.45 (2.90)	0.39 (2.60)	0.33 (2.34)	0.40 (2.40)	0.48 (3.34)	0.61 (4.19)
$LANG_{ij}$	0.22 (1.80)	0.42 (3.92)	0.76 (6.70)	0.60 (5.43)	0.69 (6.68)	0.62 (6.11)	0.68 (6.57)	0.88 (8.57)	0.80 (7.81)
FTA_{ij}	-0.01 (-0.06)	0.33 (2.00)	0.74 (3.86)	-0.10 (-0.62)	-0.78 (-4.60)	-0.42 (-2.74)	-0.41 (-2.72)	-0.23 (-1.79)	-0.08 (-0.75)
Constant	-14.94 (-29.22)	-9.19 (-4.79)	-16.16 (-10.15)	-14.72 (-11.92)	-14.50 (-12.54)	-11.59 (-10.82)	-14.40 (-13.72)	-16.48 (-18.21)	-18.65 (-19.88)
RMSE	0.9369	0.9144	1.0497	1.1262	1.1395	1.1317	1.2325	1.2222	1.2679
R^2	0.7919	0.8056	0.8024	0.7930	0.8064	0.7999	0.8063	0.8289	0.8365
No. observ.	1059	1325	1570	1947	2189	2433	2802	3073	3342

t-statistics are in parentheses. The dependent variable is natural log of the sum of nominal bilateral trade flows between *i* and *j*. Coefficient estimates and *t*-statistics for country fixed effects are not presented (for brevity).

Table 4
Summary of 1990 covariate means

	Country pairs with an FTA	Country pairs without an FTA	Difference
Log of distance	6.66	8.28	−1.62 (−23.80)
Sum of the logs of GDPs	38.55	34.69	3.87 (15.98)
Adjacency dummy	0.18	0.07	0.11 (10.46)
Language dummy	0.07	0.02	0.05 (−0.11)

t-statistics are in parentheses.

treatment effects on the *treated* (ATTs) will differ from the average treatment effects on randomly selected pairs (ATEs).

This paper employs nonparametric matching techniques to provide unbiased estimates of the long-run treatment effects of FTAs on trade flows using cross-section data. First, the key to successful matching estimates is to generate a credible counterfactual trade flow for an “untreated” matched country pair. Theoretical foundations for the gravity equation provide an excellent context for the selection of observables that can be used to generate matched untreated country pairs that are statistically indistinguishable from treated pairs, except for treatment. Matching estimation can avoid the bias introduced by non-random selection on observables and by non-linearities.

Second, while parametric techniques exist that can adjust for selection on observables bias and nonlinearities can be introduced in regressions, the virtually unlimited potential specifications suggest that matching estimates of treatment effects provide “benchmark” nonparametric estimates of long-run treatment effects. Consequently, they provide a useful “check” upon ATE and ATT estimates computed using regression analysis with alternative nonlinear structures used on nonexperimental data.

Third, there has been no study that has attempted to estimate *long-run* effects of FTAs on trade based upon cross-section data using nonparametric techniques. In international trade, several papers have used nonparametric techniques, such as propensity-score estimators, to evaluate the long-run effects of currency unions on trade. For instance, Persson (2001) found strikingly different treatment-effect estimates for currency unions using a propensity-score matching estimator in comparison to Rose (2000) using cross-section gravity equations. Rose found a treatment effect for currency unions of over 200%; Persson found treatment effects that were not (statistically) significantly different from zero.¹⁰ However, while Persson’s currency union ATEs were statistically insignificant, the specifications of his propensity-score functions lacked theoretical foundations, and many of the “balancing properties” were violated. Our study differs from Persson (2001) by examining FTA – not currency union – treatment effects, focusing on a matching technique,¹¹ and selecting covariates based upon recent theoretical developments for the trade gravity equation, cf., Anderson and van Wincoop (2003) and Baier and Bergstrand (2006, 2007b).¹²

One recent study in international trade has investigated the effects of FTAs on trade volumes and intra-industry trade indexes. Egger et al. (2008) used a difference-in-differences panel matching estimator to examine primarily the effect of FTA formations on *changes* in shares of intra-industry trade. However, with a two-year window on either side of the “events,” this study could only determine *short-run* effects of FTAs on changes in the trade volumes and degrees of intra-industry

trade among countries. Difference-in-differences techniques for 5-year windows might not nearly capture the trade-enhancing effects of FTAs. First, the vast bulk of FTAs have phase-in periods of 5 to 10 years. Second, even once the agreements are fully phased in, changed terms-of-trade have lagged effects on trade flows. Using first-differenced data and parametric techniques, Baier and Bergstrand (2007a) showed that FTAs affected trade 15 years after their first year of implementation. Thus, our study differs from Egger et al. (2008) by examining “long-run” volume of trade treatment effects of FTAs rather than short-run effects, focusing on volume-of-trade rather than intra-industry-index effects of FTAs, and selecting covariates based upon the recent theoretical developments for the trade gravity equation.¹³

3. A matching-estimator approach

The literature on matching econometrics is now well established, cf., Moffitt (2004), Cameron and Trivedi (2005), Lee (2005), and Abadie and Imbens (2006). In our context, the key is finding a (control) group of country pairs without FTAs that are virtually identical in all other respects as trade partners to a pair with an FTA. The two groups of country pairs are selected to be identical in all respects but treatment by selecting on observables for the pairs; members of the two groups are assigned based upon a set of common economic characteristics. The key is to simulate random assignment into treatment and control. The average treatment effect for the entire sample (ATE) and the average treatment effect for treated groups (ATT) are obtained by comparing the trade flows between two groups, the treated and untreated (or control group), and taking the average difference.

Interestingly, large-sample properties of such estimators have not been established until just recently, cf., Abadie and Imbens (2006). That study notes that matching estimators are highly “nonsmooth functionals” of the distribution of data, which are not readily addressed using standard asymptotic distribution theory. They find that matching estimators include a conditional bias, the stochastic order of which increases with the number of continuous matching variables. In this paper, we use nearest neighbor matching as suggested by Abadie–Imbens (nearest three neighbors) and adjust for the conditional bias using the Abadie–Imbens (A–I) technique in a robustness analysis, which does not require consistent nonparametric estimation of unknown functions. For robustness, we also estimate propensity scores, to be discussed later.

We determine the ATTs for various years and treated groups; for brevity, we ignore the *ij* variable subscripts in the following. Let TF_1 (TF_0) denote the value of trade between two countries with (without) an FTA and let *FTA* be a dummy variable assuming the value 1 (0) if the two countries have (do not have) an FTA. Let **x** denote a set of characteristics (covariates) that influence the trade flow between pairs of countries, and **X** be a random vector of dimension *k* of continuous covariates distributed on \mathbb{R}^k with compact and convex support *X*.

We follow Abadie and Imbens (2006) and make three conventional assumptions. First, for almost every $\mathbf{x} \in X$, assume *FTA* is independent of TF_1 and TF_0 – conditional upon the set of covariates $\mathbf{X}=\mathbf{x}$, which is the assumption of “conditional mean independence” or “ignorability of treatment” according to Rosenbaum and Rubin (1983); this is

¹⁰ Tenreiro also found smaller effects of currency unions on trade, taking into account self-selection of country pairs into currency unions, building upon the work of Alesina and Barro (2002) and Alesina et al. (2002).

¹¹ A limited number of covariates determining trade flows implies we do not face the “curse of dimensionality,” necessitating a propensity-score approach.

¹² As discussed in footnote 3, we focus only on estimating long-run “partial,” or treatment, effects; we do not address general equilibrium comparative statics in this analysis, an issue for future research.

¹³ As in Persson (2001), our focus is on addressing potential selection bias on observables and nonlinearities. We do not address selection on unobservables. In nonparametric estimation, the standard approach to address selection bias on unobservables is to construct difference-in-difference estimators for pre- and post-treatments over short and balanced periods, as in Egger et al. 2008; such a method would preclude estimating the long-run effects of FTAs (which can be up to 15 years), which is the focus of this paper. The standard cross-section parametric technique to adjust for selection bias on unobservables (“endogeneity” bias) is to use instrumental variables. However, parametric estimation requires specifying functional form and it is very difficult to find an instrument to predict FTAs that is not also correlated with trade flows, cf., Baier and Bergstrand (2004b, 2007a).

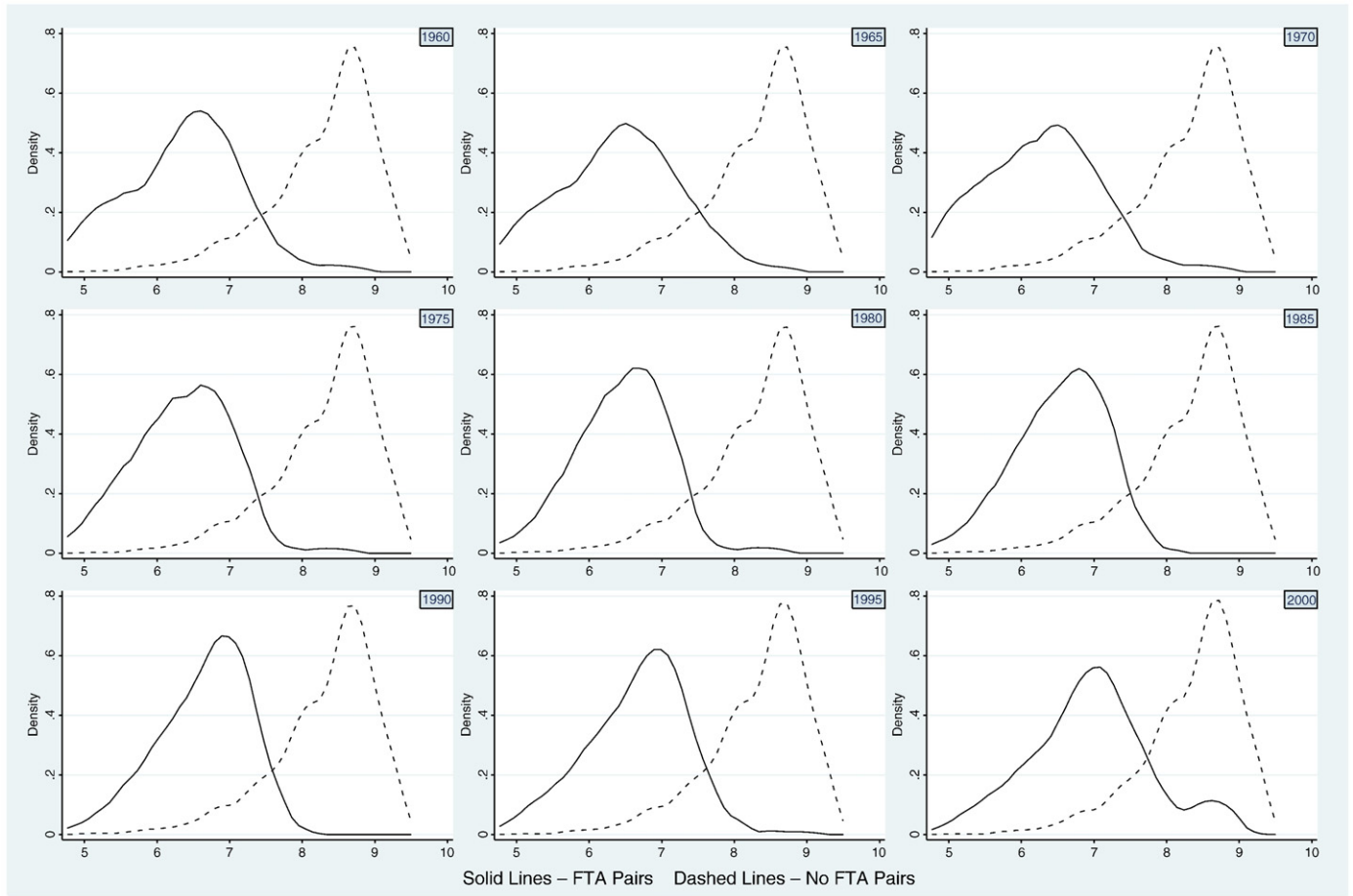


Fig. 1. Log of distance for bilateral pairs with and without an FTA.

Assumption (A1). Second, assume that $\eta < P(FTA=1|X=x) < 1-\eta$ for some $\eta > 0$, where $P(\cdot)$ denotes the probability; this is Assumption (A2). This assumption (called the “overlap” assumption) ensures that for each value of x there are both treated and untreated observations. The third assumption (Assumption A3) usually made is generally referred to as the stable-unit-treatment-value assumption (SUTVA); however, it has two parts. The first part of SUTVA (A3a) ensures that there is a unique treatment (no “multiple versions”). The second part (A3b, called “non-interference”) implies that treatment of any country pair does not influence the trade of other (untreated) pairs; that is, TF , x , and FTA are independent draws from their distributions. We now discuss the feasibility of all three assumptions in the context of international trade flows and their determinants.

3.1. Ignorability Assumption (A1)

The ignorability (also called “unconfoundedness”) assumption states that, conditioned on covariates x , FTA is independent of TF_0 and TF_1 (omitting observational subscripts for country pairs). The purpose of this assumption is to try to ensure that treatment assignment is “random.” This assumption is assured by choosing a control group for each treated pair that is matched closely to the treated pair in terms of all relevant covariates influencing trade (other than FTA). Hence, this assumption is also known as “selection on observables.”

Fortunately, theoretical foundations for the gravity equation of trade flows provide guidance for selecting observables. As Eqs. (2) and (3) above suggest, relevant covariates to use for selection on observables include the sum of logs of GDPs, log of bilateral distance, adjacency and language dummies, and measures of the “multilateral resistance” (MR)

terms. However, in Anderson and van Wincoop (2003), the MR terms are endogenous. Fortunately, Baier and Bergstrand (2006) show that a first-order log-linear Taylor-series expansion (around a symmetric equilibrium) of the system of Eq. (3) yields a reduced-form function of linear combinations of the “exogenous” variables in Eqs. (2) and (3):

$$\ln[TF_{ij}/(GDP_i GDP_j)] = \beta_0 + \beta_1 BVDIST_{ij} + \beta_2 BVADJ_{ij} + \beta_3 BVLANG_{ij} + \beta_4 BVFTA_{ij} + \epsilon_{ij} \quad (4)$$

where

$$\begin{aligned} BVDIST_{ij} &= \ln DIST_{ij} - (1/N) \sum_{j=1}^N \ln DIST_{ij} - (1/N) \sum_{i=1}^N \ln DIST_{ij} + (1/N^2) \sum_{i=1}^N \sum_{j=1}^N \ln DIST_{ij} \\ BVADJ_{ij} &= \ln ADJ_{ij} - (1/N) \sum_{j=1}^N \ln ADJ_{ij} - (1/N) \sum_{i=1}^N \ln ADJ_{ij} + (1/N^2) \sum_{i=1}^N \sum_{j=1}^N \ln ADJ_{ij} \\ BVLANG_{ij} &= \ln LANG_{ij} - (1/N) \sum_{j=1}^N \ln LANG_{ij} - (1/N) \sum_{i=1}^N \ln LANG_{ij} + (1/N^2) \sum_{i=1}^N \sum_{j=1}^N \ln LANG_{ij} \\ BVFTA_{ij} &= \ln FTA_{ij} - (1/N) \sum_{j=1}^N \ln FTA_{ij} - (1/N) \sum_{i=1}^N \ln FTA_{ij} + (1/N^2) \sum_{i=1}^N \sum_{j=1}^N \ln FTA_{ij} \end{aligned}$$

and the β s have theoretically interpretable values, cf., Baier and Bergstrand (2006, 2007b).¹⁴ This framework suggests exogenous

¹⁴ In the case of centering at a symmetric equilibrium, each of the “BV” variables is quite simple. For instance, for $BVDIST_{ij}$ the second (third) component is the mean distance of i (j) from all its trading partners and the fourth term is simply a constant. Baier and Bergstrand (2006) also show that a first-order log-linear Taylor-series expansion of Eq. (3) around alternatively a frictionless equilibrium yields terms similar to $BVDIST$, etc., except that the second, third, and fourth components are *GDP-share-weighted* (rather than simple) averages. See Baier and Bergstrand (2006, 2007b) for the economic intuition.

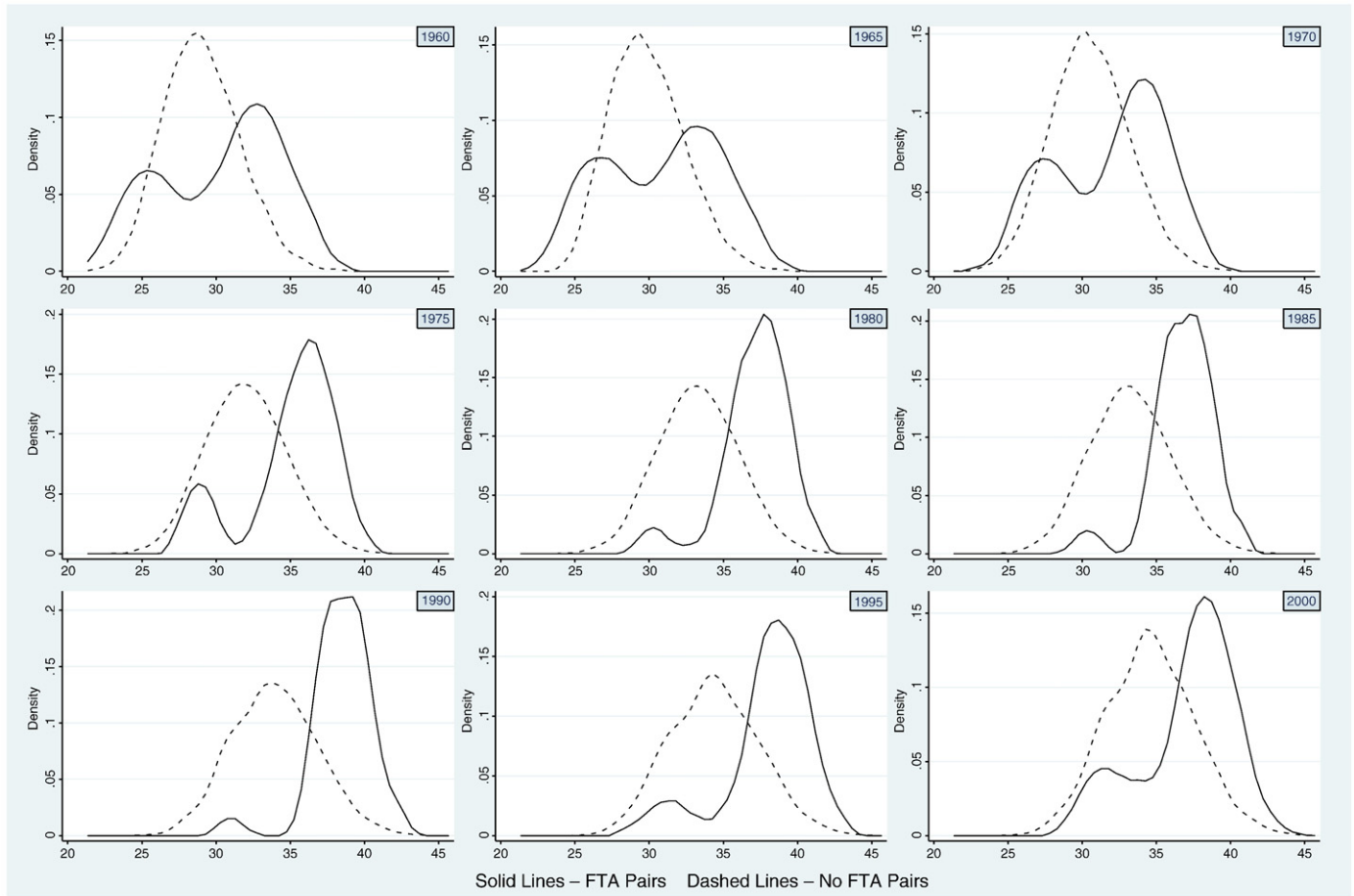


Fig. 2. Sum of logs of GDPs for bilateral pairs with and without an FTA.

covariates $SLGDP_{ij}$ (the sum of the logs of GDP_i and GDP_j), $BVDIST_{ij}$, $BVADJ_{ij}$, and $BVLANG_{ij}$ for conducting the matching estimation (relaxing the assumption of strictly unitary income elasticities).

A key consideration for successful matching is that the distributions of the covariates for the treated and matched untreated pairs be virtually indistinguishable. Fig. 3 confirms this, showing for instance the kernel density functions for the sum of the logs of GDPs ($SLGDP$) for the matched pairs. Fig. 3 reveals much greater similarity of the distributions than corresponding Fig. 2 for $SLGDP$. In fact, as conventional we estimated the differences of the means for treated and matched untreated pairs for all four exogenous covariates for all nine years. Among the 36 two-tailed t -tests for the four variables, we could reject at the 5% significance level the null of identical means in only two cases ($SLGDP$ in 1990 and $BVDIST$ in 2000). For brevity, we do not report all 36 statistics (but these are available on request).

3.2. Overlap Assumption (A2)

The second assumption (“overlap”) ensures that for each value of \mathbf{x} there are both treated and untreated observations. In our trade context, the very large number of country pairs with FTAs and without FTAs ensures that the overlap assumption is not violated.

3.3. SUTVA (A3)

In matching estimation, SUTVA is used to ensure (a) that the “treatment” (here, the FTA) is identical for each treated observation (i.e., no “multiple versions” of the treatment) and (b) that the formation of an FTA between a country pair ij does not influence

untreated trade flows (the “non-interference” assumption). Also, an FTA between another country pair ik (or kj or kl) should not influence trade between ij . In trade, many economists argue that an FTA between a country pair (say, ik) might lead to trade diversion, that is, a reduction of TF_{ij} ; this violates the non-interference assumption. Also, when FTA_{ij} is not “treaty specific” (such as EEC or CACM), then the “no-multiple-versions” of treatment assumption is likely violated as well. We will address both issues.

The history of applied matching studies in labor economics regarding job training programs suggests that researchers have justified their use of this assumption on the notion that the program is identical for treated individuals and that no single treated individual's income gains influence materially others' incomes (nor do others' treatments influence materially their incomes). Given the growing interest in matching estimators in the social sciences – and in economics, in particular – we use the theoretical foundations for the gravity equation to offer some clarity regarding this assumption in our context of international trade, in the spirit of Diamond (2006a,b). First, suppose world trade flows are determined according to the Anderson and van Wincoop (2003) structural model represented by Eqs. (2) and (3) in Section 2. By convention, the “treatment effect” might be measured by the value of α_4 . However, these equations show clearly the potential “general-equilibrium” effects that the non-interference assumption attempts to preclude. For instance, suppose the gravity-equation specification ignored $P_i^{1-\sigma}$ and $P_j^{1-\sigma}$. A change in any FTA_{ik} would alter the multilateral price term $P_i^{1-\sigma}$, confounding measurement of α_4 .

One solution, of course, for estimating the parameters is to follow the path of estimating the entire system of structural equations using,

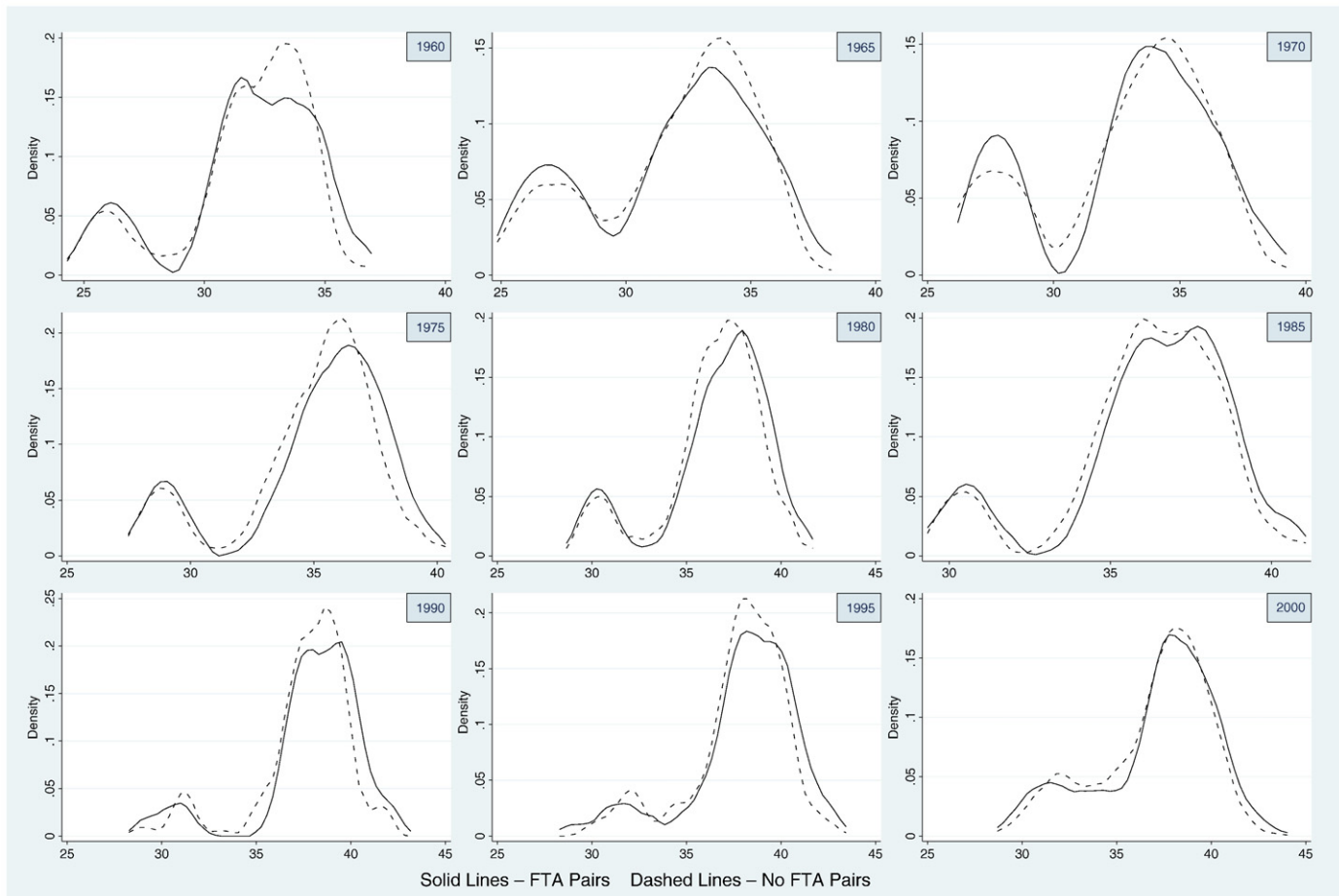


Fig. 3. Sum of logs of GDPs for matched bilateral pairs with and without an FTA.

say, nonlinear least squares, cf., [Anderson and van Wincoop \(2003\)](#) and [Heckman et al. \(1998\)](#). However, this approach imposes even more structure on the estimates of α_4 , which is exactly what nonparametric estimation is trying to avoid. Is there a theoretically feasible way to account for the variation in variables $P_i^{1-\sigma}$ and $P_j^{1-\sigma}$ given the system of nonlinear Eqs. (2) and (3) – in the context of matching econometrics – which must also address the “curse of dimensionality”?

This issue is addressed by again appealing to the [Baier and Bergstrand \(2006, 2007b\)](#) framework discussed above. By selecting on observables such as $BVDIST_{ij}$, $BVADJ_{ij}$, and $BVLANG_{ij}$, we account for the general equilibrium aspects other than FTAs among other pairs for estimating treatment effects. However, in the context of the theory, the “treatment” is $BVFTA_{ij}$. That is, the difference in the trade flows between the observed and the counterfactual for country pair ij is driven by FTAs between the country pair ij – but also between country i with any of its trading partners, between country j with any of its trading partners, and between any country pair kl , which some may argue violates the “non-interference” assumption.

However, the researcher knows *when* policies were implemented. This identifies more precisely the actual “treatment” and potentially eliminates interference from the second, third and fourth terms in $BVFTA_{ij}$. For example, later in [Section 5](#) we will consider the Treaty of Managua signed in 1960 to create the Central American Common Market (CACM). In this case, SUTVA is more likely to hold as the “treaty” and degree of integration was well-defined for every CACM member (no “multiple versions” is ensured by the common treatment of internal tariff elimination and imposition of identical external tariff

rates) *and* – because the Central American nations were both economically small and remote from the only other effective FTAs in the 1960s, 1970s, and 1980s (EEC and EFTA) – there was likely little interference (no “general equilibrium effects”). That is, the small economic size and trade of CACM countries likely ensured no perceptible influence on untreated pairs’ trade (similar to job-retrained individuals likely not having any perceptible impact on untreated individuals’ incomes), and the remoteness of the CACM countries from Europe likely ensured that the formation of EEC and EFTA had no perceptible influence on CACM members’ trade. Hence, for CACM in particular, $BVFTA_{ij}$ equals FTA_{ij} (implying no interferences) and members shared a common treatment (Treaty of Managua).

3.4. [Abadie and Imbens \(2006\)](#) matching estimator

Under these assumptions, the sample ATE with $\mathbf{X} = \mathbf{x}$ is given by:

$$ATE(\mathbf{x}) = E[TF|FTA = 1, \mathbf{X} = \mathbf{x}] - E[TF|FTA = 0, \mathbf{X} = \mathbf{x}] = E[TF_1 - TF_0 | \mathbf{X} = \mathbf{x}] \quad (5)$$

which is the ATE conditional on \mathbf{x} . By iterated expectations, averaging across the distribution of \mathbf{X} gives $ATE = E[TF_1 - TF_0]$. Similarly, the average treatment effect on the treated (ATT) is $ATT = E[TF_1 - TF_0 | FTA = 1]$.

The problem, of course, is that only one of TF_1 and TF_0 is observable. The Abadie–Imbens (A–I) matching estimator imputes the missing potential values using average outcomes for individuals with similar values for the covariates. Using notation from [Abadie and Imbens \(2006\)](#), let $\|\mathbf{x}\|_V = (\mathbf{x}'\mathbf{V}\mathbf{x})^{0.5}$ be a vector norm with \mathbf{V} a positive definite weight matrix. Let $j_m(i)$ be an index ($j=1, \dots, N$), where N

denotes the number of country pairs, which solves $FTA_j = 1 - FTA_i$, ensuring that when j has an FTA then i does not, and:

$$\sum_{l: FTA_l = 1 - FTA_i} 1\{\|\mathbf{X}_l - \mathbf{X}_i\| \leq \|\mathbf{X}_j - \mathbf{X}_i\|\} = m \quad (6)$$

where $1\{\}$ is an indicator function; the function assumes the value 1 if the bracketed expression holds and equals 0 otherwise. Then $j_m(i)$ is an index of the unit that is the m th closest unit to unit i among the covariates (among units with opposite treatment to unit i). Let $J_M(i)$ denote the set of indices for the first M matches for unit i ; in our empirical work, we will use the three nearest neighbors. Let $K_M(i)$ denote the number of times unit i is used as a match given that M matches are used per unit:

$$K_M(i) = \sum_{l=1}^N 1\{i \in J_M(l)\} \quad (7)$$

The matching estimator imputes the missing potential outcomes as:

$$TF_i^*(0) = \begin{cases} TF_i, & \text{if } FTA_i = 0 \\ \frac{1}{M} \sum_{j \in J_M(i)} TF_j, & \text{if } FTA_i = 1 \end{cases} \quad (8)$$

and

$$TF_i^*(1) = \begin{cases} \frac{1}{M} \sum_{j \in J_M(i)} TF_j, & \text{if } FTA_i = 0 \\ TF_i, & \text{if } FTA_i = 1 \end{cases} \quad (9)$$

As noted in Abadie and Imbens (2006), the estimator for the average treatment effect (ATE) is:

$$ATE_M = \frac{1}{N} \sum_{i=1}^N [TF_i^*(1) - TF_i^*(0)] \quad (10)$$

The estimator for the average treatment effect on the treated (ATT) is:

$$ATT_M = \frac{1}{N_{FTA_i=1}} \sum_{i=1}^N [TF_i - TF_i^*(0)] \quad (11)$$

If the variables in \mathbf{x} are discrete (i.e., matching is “exact”), then ATE_M and ATT_M will be unbiased estimates of the population average treatment effects. However, in our context \mathbf{x} will include some continuous variables. Consequently, both estimators will have a conditional bias in finite samples. Abadie and Imbens (2006) show that, with k continuous covariates, a term corresponding to the matching “discrepancies” will be of order $O_p(N^{-1/k})$. Abadie and Imbens (2006) propose a “bias-adjusted” estimator that adjusts for differences in covariate values within the matches, which we will provide in a sensitivity analysis.¹⁵

4. An application of matching econometrics to international trade flows and FTAs

4.1. Data description

Nominal bilateral trade flows are from the International Monetary Fund's *Direction of Trade Statistics* for the years 1960, 1965, ..., 2000 for

96 potential trading partners (zero trade flows are excluded); we will allow for zero trade flows in the sensitivity analysis. These data are scaled by exporter GDP deflators to generate real trade flows, which are also used. Nominal GDPs are from the *World Bank's World Development Indicators* (2003); these are scaled by GDP deflators to create real GDPs. Bilateral distances were compiled using the CIA *World Factbook* for longitudes and latitudes of economic centers to calculate the great circle distances. The language and adjacency dummy variables were compiled also from the CIA *World Factbook*. The FTA dummy variable was calculated using appendices in Lawrence (1996) and Frankel (1997), various websites, and FTAs notified to the GATT/WTO under GATT Articles XXIV or the Enabling Clause for developing economies; we included only full (no partial) FTAs and customs unions. A list of the trade agreements and countries used and sources is in Baier and Bergstrand (2007a).

4.2. Empirical results: average treatment effects of FTAs

Using the four covariates and the A-I matching estimator for the three nearest neighbors, we calculated the ATEs and ATTs for a sample of 96 countries for 9 years: 1960, 1965, ..., 2000. The average difference between the trade flows of the treated pairs and of their three nearest matched neighbors is an estimate of the average treatment effect on the treated. Table 5 provides the results. Columns (3)–(8) provide the baseline results. Columns (9)–(11) in Table 5 provide some sensitivity analysis to methods of estimation. Columns (12) and (13) will be addressed in detail in Section 5.

4.2.1. Baseline results for all FTAs

For convenience, columns (3) and (4) in Table 5 report the same ATE estimates reported earlier in Tables 2 and 3 using OLS and OLS with fixed effects for the traditional and theory-motivated gravity equations, respectively, for nominal bilateral trade flows for the nine years; we provide these for comparison. As noted, OLS and fixed-effects estimates vary widely across years, are often insignificant, and are significantly negative in several years.

Column (5) in Table 5 reports the ATEs estimated using the A-I matching estimator described above for nominal trade flows. Several interesting results are worth noting. First, unlike the previous numbers, for every year the ATE is positive and economically significant. This suggests our *first main result*: FTAs increase bilateral trade in every year. Second, considerable variance remains; in fact, the range of estimates is even wider than that using OLS or fixed effects. The standard deviation of OLS ATEs across the nine years is 0.46 whereas the standard deviation for the A-I ATEs is 0.62. Third, the values of some of the years' ATE estimates seem excessively “large,” i.e., even larger than the currency-union effects motivating Persson's comment on Rose (2000). For instance, in 1990, the estimated average effect of an FTA was to increase trade by *tenfold*, or 900% ($e^{2.36} = 10.59$). Fourth, the large drop in the ATE from 1995 to 2000 raises concern.

One reason the ATEs in column (5) may be excessively large is the extrapolation problem raised in Imbens (2004). ATEs provide estimates of treatment on a randomly selected pair. A second reason the ATEs in column (5) may be large is that they require two imputed values, whereas ATTs require only one imputed value, cf., Section 3.4. Column (6) provides the ATT estimates. This leads to our *second main result*: The effects of an FTA on the treated are smaller than the corresponding ATEs (although in some years the ATE and ATT estimate differences are not statistically significant). As in column (5), the ATTs are all positive, economically significant, and statistically significant from zero. The ATT estimates yield more plausible effects of an FTA on a member pair. For instance, in 1995, the A-I ATE was 2.27, implying an 867% increase in trade. By contrast, in 1995, the A-I ATT was 0.84, implying only a 132% increase in trade. The lower ATT is quite plausible economically because the ATE reflects the trade increase of a randomly selected pair and is an average of ATTs and ATUs (average

¹⁵ Abadie and Imbens (2006) show that simple matching estimators have biases related to the number of observations (N) under standard normalizations with continuous covariates. The bias is driven by “matching discrepancies” on the continuous covariates; matching on continuous variables can never be “exact.” For instance, for ATE_M in Eq. (10), the bias-adjusted estimator replaces the TF_j in Eq. (8) with $TF_j + \mu(\mathbf{X}_j, FTA=0) - \mu(\mathbf{X}_j, FTA=1)$ and the TF_i in Eq. (9) with $TF_i + \mu(\mathbf{X}_i, FTA=1) - \mu(\mathbf{X}_i, FTA=0)$ where $\mu(\cdot)$ is a consistent (regression) estimator. In our case, the number of continuous covariates is small so the bias is minimal.

Table 5
FTA treatment effects for bilateral trade flows

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Year	# of observations (# of FTAs)	Nominal level OLS ATEs	Nominal level fixed effects ATEs	Nominal level A–I ATEs	Nominal level A–I ATTs	Real level A–I ATTs	Real share A–I ATTs	Real share bias-adj. A–I ATTs	Real share P–S ATTs	Real share bias-adj. P–S ATTs	EEC real share A–I ATTs	CACM real share A–I ATTs
1960	1059 (37)	0.43* (2.02)	–0.01 (–0.06)	0.71* (1.90)	0.48** (2.16)	0.28 (1.48)	0.35** (2.11)	0.23 (1.55)	0.36 (1.23)	0.02 (0.05)	0.24 (1.03)	0.82 (1.40)
1965	1325 (42)	0.80** (4.24)	0.33** (2.00)	1.13** (2.68)	0.54** (2.92)	0.57** (3.27)	0.61** (3.24)	0.48** (2.99)	1.04** (2.89)	0.60 (1.12)	0.47* (1.80)	1.36** (2.38)
1970	1570 (41)	1.32** (6.08)	0.74** (3.86)	1.86** (3.92)	1.35** (6.16)	1.30** (6.10)	1.30** (5.75)	1.03** (5.39)	1.31** (4.08)	1.23** (2.85)	1.04** (2.71)	2.94** (2.70)
1975	1947 (86)	0.42** (2.47)	–0.10 (–0.62)	1.83** (4.18)	0.90** (4.84)	0.79** (4.29)	0.46** (2.88)	0.40** (2.73)	0.56** (2.31)	0.99** (3.20)	0.52** (1.97)	2.18** (2.32)
1980	2189 (84)	–0.11 (–0.62)	–0.78** (–4.60)	1.32** (2.51)	0.83** (5.08)	0.75** (4.77)	0.49** (3.27)	0.39** (2.94)	0.52** (2.32)	0.83** (2.90)	0.59** (2.18)	2.24** (2.54)
1985	2433 (98)	0.32* (1.98)	–0.42** (–2.74)	1.60** (3.93)	1.13** (6.88)	0.72** (5.20)	0.51** (4.20)	0.37** (3.23)	0.55** (2.78)	0.61** (2.43)	0.84** (2.43)	0.74* (1.72)
1990	2802 (129)	–0.04 (–0.27)	–0.41** (–2.72)	2.36** (4.43)	1.00** (6.85)	0.94** (6.76)	0.61** (5.47)	0.51** (5.01)	0.55** (3.27)	0.51** (2.72)	0.90** (2.20)	1.05** (2.38)
1995	3073 (163)	–0.03 (–0.27)	–0.23* (–1.79)	2.27** (5.03)	0.84** (6.67)	0.84** (6.67)	0.55** (5.20)	0.46** (4.77)	0.39** (2.49)	0.39** (2.07)	0.78** (2.09)	1.25** (2.09)
2000	3342 (263)	0.15 (1.37)	–0.08 (–0.75)	0.68** (2.34)	0.59** (5.08)	0.61** (5.20)	0.38** (3.59)	0.29** (2.88)	0.26* (1.93)	0.28 (1.38)	0.77** (2.16)	1.08** (2.18)

z-statistics are in parentheses. * (**) denotes statistical significance at the 10 (5) percent level in two-tailed tests. For 1960, 1965 and 1980, the number of (non-zero-trade) observations for real trade are 1010, 1291, and 2163, respectively. Column (3) refers to OLS results using typical gravity Eq. (1). Column (4) refers to OLS results including country fixed effects for multilateral price terms in gravity Eq. (2). Column (5) refers to ATEs using Eq. (10) for nominal trade flows. Column (6) refers to ATTs using Eq. (11) for nominal trade flows. Column (7) refers to ATTs using real trade flows. Column (8) refers to ATTs using real trade flow shares (real trade flows scaled by the product of the two countries' real GDPs). Column (9) refers to "bias-adjusted" ATTs using real trade flow shares. Column (10) refers to propensity scores using real trade flow shares. Column (11) refers to propensity scores using real trade flow shares. Columns (12) and (13) refer to ATTs for only the EEC and CACM, respectively, using real trade shares.

treatment effect on the untreated). Most country pairs without FTAs (untreated) are smaller, developing economies with much higher tariff rates, likely to have more trade created from an FTA. Consequently, the ATEs are likely to be higher, reflecting potentially large ATUs.

Our *third main result* is that the A–I ATTs have less variance across years than the A–I ATEs. The standard deviation of A–I ATTs is only 0.3, half that of A–I ATEs and two-thirds that of OLS ATEs. There is also more of a potentially explainable pattern to the average treatment effects over time, in contrast to the seemingly random pattern in the OLS and fixed-effects estimates. Although considerable variance remains, there is an economic explanation that we will discuss shortly.

Up to now, we have focused on *nominal* (bilateral) trade flows, rather than real trade flows. The reason is that most cross-sectional gravity equation analyses traditionally have used nominal trade flow and nominal GDP data. However, two reasons suggest that real trade flows and real GDPs should be examined. First, policymakers are likely to be interested more in the trade *volume* effect of an FTA, rather than the trade value effect. Second, future analyses may examine changes in trade flows, and price changes should likely be removed from the analysis. Column (7) reports the A–I ATTs using real trade flows and real GDPs. As expected, the average ATT is lower for real than nominal trade flows; inflation has been removed.¹⁶ The average ATT across all nine years for real (nominal) trade flows is 0.75 (0.85); the standard deviation of the ATTs for real (nominal) flows across the nine years are 0.29 (0.30). Thus, across all years of our sample, on average an FTA approximately *doubles* two members' real trade (i.e., $e^{0.75}=2.12$, implying a 112% increase).

Even though one of the conditioning covariates is real GDP, policymakers may be interested in the effect on the volume of trade *relative* to national outputs, denoted henceforth the *real trade share*. Here, the log of real trade share is defined as the log of the sum of the real trade flows of the two countries less the sum of the logs of their real GDPs; this is a standard scaling in the gravity equation literature (the results were robust to alternative GDP scalings). We would expect the ATTs for real trade shares to be smaller than those for real trade

levels. Column (8) confirms they are, in general. In some years, the real trade share ATTs are approximately the same as those for real trade levels; only in 1960 and 1965 were the real trade share ATTs marginally higher. The average real trade share ATT across all nine years was 0.58, implying that the real trade share was on average 79% higher with an FTA than without an FTA. There remains variance across years; the standard deviation for the ATTs in column (8) is 0.29. We have also run specifications for nominal trade shares and the results are similar (omitted for brevity).

Using the ATTs in column (8), we can now provide some explanation for the pattern of change across the nine years, taking into account phase-ins of agreements, lagged terms-of-trade effects, and the growth of regionalism. First, with the exception of 1960, 1970, and 2000, the ATT for FTA in any year is between 0.46 and 0.61. We now explain the three outliers. From 1960 to 1970, the main agreements in place were the EEC, EFTA, and CACM, which all began between 1958 and 1961. However, all of the agreements were phased-in, which partly explains that the effects of these agreements likely phased-in also over the 1960s. Moreover, terms-of-trade changes from agreements likely have a lagged effect on trade flows, contributing further to the delay in the full treatment effects from these agreements. As Table 5's column (2) shows, in 1960, 1965, and 1970, there were only 37, 42, and 41 pairs of countries with FTAs, respectively. Thus, the ATT in 1960 was lower than average because only the EEC had been in effect for more than 1 year (in fact, for only 2 years). This interpretation is confirmed in columns (12) and (13) which show that the ATTs for the EEC and for CACM were both statistically insignificant in 1960.¹⁷

The next outlier to explain is the large ATT for 1970, followed by a much lower value in 1975. While we delay a detailed explanation of the effects of the EEC and CACM until Section 5, by 1970 both agreements had been fully phased in. Moreover, note the

¹⁶ Note that in 1995 (the base year for real variables) the nominal and real ATTs are identical, as expected.

¹⁷ The economically large and marginally statistically significant (at 20%) ATT for the CACM in 1960 may seem surprising, since the treaty was only signed on December 15, 1960. However, it must be noted that after World War II, there was a wave of *bilateral* FTAs signed among Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua between 1950 and 1956, which preceded our data set, cf., U.S. Library of Congress (1993, p.2).

exceptionally large value for CACM in 1970 in column (13). While this value may seem implausible to the reader, we will discuss in Section 5 that – by historical accounts – the CACM “flourished” in the 1960s (on the heels of the Central American bilateral FTAs formed between 1950 and 1956), followed by political and civil ruptures in the subsequent two decades. The large value for the ATT for FTA in 1970 in column (8) is plausible.

But the ATT in 1975 is sharply lower than that in 1970. The addition of Denmark, Ireland, and the United Kingdom to the EEC in 1973 likely led to trade diversion from the original EEC. Moreover, the new agreements with these countries and the creation and phasing-in of FTAs between EC countries and EFTA countries between 1970 and 1975 contributed to a large expansion in the number of *new* agreements. From 1970 to 1975, the number of FTAs more than doubled from 41 to 86. As we saw in 1960 and 1965, the role of phase-ins and lagged terms-of-trade effects will cause the ATT to be much lower for new agreements. The doubling of the number of FTAs from 1970 to 1975 – and the consequent presence of brand new FTAs – had the effect of causing the ATT to decrease from its 1970 level of 1.30 to 0.46 in 1975.¹⁸

While plausible explanations such as subsequent FTA formations and the introduction of the European Monetary System can explain minor fluctuations in the level of the ATT from 1975 to 1995, the last notable change in column (8) is the decline from 0.55 to 0.38 in the ATT from 1995 to 2000. Although the number of FTAs increased from 1980 to 2000, the largest absolute and percentage increase in treated pairs in our sample in these two decades occurred between 1995 and 2000, when the number of treated pairs increased by 100 from 163 (in 1995) to 263 (in 2000). This 61% increase in the number of FTAs introduced a large number of *new* FTAs – likely all phased-in over 5-to-10 years – that probably caused the treatment effect to decline from 0.55 to 0.38.

4.2.2. Sensitivity analysis

There are three dimensions in which we examine the robustness of the estimates. First, Abadie and Imbens (2006) show that the matching estimator above will be biased in finite samples when matching is not “exact,” that is, if covariate values are not identical within matches. They show that, with k continuous covariates, a term corresponding to the matching “discrepancies” will be of order $O_p(N^{-1/k})$. Abadie and Imbens (2006) propose a “bias-adjusted” estimator that adjusts for differences in covariate values within the matches. Column (9) provides the bias-adjusted ATTs for the real trade shares. Except for one year's estimate (1970), the bias-adjusted ATTs are within 0.15 of the unadjusted ATTs in column (8). Thus, the results are largely robust to adjustment for the bias.

Second, a common alternative nonparametric estimator is the propensity-score estimator. The advantage of the propensity-score estimator is that matching is done on only one characteristic, which is a probit or logit function of several observables; this technique avoids the “curse of (high) dimensionality.” However, the reduction in dimension is at a cost of being a less precise matching procedure. Nevertheless, a useful check is to determine if the results above are largely robust to this estimator. In the propensity-score approach, one controls for all the observables by controlling for a particular function of the covariates, the conditional probability of treatment, $P[FTA=1|\mathbf{x}]$. If selection bias is absent in controlling for \mathbf{x} , it will be absent in the propensity scores (P-S). We then match on the propensity scores for the country pairs with and without FTAs. The covariates used are the same ones used for matching earlier. See Dehejia and Wahba (2002) or Cameron and Trivedi (2005) for further details. Columns (10) and (11) provide the results for real trade shares for the unadjusted and bias-adjusted P-S ATTs, respectively. The results for columns (8) and (10)

are largely similar; the results for columns (9) and (11) are largely similar. The only substantive difference between the A-I and P-S estimates are that the A-I estimates are estimated more precisely, either for the unadjusted or bias-adjusted values. In fact, every A-I estimate has a lower standard error than its corresponding P-S estimate, as expected.

Third, note that the potential sample size varies across years; this is because we only allowed non-zero trade flows for potential matched pairs. The reason is – by allowing matched pairs with zero trade – this may introduce an unobservable factor influencing potential trade that could bias results (i.e., an inappropriate extrapolation). We re-estimated the A-I results for columns (7) and (8) – real trade levels and real trade shares – allowing all 4560 pairs in each year. We summarize the results. First, the alternative ATTs were less precisely estimated; the standard errors of the estimates (z-statistics) were higher (lower) than the corresponding ATTs using the original samples. Second, the ATTs for the full sample are much more unstable across years than the original ones. Third, typically the ATTs with the full sample are considerably higher (in some cases, implausibly higher) than the baseline results.

5. Estimating the treatment effects of the Treaty of Rome and the Treaty of Managua

The main conclusion of the previous section is that the overall long-run average treatment effect of an FTA on trade between a pair of member countries is to roughly double their trade. However, as noted above, the variation across ATT estimates in Table 5 tends to display a trend that is notably absent from typical OLS estimates (Table 5, column 3). We also note that the results suggested that the average effect of an FTA tended to occur over 10–15 years, to account for lagged effects of FTAs on their terms of trade and for “phasing-in” of agreements, cf., Baier and Bergstrand (2007a) for more on such effects.

In this section, we address two issues. First, we provide *ex post* estimates of the average treatment effect of an FTA (or customs union) on the treated focusing on two specific “treaties.” By examining two particular long-standing agreements, we can explain readily most of the variation in ATTs across the nine years reported in Table 5. The only long-standing effective agreements since 1960 are the (EEC) Treaty of Rome, the (CACM) Managua Treaty, and the (EFTA) Treaty of Stockholm, so these are relevant agreements to examine and exploit our data set. In the interest of methodology and brevity, we focus on only two of the trade agreements over the period 1960–2000, the original six-member European Economic Community (treaty signed in 1957) and the original five-member Central American Common Market (treaty signed in 1960; Costa Rica joined in 1962 and Panama was excluded).¹⁹

The second issue we address is that an examination of specific treaties more likely ensures that the stable-unit-value-treatment assumption (SUTVA) is not violated. In the cases of the original six-member EEC and the original five-member CACM, the “no-multiple-versions” assumption is more likely to hold; the membership of these groups never changed and the Treaties of Rome and Managua are well defined.²⁰ Section 5.1 reports and explains the results for the EEC. However, the “non-interference” assumption may be violated in the case of the original EEC members, since EFTA was formed at approximately the same time and the EEC had subsequent enlargements to include former EFTA countries. In this case, the general equilibrium effects are consistent with standard trade diversion

¹⁹ Results for EFTA are in accord with those for the EEC, but omitted for brevity (available on request).

²⁰ Although EFTA was signed in 1960, its membership has fluctuated considerably, such that only two of the four small economies currently comprising EFTA (Norway and Switzerland) were even in the original group of seven EFTA countries. If we were to focus only on this pair, we would risk violating the overlap assumption (A2).

¹⁸ Moreover, as will be detailed in Section 5, 1969 witnessed the first political rupture in CACM with an El Salvador–Honduras war.

arguments. By contrast, the non-interference assumption is less likely to be violated for CACM. Since the CACM countries are economically small (akin to individuals in the labor economics job-training matching studies), CACM members' trade is likely to have an imperceptible impact on untreated pairs' trade. Also, since the Central American countries in CACM are remote from Europe, the EEC and EFTA agreements are unlikely to have a perceptible impact on CACM members' trade.²¹ The results for CACM are reported and explained in Section 5.2.

5.1. The original six-member European Economic Community

The EEC was formed upon the signing of the Treaty of Rome in 1957. The original six members were Belgium, Luxembourg, the Netherlands, France, (West) Germany, and Italy; the first three countries were, at the signing of the treaty, already in a free trade agreement (BENELUX). All six original members have stayed members of the European Community since its inception a half-century ago; we will address the deepening of the agreement to an economic union in the 1990s later. The EEC went into force in 1958 with a ten-year phase-in period. By 1967, tariffs on virtually all manufactured products were eliminated. Thus, taking into account the phased-in liberalizations, likely lagged effects on terms of trade of the agreement, and the absence of any expansions of the community until 1973, the likely largest effect of the EEC on members' trade should have been around 1970, as suggested in Baier and Bergstrand (2004b, 2007a).

Table 5, columns (12) and (13), provide the ATT estimates using the (bias-unadjusted) A–I real trade share matching estimator technique for EEC and for CACM, respectively. Recall, the real trade share is measured as the sum of the real trade flows between a country pair divided by the product of their real GDPs; we then logged the variable. Column (12) reports the real trade share ATTs for treated relative to untreated pairs for the original six EEC members' trade. Column (12) shows that, only 3 years into the EEC, members' trade was 27% higher ($e^{0.24}$) than their matched control pairs, suggesting a plausible 8 percent average annual increase. By 1965, EEC trade was 60% higher. By 1970, such trade was 183% higher because of phase-ins of liberalization and lagged effects on trade volume of terms-of-trade changes; 183% higher trade suggests once again an 8% average annual increase.

The European Community (or EC, as it began to be called in the 1970s) has both enlarged and deepened in degree of economic integration over time. In 1973, after 15 years of existence, the EC enlarged to include three more countries: Denmark, Ireland, and the United Kingdom. In the same year, the members of the EC and EFTA formed free trade agreements. Thus, in the period from 1973 to 1980 (and perhaps longer), the original six members of the EC (or EEC-6) should have experienced a *diversion* of trade among themselves as $BVFTA_{ijt}$ fell, and trade between the original EC members with the new EC members and with EFTA members likely expanded.

From 1970 to 1975, the ATT for the EEC-6 fell by half or 0.52, from 1.04 to 0.52, and only increased slightly to 0.59 by 1980. This result is consistent with the likely trade-diverting effects on original EEC members' trade of these expansions of free trade between the EEC-6 members with eight new countries (3 new EC members and the remaining 5 EFTA members). Using these values, the average annual decrease in EEC-6 members' trade was 4.4%.

The period 1980–1985 witnessed two policy changes within the EC. First, the EC had its second enlargement. Yet, adding Greece to the EC is not likely to have had an economically significant trade-diversion effect on the original EC members' common trade. However, the second major policy change was the entry into the European Monetary System (EMS) of all six of the original EC members. The

introduction of much smaller exchange rate variability, and consequent reduced relative price variability, is likely to have had a significant effect on increasing trade among members, as the economic literature on exchange rate variability and trade suggests, cf., Alesina et al. (2002). This effect may have dwarfed any Greece-enlargement effect. Column (12) reports that the ATT level effect rose 0.59 to 0.84.

Between 1985 and 1990, the third enlargement of the EC occurred, with the addition of Portugal and Spain in 1986. The addition of these two countries to the EC is likely to have caused little trade diversion among the original EEC-6 members, as all these countries already had free trade with Portugal. The EEC-6 level ATT actually rose slightly from 0.84 to 0.90, perhaps reflecting the lagged trade-creating effects of the EMS dominating small trade-diverting effects for EEC-6 countries of adding Spain.

The major economic policy change facing Europe between 1990 and 1995 was the political and economic liberalization of Central and Eastern Europe and the creation of free trade agreements between the EC members and Bulgaria, the Czech Republic, Hungary, Poland, Romania, and the Slovak Republic, which caused a dramatic and rapid reduction in trade barriers in these countries and fostered trade diversion from the original EEC-6 countries. In many goods, full liberalization was immediate. This liberalization of Central and Eastern European countries' trade had a material impact on the EEC level ATT in column (12), reducing it from 0.90 to 0.78.

In 1995, a fourth and even larger expansion of the EC occurred, creating the EC-15. This enlargement to include Austria, Sweden and Finland would again likely have caused a reduction in trade among the original six EEC countries. However, by 1995 the EEC-6 already had FTAs in place with these three countries. The other most notable policy development in the 1990s was the creation of the European Union (EU). The formation of the EU in 1992 deepened the degree of economic integration considerably, phased in over several years. The likely trade impact on the original EEC countries was positive, most likely not occurring until the late 1990s (1995–2000); most aspects of the EU did not take effect until January 1, 1994 or January 1, 1995. The level ATT estimate had no material change, falling slightly from 0.78 to 0.77.

The stability of the EEC ATT from 1995 to 2000 also helps to explain the fall in the ATTs in Table 5, columns (3)–(11), from 1995 to 2000. The period 1995–2000 saw an enormous increase (from 163 to 263) in the number of “treated” pairs, i.e., country pairs adopting FTAs. However, such FTAs were likely phased in (such as NAFTA and MERCOSUR) and much of their impact would have been minimal, as it was for the EEC, EFTA, and CACM in 1960. Consequently, with many country pairs having “brand new” FTAs, trade flows would not yet have adjusted substantively, and we would then expect the overall FTA ATTs in 2000 in columns (3)–(11) to be lower than those in 1995. However, we should find no such diminution in the ATT for the EEC-6, and we do not; any potential trade diversion for the EEC-6 would also not have had much effect yet.

According to our level ATT estimates, the formation of the EEC-6 and subsequent policy changes in the European Community that caused either trade diversion or more trade creation among these members left the original EEC-6 members' international trade relative to GDP on average about 116% higher by 2000 ($e^{0.77}=2.16$), or 1.9% higher per year.

5.2. The Central American Common Market

The other agreement we address in detail is the Central American Common Market (CACM), formed in 1960. Entered into force in 1961, the agreement was novel among such agreements by providing *immediate free trade* on 95% of all goods traded among El Salvador, Guatemala, Honduras, and Nicaragua; Costa Rica joined in 1962 (Panama has had only an association agreement with CACM

²¹ In last 5 years of our analysis, however, CACM members' trade may have been influenced by the formation of AFTA and MERCOSUR.

members). As for the EEC-6, the no-multiple-versions assumption is likely to hold. Yet, unlike for the EEC, the non-interference assumption is more likely to hold. First, CACM's membership was constant (*de jure*) from 1962 to 2000, so there were no enlargements and consequently potential trade diversions. The small size of trade among CACM members ensures imperceptible impact on untreated pairs' trade. CACM countries' remoteness – from Europe – ensures imperceptible impact on CACM trade of the major regional FTAs formed and altered between 1960 and 1990 (EEC and EFTA).

According to SICE (2006), CACM “flourished as the most advanced and successful regional integration scheme of Latin America in the 1960s.” Intra-regional (CACM) trade grew from \$33 million in 1960 to \$1000 million in 1980; that is, intra-regional trade grew by a factor of 30. Unfortunately, in 1969 a war between El Salvador and Honduras led to Honduras' (*de facto*) withdrawal and re-imposition of tariffs on imports from the other Central American countries. From 1970 to 1990, political unrest in the region left CACM moribund according to most observers. Nicaragua instituted the Marxist Sandanista region in 1979, which continued until 1990; during the 1980s, anti-Sandanistas were prevalent in Honduras, leading to armed cross-border conflicts in the region. In June 1990, a summit of the Central American countries' presidents called for the revival of the CACM.

The ATT estimates in column (13) of Table 5 confirm the rapid success of the CACM in the 1960s (as well as bilateral FTAs among these countries between 1950 and 1956), followed by much lower effects in the 1970s and 1980s, and finally a resurgence of effectiveness in 1990. In column (13), the real trade share ATT estimates indicate that by 1970 the CACM increased members trade by 1800% ($e^{2.94} = 18.92$). While such a figure may appear unrealistically high, one should be reminded that tariffs in Latin American economies were considerably higher than those in Northern Hemisphere countries (typically, 200% higher) in 1960; estimates put them at approximately 40%. Thus immediate and nearly full trade liberalization from high tariff levels under CACM could conceivably have had such a substantial effect, especially on the heels of the countries' bilateral FTAs formed between 1950 and 1956 (a period for which bilateral trade data are unavailable).

From 1970 to 1990, political instability, war, and the resurgence of protectionism in the region explains why CACM was less effective in that period. From 1970 to 1980 (following the 1969 El Salvador–Honduras war and the withdrawal of Honduras from CACM), the real trade share level ATT fell from 2.94 to 2.24, or a 50% decline in the effect of CACM. From 1980 to 1985 (following the 1979 ascendancy of the Marxist Sandanista regime and the Honduras–Nicaraguan border war), the CACM level ATT declined again from 2.24 to 0.74, a further 78% decline in the trade effect. Following a June 1990 summit of Central American presidents to revive the Treaty of Managua, a reversal in the trade impact of CACM occurred. By 2000, the CACM ATT on the level of the real trade share was higher at 1.08; thus, by 2000, the effect of CACM had been to roughly triple members' trade.

On net, the estimated ATTs for CACM provide economically plausible measures of the effects on trade of this agreement. The direction of changes in the effects are consistent with anecdotal evidence of its initial success, followed by a rapid dissolution of its effectiveness due to political and military conflict, and finally a resuscitation in the 1990s.

6. Conclusions

This paper has provided the first nonparametric empirical estimates using matching econometrics of the long-run effects of FTAs (and customs unions) on members' trade for a wide range of such agreements, as well as for two specific agreements that were formed in the period 1957–1960. Matching estimators provide plausible estimates of the average treatment effect of an FTA on the trade of members that actually form one (ATTs), avoiding selection bias on

observables in the presence of potential nonlinearities between the FTA variable and other covariates. We find a narrower range (across years) and more economically plausible values of the long-run effects of FTAs on members' trade using nonparametric estimates than parametric ones in cross-section. Interestingly, our results using nonparametric techniques suggest an average long-run effect of an FTA of 100%. Such an estimate is not far from those found using panel data and parametric techniques, such as the gravity equation, cf., Baier and Bergstrand (2007a) and references therein. Thus, our nonparametric estimates suggest that the gravity equation may still provide a baseline parametric framework for evaluating the effects of the FTAs on trade flows, but continued investigation is warranted.

We note that our matching estimates do not address explicitly selection bias on unobservables. Alternative nonparametric methods exist, such as “difference-in-differences” techniques, to eliminate such bias. But the construction and computation of these effects using pre- and post-treatments for 15-year balanced periods for treated and control pairs is a large endeavor beyond the scope of this particular paper, and is left for future research. Nevertheless, the results here provide a further step toward providing more precise and credible *ex post* estimates of long-run effects of FTAs on trade flows. We hope future work will continue to develop better matching estimates of FTAs and other trade policies.

Acknowledgements

The authors are grateful for excellent comments from Dan Trefler, Peter Egger, Daniel Millimet, and two anonymous referees and for financial support from the National Science Foundation under grants SES 0351018 (Baier) and SES 0351154 (Bergstrand).

References

- Abadie, Alberto, Imbens, Guido W., 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74 (1), 235–267 January.
- Aitken, Norman D., 1973. The effect of the EEC and EFTA on European Trade: a temporal cross-section analysis. *American Economic Review* 63 (5), 881–892 December.
- Alesina, Alberto, Barro, Robert, 2002. Currency unions. *Quarterly Journal of Economics* 117 (2), 409–436 May.
- Alesina, Alberto, Barro, Robert, Tenreiro, Silvana, 2002. Optimal currency areas. In: Gertler, Mark, Rogoff, Kenneth (Eds.), *NBER Macroeconomics Annual 2002*. MIT Press, Cambridge, MA.
- Anderson, James E., 1979. A theoretical foundation for the gravity equation. *American Economic Review* 69 (1), 106–116 March.
- Anderson, James E., van Wincoop, Eric, 2003. Gravity with gravitas: a solution to the border puzzle. *American Economic Review* 93 (1), 170–192 March.
- Anderson, James E., van Wincoop, Eric, 2004. Trade costs. *Journal of Economic Literature* 62 (3), 691–751 September.
- Baier, Scott L., Bergstrand, Jeffrey H., 2001. The growth of world trade: tariffs, transport costs, and income similarity. *Journal of International Economics* 53 (1), 1–27 February.
- Baier, Scott L., Bergstrand, Jeffrey H., 2002. On the Endogeneity of International Trade Flows and Free Trade Agreements. manuscript, http://www.nd.edu/~jbergstr/Working_Papers/EndogeneityAug2002.pdf.
- Baier, Scott L., Bergstrand, Jeffrey H., 2004a. Economic determinants of free trade agreements. *Journal of International Economics* 64 (1), 29–63 October.
- Baier, Scott L., Bergstrand, Jeffrey H., 2004b. Do Free Trade Agreements Actually Increase Members' International Trade? manuscript.
- Baier, Scott L., Bergstrand, Jeffrey H., 2006. Bonus Vetus OLS: A Simple Method for Addressing the 'Border Puzzle' and Other Gravity-Equation Issues. manuscript, http://www.nd.edu/~jbergstr/Working_Papers/BVOLSMarch2006.pdf.
- Baier, Scott L., Bergstrand, Jeffrey H., 2007a. Do free trade agreements actually increase members' international trade? *Journal of International Economics* 71 (1), 72–95 March.
- Baier, Scott L., Bergstrand, Jeffrey H., 2007b. Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects using the Gravity Equation. manuscript, http://www.nd.edu/~jbergstr/Working_Papers/BVOLSOctober2007.pdf.
- Baier, Scott L., Bergstrand, Jeffrey H., Vidal, Erika, 2007. Free trade agreements in the Americas: are the trade effects larger than anticipated? *The World Economy* 30 (9), 1347–1377 September.
- Baier, Scott L., Bergstrand, Jeffrey H., Egger, Peter, McLaughlin, Patrick A., 2008. Do economic integration agreements actually work? Issues in understanding the causes and consequences of the growth of regionalism. *The World Economy* 31 (4), 461–497 April.
- Baldwin, Richard E., 1994. *Towards an Integrated Europe*. CEPR, London.
- Bergstrand, Jeffrey H., 1985. The gravity equation in international trade: some microeconomic foundations and empirical evidence. *Review of Economics and Statistics* 67 (3), 474–481 August.

- Bergstrand, Jeffrey H., Egger, Peter, Larch, Mario, 2007. Gravity Redux: Structural Estimation of Gravity Equations with Asymmetric Bilateral Trade Costs. manuscript, http://www.nd.edu/~jbergstr/Working_Papers/GravityReduxOctober2007.pdf.
- Brada, Josef, Mendez, Jose, 1985. Economic integration among developed, developing, and centrally planned economies: a comparative analysis. *Review of Economics and Statistics* 67 (4), 549–556.
- Cameron, A. Colin, Trivedi, Pravin K., 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge, UK.
- Chintrakarn, Pandej, 2008. Estimating the Euro effects on trade with propensity score matching. *Review of International Economics* 16 (1), 196–198 February.
- Dehejia, Rajeev H., Wahba, Sadek, 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* 84 (1), 151–161 February.
- Diamond, Alexis, 2006a. Thoughts on SUTVA, Part I. http://www.iq.harvard.edu/blog/sss/archives/2006/02/thoughts_on_sut.shtml.
- Diamond, Alexis, 2006b. Thoughts on SUTVA, Part II. http://www.iq.harvard.edu/blog/sss/archives/2006/03/thoughts_on_sut_1.shtml.
- Eaton, Jonathan, Kortum, Samuel, 2002. Technology, geography, and trade. *Econometrica* 70 (5), 1741–1779 September.
- Egger, Hartmut, Egger, Peter, Greenaway, David, 2008. The trade structure effects of endogenous regional trade agreements. *Journal of International Economics* 74, 278–298.
- Evenett, Simon, Keller, Wolfgang, 2002. AON theories explaining the success of the gravity equation. *Journal of Political Economy* 110 (2), 281–316 April.
- Feenstra, Robert, 2004. *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton, NJ.
- Frankel, Jeffrey A., 1997. *Regional Trading Blocs*. Institute for International Economics, Washington, DC.
- Ghosh, Sucharita, Yamarik, Steven, 2004. Are regional trading arrangements trade creating?: an application of extreme bounds analysis. *Journal of International Economics* 63 (2), 369–395 July.
- Heckman, James, Lochner, Lance, Taber, Christopher, 1998. General-equilibrium treatment effects: a study of tuition policy. *American Economic Review* 88 (2), 381–386 May.
- Henderson, Daniel J., Millimet, Daniel M., 2008. Is gravity linear? *Journal of Applied Econometrics* 23 (2), 137–172.
- Imbens, Guido W., 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics* 86 (1), 4–29 February.
- Lawrence, Robert Z., 1996. *Regionalism, Multilateralism, and Deeper Integration*. The Brookings Institution, Washington, DC.
- Kenen, Peter B., 2002. Currency unions and trade: variations on themes by Rose and Persson. Reserve Bank of New Zealand Discussion Paper DP/2002/08, December.
- Lee, Myoung Jae, 2005. *Micro-econometrics for Policy, Program, and Treatment Effects*. Oxford University Press, Oxford, UK.
- Linnemann, Hans, 1966. *An Econometric Study of International Trade Flows*. North-Holland, Amsterdam.
- Moffitt, Robert A., 2004. Introduction to the symposium on the econometrics of matching. *Review of Economics and Statistics* 86 (1), 1–3 February.
- Persson, Torsten, 2001. Currency unions and trade: how large is the treatment effect? *Economic Policy* 435–448 October.
- Rose, Andrew K., 2000. One money, one market: the effect of common currencies on trade. *Economic Policy* 9–45 April.
- Rose, Andrew K., 2004. Do we really know that the WTO increases trade? *American Economic Review* 94 (1), 98–114 March.
- Rosenbaum, Paul R., Rubin, Donald B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1), 41–55.
- Santos Silva, J.M.C., Tenreiro, Silvana, 2006. The log of gravity. *Review of Economics and Statistics* 88 (4), 641–658 November.
- Sapir, Andre, 1981. Trade benefits under the EEC generalized system of preferences. *European Economic Review* 15, 339–355.
- Sapir, Andre, 2001. Domino effects in Western European regional trade. *European Journal of Political Economy* 17, 377–388.
- Schott, Jeffrey, 2005. *Free Trade Agreements: U.S. Strategies and Priorities*. Institute for International Economics, Washington, DC.
- SICE, 2006. <http://www.sice.oas.org>.
- Tenreiro, Silvana, in press. "On the Causes and Consequences of Currency Unions." Harvard University unpublished working paper.
- Trefler, Daniel, 1993. Trade liberalization and the theory of endogenous protection: an econometric study of U.S. import policy. *Journal of Political Economy* 101, 138–160 February.
- The World Bank, 2003. *World Development Indicators 2003*. The World Bank, Washington, DC.