CGPA WORKING PAPER
2025-03

The European Single Market and Intra-EU Trade: An Assessment with Heterogeneity-Robust Difference-in-Differences Methods

Arne J. Nagengast
Fernando Rios-Avila
Yoto V. Yotov

CENTER FOR GLOBAL POLICY ANALYSIS
LeBow College of Business
Drexel University

The European Single Market and Intra-EU Trade: An Assessment with Heterogeneity-Robust Difference-in-Differences Methods*

Arne J. Nagengast
Deutsche Bundesbank

Fernando Rios-Avila
Levy Economics Institute

Yoto V. Yotov
Drexel University

January 2, 2025

Abstract

We use heterogeneity-robust difference-in-differences (DiD) methods to evaluate the impact of membership in the European Union (EU) Single Market on international trade. On the policy front, we provide evidence that: (i) On average, the EU has been very effective in promoting trade among its member states; (ii) The trade effects of the EU have been long-lasting, but heterogeneous across EU cohorts; and (iii) While the EU has benefited both 'old' and 'new' members, the increase in the exports from the 'old' members to the 'new' joiners has been disproportionately larger. From a methods and practical perspective, the contribution of this paper is to introduce a new, fast, and flexible estimation command that combines leading estimation techniques from the gravity literature with recent methods from the heterogeneity-robust DiD literature.

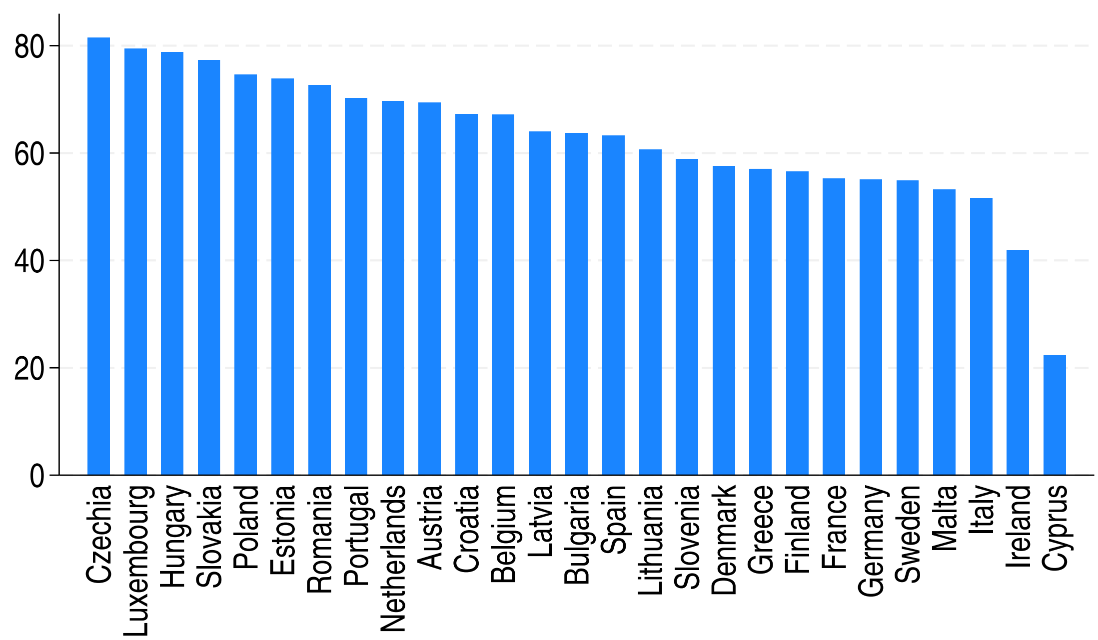
JEL classification: C13, C23, F10, F13, F14.
Keywords: EU membership, Staggered Difference-in-Differences, Gravity Model, Estimation Command

*The views expressed in this paper are those of the author(s) and do not necessarily coincide with the views of the Deutsche Bundesbank or the Eurosystem. Contact information: Nagengast: Deutsche Bundesbank, E-mail: arne.nagengast@bundesbank.de; Rios-Avila: The Levy Economics Institute, Bard College, E-mail: friosavi@levy.org; Yotov: School of Economics, Drexel University; ifo Institute; CESifo, E-mail: yotov@drexel.edu.

1 Introduction

On May 1, 2024, the European Union will celebrate the 20th anniversary of its largest enlargement in terms of number of states and population.[1] On the eve of this anniversary, the European Union is considered the most prominent and successful international integration effort in the world. "It is also the world's largest single market area, and free trade among its members was one of the EU's founding principles."[2] Thus, perhaps not surprisingly, the most important trade partners of most EU members are other EU states. Figure 1 captures

Figure 1: Intra-EU Exports of Goods in 2023 (% of Total)



Notes: The figure reports the share of intra-EU exports of goods in 2023 as % of total exports. The data comes from Eurostat at https://ec.europa.eu/eurostat/databrowser/bookmark/49733444-90e5-483b-b914-de554b5ca12f?lang=en.

two notable patterns of intra-EU trade of goods in 2022. First, more then 50% of the exports for 25 of the 27 EU members are directed to other EU members. The exceptions are Cyprus and Ireland. Second, the country-specific shares of intra-EU trade are very heterogeneous,

[1]The ten countries, which joined the EU in 2004 were Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia. See Figure 2 for details.

[2]https://european-union.europa.eu/priorities-and-actions/actions-topic/trade_en.

ranging from 22% and 42% for Cyprus and Ireland, respectively, to 81% and 79% for Czechia and Luxembourg.

While few would doubt the significant role of the EU Single Market to promote trade among its members, the following two, more nuanced, questions should be of interest to policy makers and academics alike: (i) What fraction of the large shares of intra-EU trade are actually due to joining the Single Market as opposed to other forces, e.g., geographic and cultural proximity? and (ii) Has the Single Market succeeded in promoting intra-EU trade uniformly or did some members enjoy larger gains from free trade within the EU? Answering these questions is important for ex-post evaluation of the success of EU trade integration but also for ex-ante evaluation of the potential implications of further enlargement of the EU. Moreover, obtaining reliable partial estimates of the effects of the Single Market on trade is a crucial prerequisite for estimating sound general equilibrium and welfare implications of joining or leaving the EU (Dhingra et al., 2017).

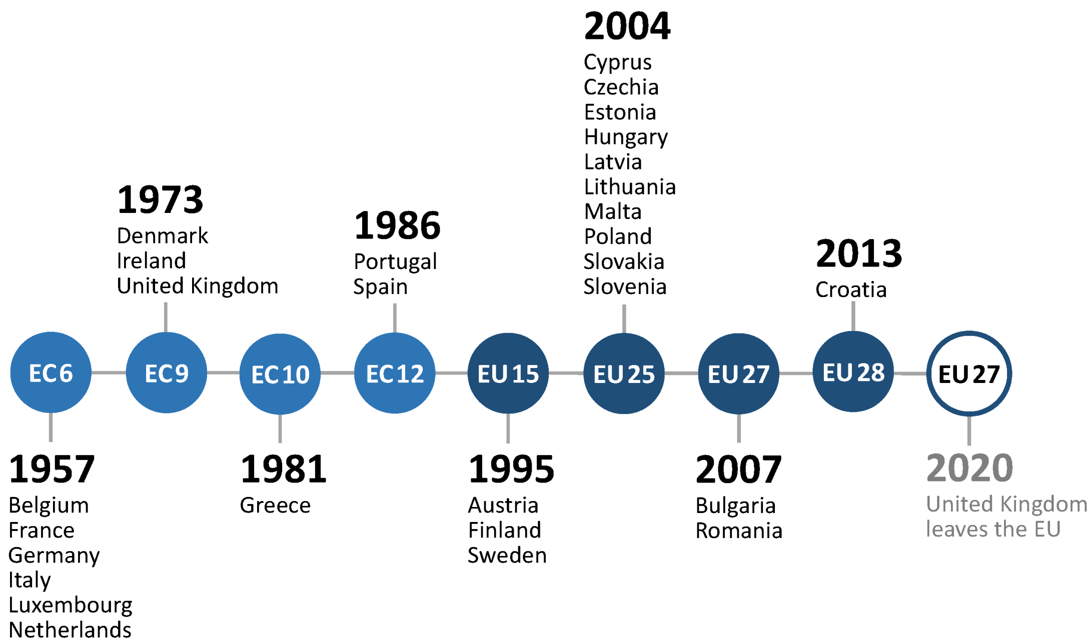
Despite the economic and political importance of the EU as the largest free trade area in the world, there are surprisingly few studies that have evaluated the impact of EU membership on intra-EU trade (e.g., Hamilton and Winters, 1992; Fontagné et al., 1998; Head and Mayer, 2000; Papazoglou et al., 2006; Iliev et al., 2016). Moreover, all studies of the impact of the EU that we reviewed relied exclusively on traditional 'gravity-type' methods,[3] and we are not aware of any investigation that accounts for possible biases in the presence of treatment effect heterogeneity, e.g., due to so-called 'forbidden comparisons' that (mis)use already-treated units in the control group (e.g., Borusyak and Jaravel, 2017; de Chaisemartin and D'Haultfœuille, 2020).

This caveat may be particularly important for the evaluation of the effects of EU membership, which is a prominent example of a staggered policy adoption design, where new members have joined in waves of accession that, as depicted in Figure 2, also vary in the

[3]Consistent with the recent econometrics DiD literature (e.g., de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Wooldridge, 2021; Borusyak et al., forthcoming), we will refer to these traditional methods as *"two-way-fixed-effects (TWFE) estimators"*.

number of new members, e.g., the large 2004 wave vs. the accession of Bulgaria and Romania in 2007 vs. the accession of Greece in 1981 or Croatia in 2013.

Figure 2: EU Accession Timeline



Notes: The figure traces the accession timeline for new European Union members. EC denotes the European Community, which was the predecessor of the European Union (EU). The layout and style of the figure is based on the corresponding figure from the German Federal Statistical Office at https://www.destatis.de/Europa/EN/Country/EU-Member-States/_EU_EZ_Zeitverlauf_en.html.

Against this backdrop, the contribution of our paper is twofold. First, from a policy and academic perspective, we combine established methods from the structural gravity literature with recent heterogeneity-robust difference-in-differences (DiD) methods to re-evaluate the impact of membership in the EU Single Market on international trade, and to explore the heterogeneity of the EU effects across various dimensions. Second, on the methods and practical front, our contribution is to introduce a new, fast, and flexible estimation command that combines leading estimation techniques from the gravity literature with recent methods from the heterogeneity-robust DiD literature and fast computation approaches to handle high-dimensional fixed effects in linear and non-linear econometric models.

To build our command, we capitalize on developments from three influential strands of the literature. First, we take into account many recent contributions from the vibrant difference-in-differences econometrics literature (e.g., Hull, 2018; de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Wooldridge, 2021; Goldsmith-Pinkham et al., 2022; de Chaisemartin and D'Haultfœuille, 2023; Borusyak et al., forthcoming; Wooldridge, 2023). Second, we take advantage of the recent developments for fast computation and convergence of linear and non-linear models with high-dimensional fixed effects (e.g., Correia, 2016b; Correia et al., 2020). Third, we rely on the empirical and theoretical trade gravity literature (e.g., Eaton and Kortum, 2002; Anderson and van Wincoop, 2003; Arkolakis et al., 2012; Yotov et al., 2016; Nagengast and Yotov, 2023) for guidance and recommendations for some of the main options of the command.

The result is a fast and flexible estimation command for heterogeneity-robust difference-in-differences estimations – `jwdid`.[4] Our current analysis focuses on the effects of EU membership on international trade. However, the proposed command is much more general. Naturally, it can be applied to quantify the impact of other determinants of trade flows within the same trade-gravity setting (e.g., WTO membership, Currency unions, etc.). Moreover, it can be used to estimate gravity equations for other bilateral flows and relationships (e.g., migration, FDI, patents, international M&As, etc.). Finally, the command can also be used in more general (non-gravity) panel and cross-sectional settings (e.g., Liebersohn (2024) and Ahmadi et al. (2024)).

To perform the empirical analysis, we rely on several standard trade data sources. First, we secure a maximum number of observations over a long period of time by using a combination of the *Direction of Trade Statistics* (DoTS) of the International Monetary Fund (IMF) and the United Nations' *Commodity Trade Statistics* (COMTRADE) databases. The resulting estimating sample covers aggregate trade flows for 260 countries over the period 1950-2019. The long time coverage of these data enables us to obtain estimates of the EU effects for recent members as well as the founding countries.

[4] To install the `jwdid` command from SSC use: ssc install jwdid, replace. To access the most uptodate version, use: net install jwdid, replace from(http://friosavila.github.io/stpackages).

In addition to the aggregate trade data, we also employ the *Structural Gravity Database* (SGD) of the World Trade Organization (WTO), which covers manufacturing trade for 229 countries over the period 1980-2016. An advantage of the SGD sample is that it complements our aggregate estimates with sectoral results for manufacturing, which is by far the most traded sector within the EU.[5] Another advantage of the SGD is that it includes consistently constructed domestic trade flows, which may be important for quantifying the impact of trade policies (Yotov, 2022).

We draw several conclusions based on the empirical analysis. First, we find that membership in the EU has led to a significant increase in bilateral trade flows among its member states. Specifically, our preferred estimate implies that membership in the EU has led to an increase of 52% in the bilateral trade flows between its member states. From a methodological perspective, our preferred ETWFE estimate is significantly larger than the corresponding TWFE result that is obtained with traditional gravity methods. We also note that both the ETWFE and TWFE EU estimates that we obtain are larger than recent estimates of the effects of trade agreements, thus reinforcing the perception of the EU as one of the most successful free trade areas in the world. Finally, an event-type analysis reveals that the effects of the EU on members' trade are also significantly more long-lasting as compared to the corresponding effects of trade agreements from the related literature.

The average EU estimate that we obtain is masking significant heterogeneity across several dimensions. For example, we see significant differences in the EU effects on exports across cohorts. Specifically, with the exception of the 1981 cohort, the effects for the early cohorts (1957, 1973 and 1986) are larger than those for the late cohorts (1995, 2004, 2007, and 2013). We also observe heterogeneity in the evolution of the EU effects across cohorts. For example, the effects for the 1973 and 1986 cohorts are large and are exhausted relatively fast, while the effects for the 1995 cohort are smaller, but more long-lasting. Finally, we find that the effects of the EU have been disproportionately stronger for the exports from 'old' members

[5] According to Eurostat, "Four fifths of total exports of goods within the EU in 2022 were manufactured products." (Eurostat, 2023).

to 'new' members than in the opposite direction, while the effects between 'new' members have been almost as strong as those between 'old' and 'new' members.

Our overall conclusion is that the effects of the EU have been very strong on average and that the EU has benefited all members, however, the EU effects have been very heterogeneous. In addition, our asymmetric estimates point to possible areas for policy intervention.

We also test and confirm the robustness of our results and main conclusions to different datasets, the degree of heterogeneity of the ETWFE model and alternative estimators, as well as DiD- and ETWFE-specific assumptions. First, we consider the WTO SGD as an alternative dataset, which includes domestic trade flows and which focuses on manufacturing trade, and we also perform several robustness checks related to the construction of our baseline dataset. Second, we consider restricting the heterogeneity of the ETWFE model, but we also use an imputation estimator that is more flexible. In addition, we consider a jackknife bias correction for potential incidental parameter problems, alternative standard error clusterings, and Ordinary Least Squares (OLS) as an alternative estimator. Third, we consider the effect of methodological choices related to the definition of treatment onset, the potential effect of the presence of other treatments, the exclusion of long-term effects, a variant relaxing the parallel trends assumption, and alternative weighting schemes for obtaining the aggregate treatment effect coefficient. Although the size of the coefficient varies slightly in some cases, all robustness tests leave the estimates of EU membership on trade by and large unchanged in terms of sign and significance.

The rest of the paper is organized as follows. Section 2 discusses the current methods to estimate trade gravity models and draws implications and recommendations for the key features of our estimation command. Section 3 introduces the new command by describing its key building blocks and main features. Section 4 offers a brief description of our data, the data sources, and some key summary statistics. Section 5 presents our main findings and the results from a series of validity checks and sensitivity experiments. Section 6 describes the results from a series of robustness experiments. Section 7 concludes. Finally, a Supplementary

Appendix includes a detailed description of the new command including all of its options and some additional estimation results.

2 Estimating the effects of the EU on trade

The objective of this section is to review the current and established techniques and developments from the empirical trade literature for estimating the effects of the EU (and other policies) on international trade and to draw implications for the new estimation command, which we introduce in the next section. To this end, we rely on the current recommendations from the structural gravity and the staggered difference-in-differences literature, as recently summarized by Nagengast and Yotov (2023), and we specify the following *structural*[6] estimating gravity equation that is used to obtain our estimate of the effects of EU membership on international trade:

$$X_{ij,t} = \exp[\phi_{i,t} + \psi_{j,t} + \overrightarrow{\gamma_{ij}} + \sum_{g=q}^{T}\sum_{s=g}^{T}\delta_{gs}D_{gs}] + \epsilon_{ij,t}. \quad \forall i, j, t. \tag{1}$$

The dependent variable in equation (1), $X_{ij,t}$, denotes nominal exports (at delivered prices) from exporter i to importer j at time t.[7] Consistent with the staggered difference-in-differences methods, and due to availability of better data, most of the recent gravity applications are performed in panel settings. In addition, we follow the recommendations of Egger et al. (2022) to estimate our model with data for consecutive years. Historically, in response to the critique of Cheng and Wall (2005, p. 52, fn. 8) that "[f]ixed-effects estimation

[6]We use the term 'structural' because gravity equation (1) is representative of many influential theoretical models of international trade, e.g., Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Arkolakis et al. (2012). However, we also note that the estimation techniques that we discuss and implement in this paper are much broader and they apply to a wide range of gravity models for different types of bilateral trade flows, e.g., migration, FDI, patents, international M&As, etc.

[7]Consistent with theory, $X_{ij,t}$ may also include domestic trade flows. As summarized by Yotov (2022), using domestic trade flows in gravity regressions has significant implications for estimating the effects of various country-specific as well as bilateral policies, e.g., the EU. The intuition is that one of the main channels through which the EU has promoted international trade is at the expense of domestic sales. For robustness, we consider estimation results for the WTO SGD in Section 6.1, and our results are consistent with the existing literature.

is sometimes criticized when applied to data pooled over consecutive years on the grounds that dependent and independent variables cannot fully adjust in a single year's time", many gravity applications (e.g., Baier and Bergstrand (2007) and Olivero and Yotov (2012) among others) have used interval data instead of data for consecutive years. However, as discussed in Egger et al. (2022), the use of intervals instead of consecutive years requires dropping data randomly, and may prevent estimation of the effects of some policies, especially when the focus is on their heterogeneous effects, which often are identified off relatively few observations. The implication of using consecutive-year data for our new command is that it should be able to handle 'long' panel data and the corresponding fixed effects.

Due to the influential work of Santos Silva and Tenreyro (2006), the Poisson pseudo maximum likelihood (PPML) estimator has established itself as the leading gravity estimator. As originally motivated by Santos Silva and Tenreyro (2006), the two main advantages of the PPML estimator for gravity estimations are that (i) it can handle heteroskedasticity in the trade data and that (ii) due to its multiplicative form, it can handle zero trade flows. More recently, several papers document additional favorable properties of the PPML estimator for gravity estimations, which have reinforced its use as the leading estimator for gravity estimations. An important implication for our command is that it has to allow for PPML, while also providing the option for the estimation to be performed by OLS (and other estimators) for robustness.

Turning to the covariates in our model, equation (1) includes three sets of fixed effects. The theoretical motivation behind the exporter-time and importer-time fixed effects ($\phi_{i,t}$ and $\psi_{j,t}$, respectively) is that they control for the structural multilateral resistances (MR) terms of Anderson and van Wincoop (2003), whose omission has been labeled *'the gold medal mistake'* in gravity regressions by Baldwin and Taglioni (2006).[8] In addition to the MR terms, and from a broader econometric perspective, the exporter-time and importer-time fixed effects

[8] Intuitively, the MRs capture the fact that trade between two countries depends not only on their sizes and the bilateral trade costs between them but also on how remote (geographically and economically) these countries are from the rest of the world. Yotov et al. (2016) summarize several attractive features of the structural multilateral resistance terms.

account for any other country-specific characteristic, e.g., country-size, that may impact bilateral trade flows. The implication for our command is that gravity should be estimated with exporter-time and importer-time fixed effects.[9]

The third set of fixed effects, $\overrightarrow{\gamma_{ij}}$, includes directional-country-pair fixed effects. This follows directly from the difference-in-differences specification under consideration, however, there are also several appealing reasons to include those fixed effects in gravity regressions. First, as demonstrated by Egger and Nigai (2015) and Agnosteva et al. (2019), the pair fixed effects match the bilateral trade costs in the gravity model much better than the 'standard gravity variables' (e.g., distance, contiguity, etc.). The latter do well in predicting relative trade costs, but they fail to capture the level of bilateral trade costs. Specifically, they under-predict the bilateral trade costs for the poor countries and over-predict them for the more developed countries. Second, on a related note and following Wooldridge (2010), Baier and Bergstrand (2007) argue that the country-pair fixed effects control for and absorb most of the unobserved correlation between the potentially endogenous bilateral policy variables in the error term in gravity models, thus mitigating endogeneity concerns. Finally, in order to allow for asymmetric bilateral trade costs and asymmetric trade policy effects, Baier et al. (2019) argue that the pair fixed effects should be directional, i.e., ij vs. ji.

The use of asymmetric country-pair fixed effects has two implications for our command. First, the command should be able to handle a very large set of fixed effects, e.g., with a balanced dataset for 100 countries, the number of relevant country-pair fixed effects is 9,900.[10] Second, as demonstrated by Weidner and Zylkin (2021), PPML estimations with exporter-time, importer-time, and country-pair fixed effects may suffer from the incidental parameter problem (IPP).

[9] Gravity model with exporter-time and importer-time fixed effects are often referred to as two-way gravity models. Similarly, those with exporter-time, importer-time, and exporter-importer fixed effects are usually referred to as three-way gravity models. To avoid potential confusions with the term two-way fixed effects models (i.e., with unit and time fixed effects) from the DiD literature, we refrain from using these terms in this paper.

[10] Some of the most recent gravity datasets, e.g., the International Trade and production Database for Estimation (ITPDE) of the U.S. International Trade Commission, include more than 200 countries.

Weidner and Zylkin (2021) propose an adjustment based on Taylor expansions to correct for such biases. Relatedly, Jochmans (2017) uses a quasi-differencing approach and Pfaffermayr (2021) implements jackknife and bootstrap confidence interval methods. Following Weidner and Zylkin (2021), Nagengast and Yotov (2023) implement a Monte Carlo simulation to study the IPP bias in a staggered DiD gravity setting, and they find that, given the relatively large time dimension of their data, IPP is not be such a big issue for the coefficient estimates, however, the standard errors could be downward biased. To address potential IPP concerns, Nagengast and Yotov (2023) follow Dhaene and Jochmans (2015), Pfaffermayr (2021), and Weidner and Zylkin (2021) to implement a split-panel jackknife bias-correction procedure.

In general, equation (1) and the command also allow the inclusion of additional time-varying bilateral control variables. While this is a common practice in the gravity literature, from an econometric perspective, additional time-varying variables may only be included if they are unaffected by the treatment and strictly exogenous. For example, in the presence of domestic trade flows, and following Bergstrand et al. (2015), the specification should also include a set of time-varying border variables, which take a value of one for international trade and are equal to zero otherwise. These time-varying border variables would capture common (across countries) globalization trends, e.g., improvements in communication and transportation technologies. In principle, other treatments such as tariffs, free trade agreements, economic sanctions, etc. could also be included in the specification. However, in general, the current estimation framework is not designed to handle the additional complications introduced by the presence of multiple treatments or continuous treatments such that it not advisable to include them.[11] An alternative approach to testing the robustness of the results to the presence of other treatments that may potentially confound the main effect of interest is to omit conspicuous episodes from the analysis. We demonstrate this approach in

[11] See, for example, de Chaisemartin and D'Haultfœuille (2023) for an approach to estimate heterogeneous treatment effects in the presence of multiple treatments in the linear setting. Similarly, see de Chaisemartin et al. (2023) and Callaway et al. (2024a,b) for estimation methods for continuous treatments in the linear setting.

Section 6.3.

Finally, and most important for our purposes, the key term ($\sum_{g=q}^{T}\sum_{s=g}^{T}\delta_{gs}D_{gs}$) that capture the effects of EU membership on trade in our model, is motivated by the staggered difference-in-differences literature. Consistent with most existing studies, in our main analysis, we focus on the impact of the EU on trade between member countries, i.e., a country pair ij belongs to treatment cohort g if the condition that both countries were EU members was for the first time fulfilled in year g, q is the first year of the treatment of cohort g, T is the last year of the panel, D_{gs} is a time-varying treatment indicator equal to 1 for cohort g for $s = t$ in post-treatment years and 0 otherwise, and δ_{gs} captures the cohort-year specific treatment effects. For example, $D_{2004,2004}$ would be equal to one for the country pair Austria-Poland in the year 2004, since Poland joined the EU in 2004 and Austria had already been an EU member since 1995. $\delta_{2004,2004}$ would capture the EU effect on trade between all country pairs of the 2004 cohort in the first year. $\delta_{2004,2005}$ would capture the EU effect on trade between all country pairs of the 2004 cohort in the second year, etc.

In addition to obtaining a single/common EU estimate, which is the most common practice in the existing literature, we also obtain a series of other EU estimates across various dimensions that have been of interest to scholars and policy makers. Specifically, we explore the heterogeneity of our estimates by event time, by cohort, and we also obtain directional estimates on trade between 'old' and 'new' EU members depending on the recency of their EU membership.

While there is no universally accepted best practice for clustering the standard errors in gravity models, Egger and Tarlea (2015), Pfaffermayr (2019), and Pfaffermayr (2022) demonstrate that multi-level clustering could make a big difference. Accordingly, we experiment with two alternative clusterings. Specifically, to obtain our main results, we follow most of the existing literature and we cluster the standard errors by country pair (i.e., $Cov[\epsilon_{ij,t}, \epsilon_{ij,d}] \neq 0$, for all t, d, and zero else). In addition, we also obtain estimates with 3-way clustering (by exporter, importer, and year, i.e., $Cov[\epsilon_{ij,t}, \epsilon_{il,d}] \neq 0, Cov[\epsilon_{ij,t}, \epsilon_{kj,d}] \neq$

$0, Cov[\epsilon_{ij,t}, \epsilon_{kl,t}] \neq 0$, and zero else). The implication for our command is that it should include an option for alternative clusterings.

Before we continue, we note that our econometric model (1) can be estimated at any desired level of aggregation, i.e., product vs. industry vs. sector vs. aggregate. For expositional simplicity, we do not include any notation to denote products, sectors, etc. However, since the disaggregated gravity equation has solid theoretical foundations (e.g., Anderson and van Wincoop (2004), on the demand side, and Costinot et al. (2012), on the supply side), we can rely on clear guidance for some aspect of the disaggregated gravity estimations with pooled data, e.g., the multilateral resistances should be of dimensions exporter-product-time and importer-product-time.

As a result, the corresponding three sets of fixed effects in our model would become exporter-product-time ($\phi^k_{i,t}$), importer-product-time ($\psi^k_{j,t}$), and country-pair-product ($\overrightarrow{\gamma_{ij}}$), where the superscript k is used to denote the product level (or any other level of aggregation). In addition, some bilateral policies (e.g., sanctions) may be implemented at the product level too. Moreover, regardless of whether a given policy, including EU membership, is implemented at the disaggregated or at the aggregate level, it is very likely that the corresponding effects on trade would be heterogeneous across products, sectors, etc. Finally, the disaggregated dimension should also affect the clustering in our estimations. The main implication for our command is that it should be able to handle even more fixed effects and a larger number of gravity covariates.

3 DiD with Extended High Dimensional Fixed Effects

This section describes a new (Stata) command that can be used to estimate staggered difference-in-differences (DiD) models using an extended two-way fixed effect approach while allowing for high-dimensional fixed effects (`jwdid`).[12] To demonstrate the usefulness of the

[12] To install the `jwdid` command from SSC use: ssc install jwdid, replace. To access the most uptodate version, use: net install jwdid, replace from(http://friosavila.github.io/stpackages).

new command, we rely on it in subsequent sections to obtain all estimation and post-estimation results and figures in the paper. To build the command, we capitalize on three influential strands of the literature. First, we take into account many recent contributions from the vibrant difference-in-differences econometrics literature (e.g., Hull, 2018; de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Wooldridge, 2021; Goldsmith-Pinkham et al., 2022; de Chaisemartin and D'Haultfœuille, 2023; Borusyak et al., forthcoming; Wooldridge, 2023). Second, we take advantage of the recent developments for fast computation and convergence of linear and non-linear models with high dimensional fixed effects (e.g., Correia, 2016b; Correia et al., 2020). Third, we rely on the empirical and theoretical trade gravity literature (e.g., Eaton and Kortum, 2002; Anderson and van Wincoop, 2003; Arkolakis et al., 2012; Yotov et al., 2016; Nagengast and Yotov, 2023). The result from synthesizing the developments from these three strands of the literature is a fast, flexible, and general estimation command – `jwdid`. As we demonstrate below, our command can easily accommodate all established methods for estimating trade gravity models. In addition, the command is quite general as it can be applied directly to gravity models for any bilateral flows and, even more broadly, to any two-dimensional DiD model.

3.1 The Baseline Model and Estimation Command

The analysis in this paper is based on the bilateral gravity model of trade. However, in order to highlight the generality of the command and link it more directly to the recent DiD literature, we introduce it and describe its features based on the following ETWFE estimator proposed by Wooldridge (2021):

$$Y_{i,t} = \alpha + \sum_{g \in G} \sum_{s=g}^{T} \theta_{g,s} D_{g,s} + \xi_i + \xi_t + \varepsilon_{i,t}, \tag{2}$$

where $Y_{i,t}$ is the dependent variable, $D_{g,t}$ is a dummy that takes the value of 1 if the observation is in the treatment group g, on period t, and 0 otherwise. G is a set that indicates at what time treatment started for all units, and T is the last period of the analysis. ξ_i and ξ_t are sets of fixed effects for the individual and time dimensions, respectively.[13] In this setup, the $\theta_{g,t}$ coefficients represent the average treatment effect that the treatment group g experiences at time t ($ATT(g,t)$). As described in Wooldridge (2021), allowing for a flexible specification of the $\theta_{g,t}$ avoids the problem of bad controls and negative weights that have been identified in the literature as potential problems in the estimation of DID models using traditional TWFE estimators.

Based on Equation (2), and accommodating a series of recommendations and developments from the related literature, the benchmark syntax for the proposed `jwdid` command, including its main options as follows:

```
jwdid y, ivar(i) tvar(t) gvar(g) method() never exovar(x_ex) xtvar(x_t) ///
xgvar(x_g) fevar() hettype(option) cluster(cvar)
```

Here, `y` is the dependent variable (e.g., bilateral trade in our case), `ivar(i)` is used to identify the individual panel data dimension or unit (e.g., a country-pair identifier in the gravity settings).[14] `tvar(t)` identifies the time dimension (e.g., the year in our case), and `gvar(g)` identifies the treatment cohorts (e.g., country pairs between EU members). Specifically, for observation it, g would take the value of zero if the unit is never treated (within the window of the analysis), and would take a value different from zero to indicate the year that treatment started for unit i. Following standard assumptions, this specification assumes that the treatment is an absorbing state, i.e., once a unit is treated, it remains treated for the rest of the analysis.

As described in Wooldridge (2023), the standard ETWFE model from Equation 2 identi-

[13]Often, one can use group fixed effects instead of individual fixed effects, and would still obtain numerically identical results in the linear model case.

[14]While the command does not impose the assumption that the data is a panel, the methodology is designed to work with panel data. In case of repeated cross-sectional data, one should exclude `ivar(i)`.

fies the average treatment effect by imposing a linear parallel trends assumption. However, such an assumption may not be valid in some cases, e.g., when the dependent variable follows some limited distribution. Roth and Sant'Anna (2023) discuss a similar problem, stating that the choice of transformation of the dependent variable is crucial for the identification of the average treatment effect, and only under certain conditions would the ATT be identified for any transformation. This is particularly important in gravity settings where, as discussed earlier, the non-linear PPML estimator is the established estimator.

Wooldridge (2023) addresses this issue by demonstrating that the linear ETWFE models can be adapted to allow for non-linear models, by simply imposing the linear parallel trends assumption only on the latent variable of the model, but not on the outcome itself. Capitalizing on that, the `jwdid` command allows the user to specify an estimator of choice using the `method()` option. By default, `jwdid` will use the `reghdfe` command of Correia (2016a), however, there are no restrictions on the type of `method`. Thus, for example, to estimate gravity models with PPML and high-dimensional fixed effects, as we do here, one can use the `ppmlhdfe` command of (Correia et al., 2020), i.e., `method(ppmlhdfe)`.[15]

Specification (2) makes the implicit assumption that parallel trends are satisfied, using all never treated and not-yet treated observations as controls (not included category) for the identification of treatment effects. If one instead wants to relax this assumption, the user can specify the option `never`. In this case, the only observations that are used as controls are the ones that were never treated. In principle, this is the same as the strategy proposed by Sun and Abraham (2021), allowing for full heterogeneity across all groups and all event time periods. Such specification is also numerically identical to the one proposed by Callaway and Sant'Anna (2021), for the case where there are no covariates.

As described in Wooldridge (2021), it is possible to include covariates in the model, by simply adding corrections that enable to easily identify the average treatment effect. However,

[15] The `jwdid` command has not been tested with all possible models. The user should be aware that the `method()` option is passed directly to the model estimation step, and the user should be familiar with the syntax of the model being estimated. Other 'methods' options can be used as well following the syntax method(cmd, cmdoptions).

following the literature on DID models, the implicit assumption is that covariates are time-invariant. `jwdid` allows for adding covariates directly to the benchmark syntax, immediately after the dependent variable, and it does not impose any assumption on the covariates, but the user should be aware of the implications. If one uses the option `xasis`, the command will use the covariates without demeaning them, which may save some computation time.

The default option when introducing additional covariates is to interact all of them (or the demeaned transformations) with the same level of covariate heterogeneity. Sometimes, however, one may not be interested in estimating the same level of heterogeneity for all covariates. It may be possible, for example, to consider separate sets of covariates that could be interacted only with the time or the group dimensions. Specifically, assume there are no variables we wish to consider for the treatment heterogeneity, but instead consider three sets of covariates: variables without further interactions (x^{EX}); variables that would be interacted with the time variables only (x^{T}), and variables that would be interacted with group indicators only (x^{G}). `jwdid` can accommodate such scenarios with the options `exovar(x_ex), xtvar(x_t)`, and `xgvar(x_g)`, respectively.

Sometimes, especially in gravity settings such as ours, it may be beneficial to include high-dimensional fixed effects (and interactions with fixed effects) that are different from the individual and time fixed effects. It is possible to request the inclusion of those types of fixed effects using the option `fevar()`, which is only valid if one is using the default estimator method `reghdfe` or `ppmlhdfe`. In both cases, the additional fixed effects (or interactions) are included in the estimation of the model without further interactions. This is particularly convenient for gravity-type estimations, which often require the use of source-time and destination-time fixed effects.

The number of estimated parameters in Equation 2 can grow quickly with the number of groups/cohorts, time periods, and covariates. This could lead to increasing computational burden of the estimation. An alternative, which is already implemented via `xthdidregress` and `hdidregress` in Stata 18, is to estimate a model that reduces the heterogeneity of the treatment effects. Specifically, it allows treatment effects to vary across cohorts, across absolute time, or across relative time. The `jwdid` command allows the user to specify/impose restrictions along each of these dimensions with `hettype(option)`, where `option` can be `time`, `cohort`, or `event`. If no option is selected, the command will estimate the model by allowing for full cohort-time heterogeneity. For convenience, `twfe` can also be specified, which imposes treatment effect homogeneity across time and cohorts, i.e., the TWFE specification.

Finally, by default, the `jwdid` command assumes clustered standard errors at the `i` level, i.e., at the country-pair dimension in our case, which is consistent with the standard approach in the gravity literature. However, if a different level of clustering is desired, the user can specify the `cluster(cvar)` option. The `cluster(cvar)` option is not required, but can be used to request the standard errors to be clustered at the level `cvar` when using `reghdfe` or `ppmlhdfe`. This is useful for gravity estimations where sometimes three-way clustering (e.g., by source, destination, and time) may be recommended (Egger and Tarlea, 2015).

3.2 Post-estimation Options and Analysis

After the estimation of the model, under the default options, one can recover directly the estimates of the group and time specific average treatment effects on the treated. However, one may also be interested in obtaining aggregated ATTs for the overall data, across groups, or periods, or dynamic effects. Furthermore, when the underlying method is a non-linear model, the estimated coefficients cannot be directly interpreted as the average treatment effect on the outcome, but only on the latent variable. To accommodate such needs, the `jwdid` command comes along with the post estimation command `jwdid_estat/estat` or just `estat`:

```
estat [aggregation] [pw = weight], [vce(unconditional)] [margins_options]
```

`estat` can implement four types of 'aggregations' including (i) an average treatment effect on the treated for all observations that were treated at some point in time (`estat simple`),

(ii) an average treatment effect for observations that were treated at time g (`estat group`), (iii) an average treatment effect at time t for all observations that were effectively treated at that point (`estat calendar`), and (iv) dynamic treatment effects, also known as event studies (`estat event`).

The default option for the estimation of the aggregated ATTs is to use the weights that were used in the estimation of the model. However, if the user wants to use different weights, it is possible to do so using the standard `[pw = weight]` option.

Wooldridge (2021) suggests than when one estimates standard errors for the aggregated ATTs, one should use `vce(unconditional)` option in `Stata`, to allow for uncertainty in the explanatory variables. `jwdid_estat`/`estat` does not use this approach by default, because it requires that the underlying command is able to produce scores for the estimated model. For example, if the model was estimated using `method(regress)`, the scores will be available, and unconditional standard errors for the aggregated ATTs can be estimated. However, this is not possible if one uses `reghdfe` or `ppmlhdfe`.

In addition, standard `margins` options can be used after the post-estimation command `estat`. In particular, given our focus on the gravity setting and the `ppmlhdfe` method, we use the option `predict(xb)`. This yields (approximate) proportional treatment effects (Wooldridge, 2023; Nagengast and Yotov, 2023). Because of this, standard errors are typically estimated using the delta method. If using methods such as `regress` or `logit`, one can also request `vce(unconditional)` so standard errors also account for sampling variation of covariates.

As described earlier, the default aggregation considers all treated observations, imposing restrictions only in terms of time, group, or event dimensions. However, one may be interested in imposing further restrictions that could leverage on the use of covariates, e.g., a dummy variable `dx`, in the model specification. As usual, one can request the estimation of the aggregated ATTs for the whole sample. Alternatively, one can also impose the added restriction that the covariate `dx` is zero or one, using the option `orestriction()`, e.g., `orestriction(dx==1)`. The expression inside the parenthesis should be a valid Stata expression that is used when calculating the aggregated ATTs.

After aggregate effects have been estimated, the user may want to store the results for further analysis or reporting. Because `estat` uses `margins` in the background, the default option is to store the results in memory as `r()` elements. Alternatively, `jwdid_estat`/`estat` allows the user to store the output of the command using three different options: (i) store them as the current estimations in memory `e()` (using the option `post`), (ii) store them in memory as `name` (using the option `estore(name)`), and (iii) store them in a file `filename`, as a `ster` file (using the option `esave(filename)`).

After **time**, **group** or **event** aggregations are estimated (or **simple** with the **over()** option), it is possible to request plotting those results using `estat plot`. The basic syntax is to type it after the aggregation command:

```
estat plot,  style(style)  twoway_options
```

where the option `style` allows the user to select the style of the plot and standard `twoway` graph options can be used after the `estat plot` command. The **plot** command allows for a lot of further flexibility with only minimal command specific options, which we describe in detail in Section A.1 in the Appendix, where we also offer further details on the `jwdid` command and its options. We also provide specific examples for the implementation of the command in our gravity setting in the results section.

4 Data

We rely on standard/established data sources to compile several estimating samples for our empirical analysis. Our largest estimating sample is an unbalanced panel dataset of aggregate bilateral trade flows at the country-pair level, which covers 260 countries over the period 1950-2019. In addition to all goods trade flows, in Section 6.1, we also obtain results with data for manufacturing trade, which covers 229 countries over the period 1980-2016. We also use data on several gravity variables. Finally, we note that, to limit the influence of other events on the estimates and to ensure the same number of pre-treatment observations for all cohorts, we limit the estimating samples to six years before EU accession for treated country pairs.

4.1 Data sources

Aggregate trade flows. The data on nominal aggregate bilateral trade flows was constructed by Felbermayr et al. (2020), who combined the data on aggregate bilateral trade flows at the country-pair level from the two most comprehensive and most widely used trade datasets: (i) the *Direction of Trade Statistics* (DoTS) of the International Monetary Fund (IMF) and (ii) the United Nations' *Commodity Trade Statistics* (COMTRADE) database.[16] To obtain the maximum number of trade flow observations, Felbermayr et al. (2020) use a mirroring procedure. Specifically, consistent with trade theory, where trade volumes are measured at delivered prices (e.g., Eaton and Kortum (2002), Anderson and van Wincoop (2003), Anderson and van Wincoop (2004)), CIF imports are used as the baseline data. Then, missing values are replaced with data on FOB exports.

The trade data of Felbermayr et al. (2020) covers the period 1950-2019, and its main advantage for our purposes is that covers a very long period of time. To obtain our main estimates, we rely on a subsample of this sample, which covers 98% of bilateral trade flows in the world, however, we also demonstrate that our results are robust to using a medium-size sample, which covers 99% of trade flows, and also the full sample. We also experiment with several variants of the dataset to test the robustness of our results to the treatment of missing values in the IMF DoTS dataset.

Manufacturing trade flows. The manufacturing trade flows data come from the Structural Gravity Database (SGD) of the World Trade Organization (Larch et al., 2019).[17] The SGD includes aggregate manufacturing trade data for 229 countries over the period 1980-2016, and

[16] Detailed information about the DoTS database can be found at https://data.imf.org/?sk=9D6028D4-F14A-464C-A2F2-59B2CD424B85, and for the COMTRADE Database at https://comtradeplus.un.org/.

[17] The SGD is available at https://www.wto.org/english/res_e/reser_e/structural_gravity_e.htm.

it has three advantages for our purposes. First, as noted earlier, according to the Eurostat, four fifths of all intra-EU trade is in manufacturing goods (Eurostat, 2023). Thus, this is by far the most important sector for intra-EU trade. Second, on a related note, the sectoral estimates for manufacturing will complement our aggregate results. Finally, the SGD includes consistently constructed domestic trade flows, which will enable us to obtain and compare estimates of the EU effects with and without domestic trade flows.

Other data. We also coded the data on EU membership according to Figure 2, and we use data on several standard gravity variables in our analysis, which come from the Dynamic Gravity Dataset (DGD) of the United States International Trade Commission (Gurevich and Herman (2018)).[18]

There are various reasons to expect that the onset of the EU effects on trade would precede the date of the official EU accession. For example, accession to the EU is a multi-stage process with a gradually increasing degree of commitment. Specifically, the process involves 6 steps, including (i) Application for accession, (ii) Opinion from the Commission, (iii) Granting of candidacy, (iv) Accession negotiations, (v) Signing of an accession treaty, and (vi) Entering into force of the accession treaty (Leppert, 2022). It is quite possible to observe certain adjustments in trade costs as well as in the behavior of economic agents at each stage of the accession process. However, we believe that it is most likely to observe such changes during the period between the signing of the accession treaty and its official entry into force. This period varies by cohort, and it is usually between about 1 year (e.g., for the 2004 cohort) and about 2 years (e.g., for the 2007 cohort).

Against this backdrop, for our main analysis, we will assume an 'onset' of the EU effects two years before the official entry into force. In addition, in the robustness analysis, we consider two alternative strategies: (i) We omit periods with anticipation effects; and (ii) We use the year of EU accession as the treatment onset.

[18] The DGD can be downloaded at https://catalog.data.gov/dataset/dynamic-gravity-dataset-1948-2016.

4.2 Descriptive statistics

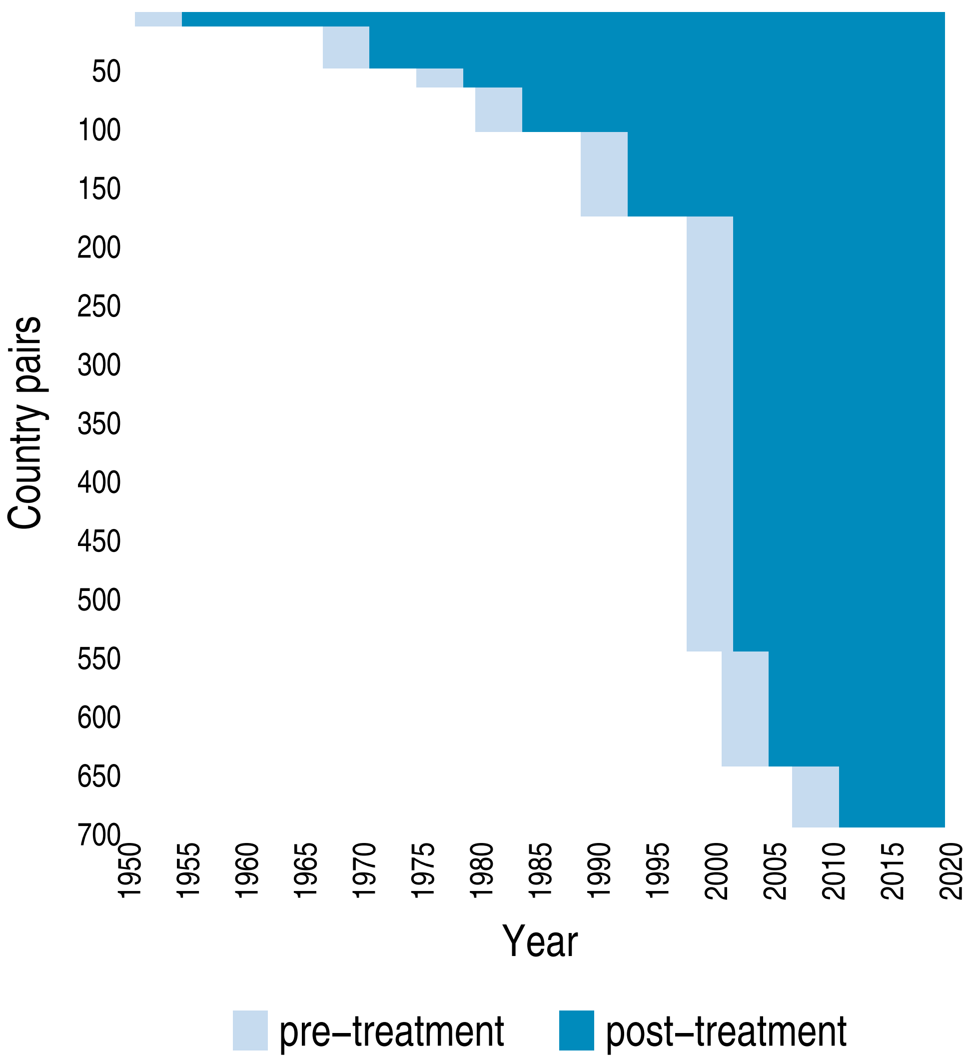
Figure 2 shows all countries that were, at some point, part of the EU by accession year.[19] It shows that EU enlargement occurred in several waves usually with several countries entering the EU at the same time. Note that separate foreign trade data for Belgium and Luxembourg is only available from 1999 onwards. To avoid breaks, we consider Belgium-Luxembourg as one exporter/importer throughout the analysis. Note also that data for Belgium-Luxembourg is only available in our dataset from 1960 onwards such that Belgium-Luxembourg is not included in the 1957 cohort.

Figure 3 shows the observations of country pairs in the treatment group by pre- and post-treatment status over time. The treatment cohorts are of different sizes – in accordance with the different waves of EU enlargement from Figure 2 – and also of different lengths depending on the year of treatment onset. Note that country pairs are combinations of exporters and importers. Therefore, cohorts are composed not just of new members, but also of combinations of new and old members, an aspect that we analyze in detail in Section 5.3.

For completeness, Table 1 reports the number of observations, country pairs, exporters, importers, and years for different groups in the dataset. Similarly, Table 2 reports the average of the variables Distance (in kilometers), Contiguity, Language, and Colony for different groups in the dataset. Note that, not surprisingly, treated country pairs differ quite substantially with regard to standard gravity variables. Country pairs in the EU are closer together, more likely to share a common border, but less likely to share a common language or a past colonial relationship. The most salient difference between treated and never-treated country pairs is the substantially smaller distance between EU member states. In Section 6, for robustness, we consider an extension of the ETWFE model relaxing the parallel trends assumption, which now only needs to hold conditional on distance.

[19] We include the United Kingdom in the analysis since it only left the EU in the 2020, while our sample already ends in 2019.

Figure 3: Cohorts and treatment status over time.



Notes: This figure shows the treatment status of treated country pairs over time. Pre-treatment years are colored in light blue. Post-treatment years are colored in dark blue. Note that, as explained in the main text, the treatment onset was shifted forward by two years to capture anticipation facts. As a result, for example, the post-treatment years for the 1957 cohort already begin in 1955. The figure was generated using the Stata module `panelview` (Mou and Xu, 2022).

Table 1: Descriptive statistics: Observations along different dimensions

Group	Observations	Pairs	Exporters	Importers	Years
1957 cohort	780	12	4	4	65
1973 cohort	1,764	36	8	8	49
1981 cohort	656	16	9	9	41
1986 cohort	1,368	38	11	11	36
1995 cohort	1,944	72	14	14	27
2004 cohort	6,660	370	24	24	18
2007 cohort	1,470	98	26	26	15
2013 cohort	468	52	27	27	9
Treated	15,110	694	27	27	65
Not-yet treated	2,776	694	27	27	31
Never treated	399,245	7,232	90	90	70

Notes: The table reports the number of observations, country pairs, exporters, importers, and years for different groups in the baseline estimation sample from the ETWFE estimate in column (2) of Table 3. 'Cohort' refers to all post-treatment observations of new EU country pairs in a particular year. 'Treated' refers to all post-treatment observations of all cohorts. 'Not-yet treated' refers to all pre-treatment years of all cohorts. 'Never treated' refers to all observations of country pairs that did not enter the EU during the sample period.

Table 2: Descriptive statistics: Summary statistics of covariates for different groups

Group	Distance	Contiguity	Language	Colony
1957 cohort	926	0.50	0.33	0.00
1973 cohort	991	0.11	0.33	0.06
1981 cohort	2,075	0.00	0.75	0.00
1986 cohort	1,702	0.11	0.00	0.00
1995 cohort	1,617	0.08	0.14	0.00
2004 cohort	1,503	0.08	0.14	0.01
2007 cohort	1,473	0.06	0.16	0.00
2013 cohort	1,124	0.08	0.38	0.00
Treated	1,458	0.10	0.19	0.01
Never treated	7,945	0.02	0.34	0.03

Notes: The table reports the average of the variables Distance (in kilometers), Contiguity, Language, and Colony for different groups in the baseline estimation sample from the ETWFE estimate in column (2) of Table 3. 'Cohort' refers to all new EU country pairs in a particular year. 'Treated' refers to all cohorts. 'Never treated' refers to all country pairs that did not enter the EU during the sample period.

5 Estimates of the EU effects on trade

This section presents our main results. In Section 5.1, we test the identifying assumptions of the ETWFE model. In Section 5.2, we obtain and discuss our main results. Finally, in Section 5.3, we explore the heterogeneity of the EU effects across several dimensions. Importantly, all results and all figures that we present in this section are obtained with the new estimation command. To highlight this, we have included the command syntax used to obtain each table and figure in this section in the notes for the corresponding exhibit. In this regard, we define `trade` to be exports, which vary over the exporter-importer-year dimension, `id_ci_cj` is a country-pair id, `year` indicates the year in the dataset, `FT_EU` indicates the first year, in which both the exporter and importer of a country pair were part of the EU, and `idt_ci` and `idt_cj` are exporter-time and importer-time identifiers.

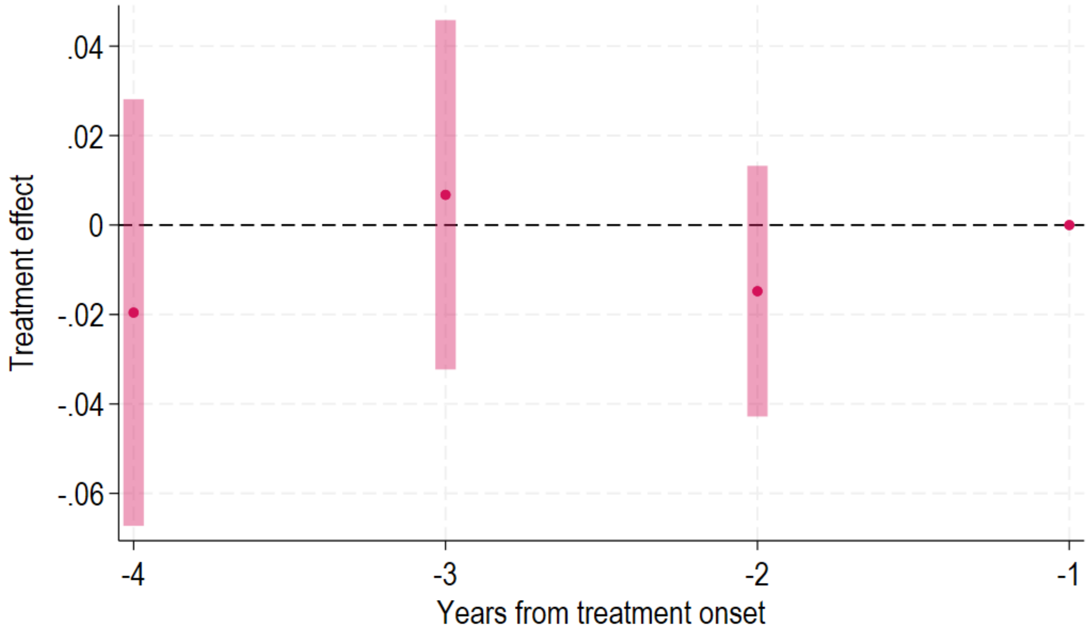
5.1 Test of identifying assumptions

In this section, we test the no anticipation assumption and the parallel trends assumption of the ETWFE model (Wooldridge, 2021, 2023; Nagengast and Yotov, 2023). We do so by following Wooldridge (2021, 2023) by augmenting the estimating equation 1 with cohort-year-specific placebo effects in the years before treatment onset.[20] Figure 4 shows the resulting placebo treatment effects aggregated across cohorts by event time. The results provide evidence against the violation of the identifying assumptions before EU accession. The pre-treatment effects are statistically insignificant in the years before treatment onset.[21] We, therefore, proceed with our main analysis.

[20] Alternatively, using only untreated observations for the pre-treatment test in the spirit of Borusyak et al. (forthcoming) yields very similar results (Figure A1).

[21] Note, however, that a joint significance test yields a p-value of 0.090 reflecting the small negative placebo effects in the years -4 and -2.

Figure 4: Pre-treatment effects



Notes: The figure reports pre-trend estimates from a PPML estimation of equation (1) augmented with cohort-year-specific placebo effects in the four years before treatment onset. To avoid misleading visual inferences, we follow the recommendation by Roth (2024) and put our pre-treatment estimates in a different plot from the post-treatment estimates. The regression is estimated using the full sample following Wooldridge (2021, 2023). The cohort-year-specific treatment effects were aggregated across cohorts to obtain event-time-specific treatment effect estimates. 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) never
estat event, predict(xb) pretrend window(-4 -1)
estat plot, ytitle("Treatment effect") xtitle("Years from treatment onset") pstyle1(p2)
```

5.2 Main estimation results

Our main aggregate EU estimate appears in column (2) of Table 3, and they are obtained from an estimating sample that covers 98% of world trade. The main implication of these results is that, on average, membership in the EU has led to a significant increase in bilateral trade flows among its member states. Specifically, the EU estimate in column (2) implies that the EU has led to an increase of more than 52% in the bilateral trade flows between its member states.[22] Our estimate is significantly larger then recent estimates of the effects of trade agreements (e.g., Baier et al., 2019; Larch and Yotov, 2022; Nagengast and Yotov, 2023). Moreover, it is also very likely that the average estimates in Table 3 may be masking significant heterogeneity across various dimensions, e.g., over time, across cohorts, and across

[22] Calculated as $(\exp(0.422) - 1) \times 100 = 52.5$.

EU member states. In subsequent analyses, we demonstrate that this is indeed the case.

Table 3: Average EU effects on trade: TWFE vs. ETWFE for different samples

	(1) TWFE	(2) ETWFE	(3) TWFE	(4) ETWFE	(5) TWFE	(6) ETWFE
$EU_{ij,t}$	0.332***	0.422***	0.337***	0.435***	0.338***	0.431***
	(0.035)	(0.037)	(0.035)	(0.037)	(0.035)	(0.036)
Sample	Baseline	Baseline	Medium	Medium	Large	Large
Observations	417,131	417,131	642,570	642,570	2,303,571	2,303,571
Exporters	90	90	112	112	260	260
Importers	90	90	112	112	260	260
Years	70	70	70	70	70	70
Coefficients	1	260	1	260	1	260
p-value H0:TWFE=ETWFE		0.0014		0.0004		0.0005
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents PPML regression results using a TWFE model, in which the cohort-year specific effects in equation (1 were substituted with a homogeneous EU effect, and the ETWFE model (equation (1)), for which the cohort-year-specific treatment effects were aggregated to obtain an aggregate treatment effect estimate. The dependent variable is exports, which vary over the exporter-importer-year dimension. The 'Small' sample contains 90 countries, accounting for 98% of world exports. The 'Medium' sample, which is used to obtain the results in columns (3) and (4), contains 112 countries, accounting for 99% of world exports. The 'Large' sample, which is used to obtain the results in columns (5) and (6), contains the full set of countries from our dataset. 'Coefficients' reports the number of estimated coefficients apart from the fixed effects. Standard errors in parentheses are clustered by country pair. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. The row labeled "p-value H0: TWFE=ETWFE" reports the results of a standard F-test of the equality of the corresponding TWFE and ETWFE coefficients, where the covariance between the two coefficients is accounted for using a seemingly unrelated regression specification.
Stata code:

```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) ///
            hettype(twfe) // Columns (1), (3), and (5)
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) // Columns (2), (4), and (6)
estat simple, predict(xb) // All columns
```

To highlight the policy and methodological implications of our ETWFE estimates, we compare them to those from a standard TWFE gravity model, which are reported in column (1) of Table 3. While the difference between the estimates does not seem to be very large, they are statistically significantly different from each other. Moreover, the difference between the implied trade-volume effects between the two specifications is 13 percentage points (39.4% from the TWFE model vs. 52.5% from the TWFE specification), i.e., not negligible.[23] We confirm the robustness of these results in columns (3) and (4) of Table 3, where

[23]To learn more about the reasons behind the differences between the TWFE and ETWFE results, a decomposition analysis of the *OLS* TWFE estimate in the spirit of de Chaisemartin and D'Haultfœuille (2020) can be performed. The underlying assumption is that the OLS decomposition results are likely informative about the PPML estimates as well. Given that the pattern of differences between the TWFE and ETWFE results is relatively similar for OLS and PPML (Table 5), we hypothesize that this is indeed

we use a medium sample that covers 99% of trade flows in the world, and in columns (5) and (6), where we use the full sample. The EU estimates are very similar across all ETWFE specifications and, in each case, they are also larger than the corresponding TWFE results.

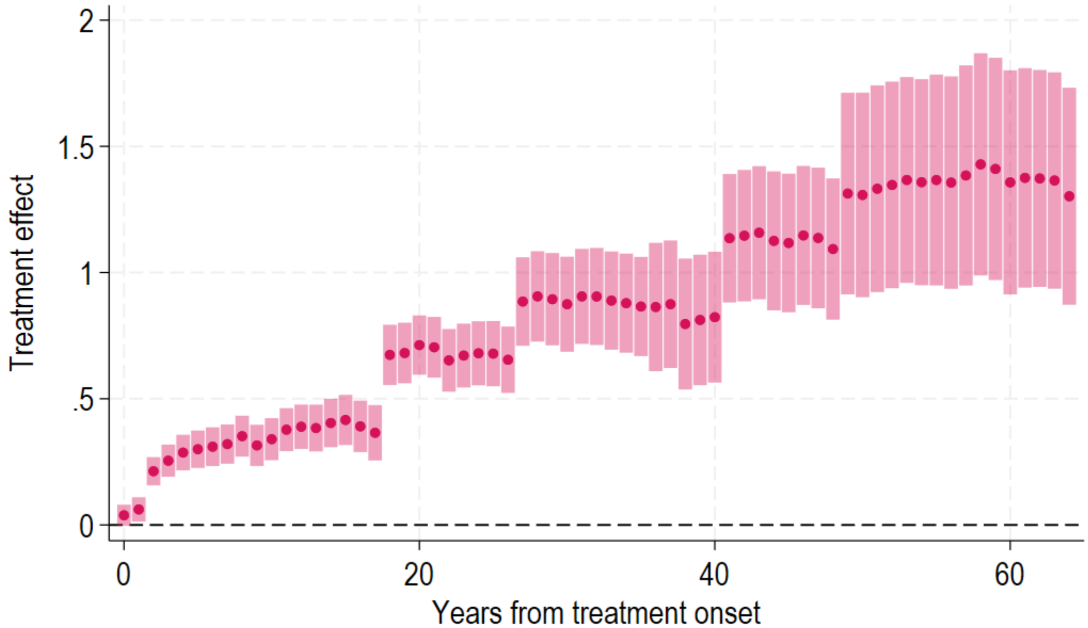
We complement the aggregated EU estimates from Table 3 with a set of event-time-specific ETWFE EU effects, which are visualized in Figure 5. Several patterns stand out from the evolution of the EU effects over time. First, in line with the literature and the discussion in Section 4.1, we find positive and significant anticipation effects in the first two years before EU accession. While the anticipation effects are statistically significant, they are at the same time substantially smaller than the effects after the actual EU accession (i.e., from period 2 onwards in Figure 5). Second, consistent with our aggregate estimates, all estimates of the phasing-in effects of the EU on trade are positive and statistically significant. Third, the EU effects increase over time throughout the period of investigation. A possible interpretation of these results is that the EU effects are long-lasting. However, it is also possible that the average event-time-specific effects hide certain cohort-specific patterns in the evolution of the EU effects. The few 'jumps' that we see in Figure 5 are consistent with this hypothesis. As we will see in the next section, those 'jumps' are due to composition effects, because the underlying event-study plots by cohort are smooth and have no 'jumps'.

5.3 Disaggregated EU effects on trade

Motivated by the results in the previous section, and in order to demonstrate the importance of the heterogeneity-robust DiD approach for estimating the EU effects, in this section, we explore several dimensions of the underlying heterogeneity in the EU effects. In particular, we consider heterogeneity across cohorts, cohort-specific event-study results, and we also obtain directional estimates on trade between 'old' and 'new' EU members depending on the recency of their EU membership.

the case. The OLS TWFE decomposition results suggests that the TWFE is likely downward biased since it places larger weights on short-term effects and on the effect of late cohorts, which are both associated with smaller treatment effects.

Figure 5: Event-time-specific EU effects



Notes: The figure reports event-time-specific treatment effects, for which the cohort-year effects (corresponding to the aggregate EU effect in column (2) of Table 3) were aggregated across cohorts. Note that the treatment onset was shifted forward by two years (Section 4.1) such that the year of the EU accession corresponds to time period 2. 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj)
estat event, predict(xb)
estat plot, ytitle("Treatment effect" xtitle("Years from treatment onset").
```

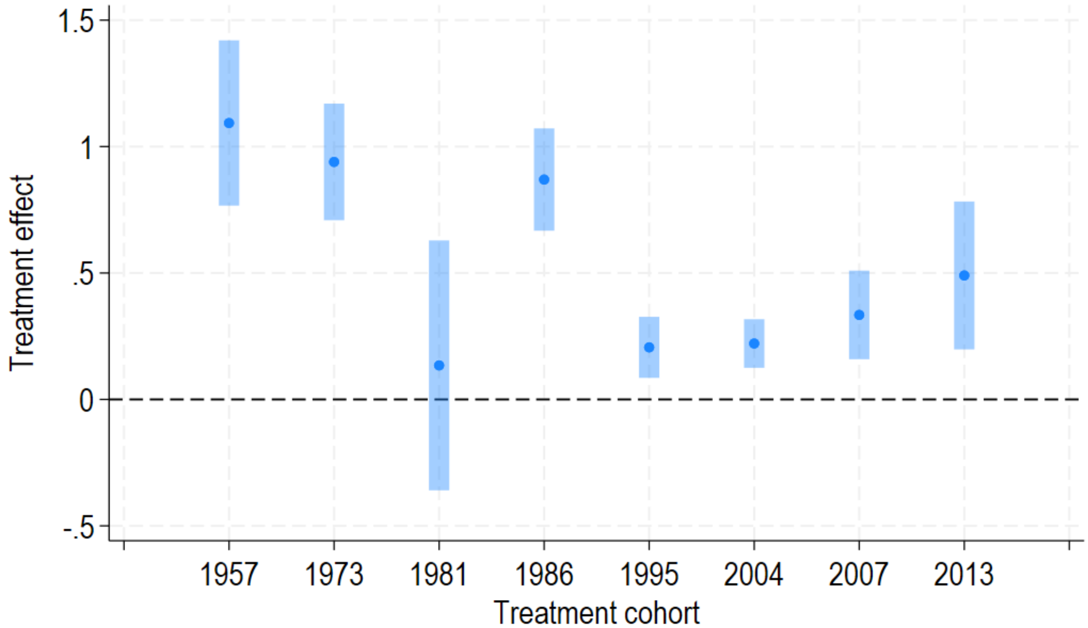
EU effects by cohort. We start with an analysis by cohort. In this regard, the heterogeneous waves of EU accession (see Figure 2) offer a very interesting setting. Our estimates are visualized in Figure 6. Two main results stand out.

First, all cohort estimates are positive and all but one of them are statistically significant. Thus, consistent with our main findings, the implication is that the EU has benefited its members' exports. The single exception is the 1981 cohort, which only consists of all country pairs of previous EU members with Greece.[24]

The second conclusion that we draw based on the results in Figure 6 is that the effects of the EU have been quite heterogeneous across cohorts. The results in Figure 6 suggest that,

[24]Note that this is not to say that Greece's trade has not benefited from its EU membership. In unreported country-specific results, we find that the EU effect for Greece's exports is indeed positive. This stems from Greece being part of subsequent cohorts (as an 'old' member).

Figure 6: Cohort-specific treatment effects



Notes: The figure reports cohort-specific treatment effects, for which the cohort-year effects (corresponding to the aggregate EU effect in column (2) of Table 3) were aggregated across event time. 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj)
estat group, predict(xb)
estat plot, ytitle("Treatment effect") xtitle("Treatment cohort")
```

apart from the 1981 cohort, the earlier cohorts, i.e., the 1957 cohort (country pairs consisting of the founders), the 1973 cohort (country pairs consisting of previous EU members with Denmark, Ireland, and the United Kingdom) and the 1983 cohort (country pairs consisting of previous EU members with Portugal and Spain), have benefited more from the corresponding EU enlargements.

A possible explanation for this is that the benefits from the most recent rounds of enlargements have not materialized yet, i.e., the differences that we capture can be due to composition effects because later cohorts have fewer long-term effects than early cohorts. However, the overall pattern of results remains qualitatively similar – apart from a very strong effect for the 2013 cohort comparable to those of the 1973 and 1986 cohort – when conditioning on the same number of years for every cohort (Figure A2 in the Appendix). Subsequently, we explore whether differences in phasing-in effects might leave room to expect additional trade effects for late cohorts in the years after 2019.

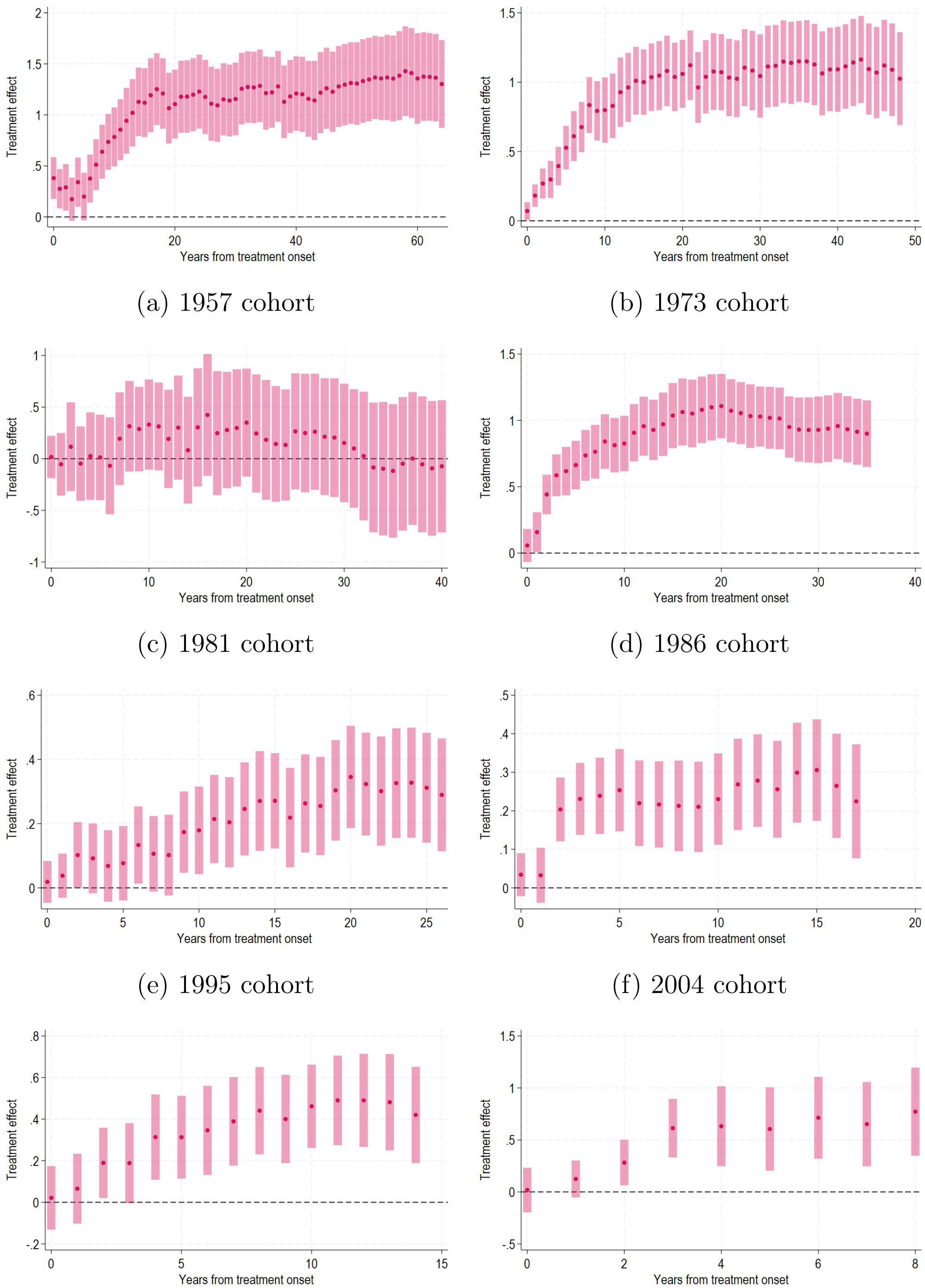
Event-study results by cohort. In our next experiment, we obtain event-study results by cohort, i.e., we trace the evolution of the EU effects over time for each of the cohorts in our sample. Our findings are reported in Figure 7 and, overall, they demonstrate that the more aggregated estimates that we discussed thus far indeed concealed significant heterogeneity. Several results are noteworthy.

First, we observe that all cohort-specific event-study plots are smooth over time and that there are no 'jumps'. However, note that estimates of event-study effects far away from treatment onset are only available for earlier cohorts. In combination with the heterogeneous EU effects that we observe across cohorts (Figure 6), this results in the discontinuities that were apparent in the aggregate event-study plot (Figure 5), i.e., the 'jumps' are due to composition effects.

Second, we see that the positive effects for the 1973, and 1986 cohorts accrue relatively quickly. We should also note that even if these effects (which last for 10-15 years) seem to be exhausted relatively quickly, they are still significantly more long-lasting than recent estimates of the duration of the phasing-in effects of FTAs, e.g., Egger et al. (2022). For the 1957 cohort, positive effects also appear quickly and then begin to rise again strongly around 5 years after the foundation of the European Community perhaps reflecting that the benefits for the initial founders set in with a certain delay as the initial integration took more time than for later cohorts.

Third, we also see from Figure 7 that both the magnitude and the evolution of the EU effects vary significantly across cohorts. Thus for example, the effects for the 1973 and 1986 cohorts are large and are exhausted relatively fast, while the effects for the 1995 cohort are smaller but more long-lasting. There were hardly any anticipation effects for the 2004 cohort, while the remaining effects have been relatively immediate, small, and stable over time. A possible caveat with the interpretation of the results for the 2004 cohort is that these estimates may be masking significant heterogeneity across the large number of countries that are in this cohort. The effects for the 2007 cohort (Bulgaria and Romania) are relatively large and gradual, and they seem to be exhausted already, about 10 years after accession. Finally, the impact of EU membership on Croatia's trade (the single country in the 2013 cohort) has been strong and it is ongoing 6 years after its accession. Therefore, from this perspective, the the cohort heterogeneity observed in Figure 6 also cannot be explained by differences in phasing-in effects, while the final effect of the 2013 cohort might very well end up in the ball park of the 1957, 1973 and the 1986 cohort.

Figure 7: Event-time specific EU effects by cohort



(a) 1957 cohort

(b) 1973 cohort

(c) 1981 cohort

(d) 1986 cohort

(e) 1995 cohort

(f) 2004 cohort

(g) 2007 cohort

(h) 2013 cohort

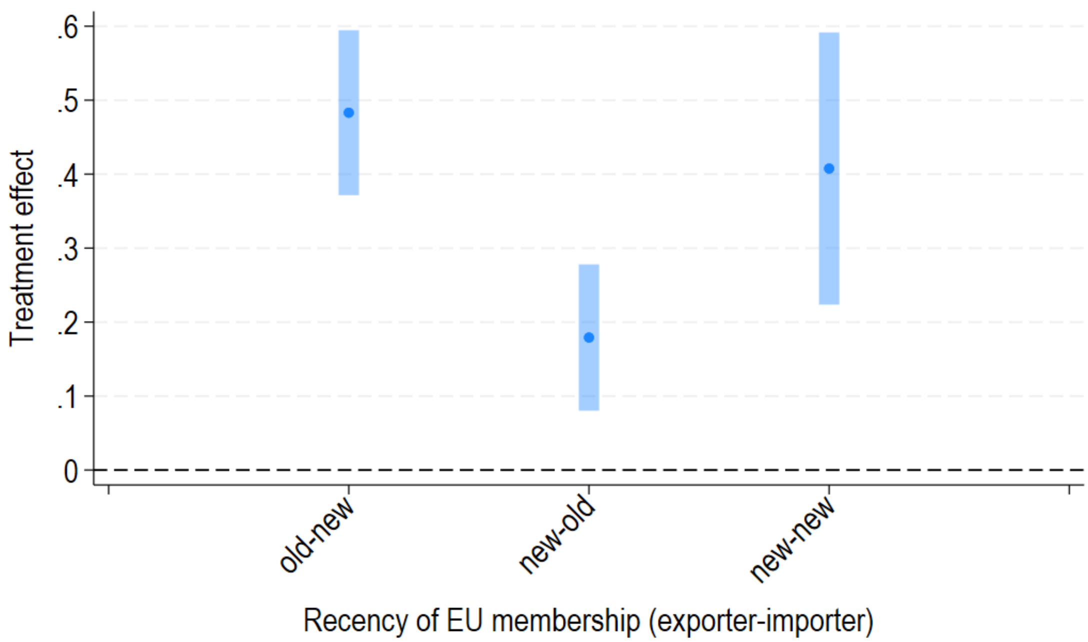
Notes: The figure reports event-time specific EU effects by cohort, i.e., all estimated cohort-year effects (corresponding to the aggregate EU effect in column (2) of Table 3). 95% confidence intervals are shown using standard errors clustered by country pair.

Stata code:

```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj)
estat event, predict(xb) ores(FT_EU==‘year’) // ‘year’ corresponds to the first post-treatment year of a cohort
estat plot, ytitle("Treatment effect") xtitle("Years from treatment onset")
```

EU effects by recency of membership. We conclude with an exploration of possible directional asymmetries in the effects of the EU based on recency of membership. Specifically, we split the countries in each cohort into two groups: 'old' members that were already part of the EU and 'new' members that joined the EU in that particular year. Note that this definition is dynamic with regard to countries and always relative to a particular cohort, i.e., a country will initially be a 'new' member in the cohort of the year in which it entered the EU, but it will be an 'old' member in subsequent cohorts. Then, we obtain estimates of the effects of the EU depending on the direction of trade flows, e.g., old-to-new, new-to-old, and new-to-new. Note that, by definition, we cannot estimate the direction old-to-old as these country pairs are already part of the other categories (in previous cohorts).

Figure 8: EU trade effects by recency of membership



Notes: The figure reports the average EU effect by recency of membership from a specification allowing for additional heterogeneity, for which the cohort-recency-year effects were aggregated across cohorts and event time. 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

```
jwdid trade i.newold, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj)
estat simple, predict(xb) over(newold) // 'newold' indicates the direction of trade flows by recency
estat plot, ytitle("Treatment effect",size(large)) xtitle("Recency of EU membership (exporter-importer)") ///
          xlabel(0 " " 1 "old-new" 2 "new-old" 3 "new-new" 4 " ", angle(45)) ylabel(0(0.1)0.6
```

Our results appear in Figure 8, and we draw the following conclusions based on them.

First, trade has increased for all groups and in each direction, and all estimates are statistically significant. Second, the effects have been asymmetric. Specifically, the estimate of the effects of the EU on the exports from old-to-new members (0.483, std.err. 0.057) and from new-to-new members (0.408, std.err. 0.094) is substantially larger (more than twice as large) than the corresponding estimate for the exports of new-to-old members (0.179, std.err 0.050). We note that this pattern also mostly holds across cohorts (Figure A3) and cannot be explained by differences in pre-trends (Figure A4). Furthermore, the differences between the new-to-old members and the other two categories seem to be particularly pronounced for the EU accession waves from 1995 onwards.

These asymmetries are consistent with the estimates from Anderson and Yotov (2022), who explore the EU impact on the extensive margin of trade and also find significantly stronger impact on the number of products traded from 'old' to 'new' members than from 'new' to 'old' members (here defined at the country level). As noted in Anderson and Yotov (2022), a possible explanation for this result is that the new members have not been very successful in placing their products on the more competitive Western European market.

Overall, the results in Figure 8 reinforce our previous conclusion that the EU has promoted trade among its members significantly. However, we now also see that the increase in exports may have been asymmetric and in favor of the 'old' member states. We do note that (i) the new members have also enjoyed larger exports due to their EU membership, particularly with other 'new' member states (ii) it is possible that their exports to old members will keep increasing over time and that they will also further profit from future enlargements as by now seasoned ('old') EU members, and (iii) they have further benefited from their imports from the rest of the EU. Nevertheless, we believe that our results may have important policy implications.

6 Robustness analysis

This section describes a series of robustness checks, which are split into three groups: (i) different datasets (Section 6.1), (ii) degree of heterogeneity and alternative estimators (Section 6.2), and (iii) DiD- and ETWFE-specific assumptions (Section 6.3).

6.1 Dataset

With regard to the dataset, first, we use the WTO's Structural Gravity Database (Larch et al., 2019) as a different dataset and, second, we conduct several robustness checks concerning the treatment of missing values in the IMF DoTS dataset that was used to generate our baseline dataset.

WTO's Structural Gravity Database. First, we use the WTO's Structural Gravity Database (Larch et al., 2019), which contains aggregate manufacturing trade covering the years 1980-2016.[25] The main differences relative to our main dataset are that the SGD includes domestic trade flows, it covers a shorter time period, and it only includes manufacturing trade rather than all merchandise trade as our baseline dataset.

The ETWFE results for the SGD are presented in Table 4, where the estimates are based on a small sample that contains 72 countries, accounting for 98% of world exports (columns (1)-(3)), a medium sample that contains 92 countries, accounting for 99% of world exports (columns (4)-(6)), and a large sample that contains all 229 countries (columns (7)-(9)). In addition, we capitalize on the fact that the SGD includes domestic trade flows, and we explore their importance for the intra-EU estimates.

We proceed in four steps. First, in column (1), we reproduce our main estimates, which are obtained with the small sample including domestic trade flows and with the full set of border-time dummies, which will account for common globalization trends. Next, in column (2), we use the same sample but without the common globalization trends. Consistent with the main

[25] The SGD is available at https://www.wto.org/english/res_e/reser_e/structural_gravity_e.htm.

argument from Bergstrand et al. (2015), the EU estimate is significantly larger. A possible explanation is that it now also captures some common globalization effects. Third, the results in column (3) are obtained from a sample without domestic trade flows. The estimate of the EU effects is positive and statistically significant, but smaller than the corresponding result form column (1). Consistent with Dai et al. (2014), the explanation for the difference is that the specification with domestic trade flows from column (1) enables us to explicitly account for diversion from domestic toward international sales.

We confirm these conclusions with our medium and large samples in columns (4)-(6) and (7)-(9), respectively. Finally, we note that the coefficients in columns (1), (4), and (7) are substantially smaller than those based on our baseline dataset. This is likely related to the other two salient differences between the datasets. First, the SGD does not include the 1973 cohort, which was found to have the largest treatment effect (Figure 6b). Second, the effects of manufacturing trade tend to be slightly smaller than those of trade in agricultural goods (Fontagne and Yotov, 2024), which are also included in our dataset.

Robustness checks related to baseline dataset. Next, we consider a series of robustness checks concerning the treatment of missing values in the IMF DoTS dataset that was used to generate our baseline dataset. In columns (10)-(12) of Table 4, we replicate our main specification using the small baseline sample while progressively dropping observations associated with different assumptions imposed during the construction of our dataset. First, in column (10), we omit observations, in which missing values were replaced with zeroes (60,975 out of 417,131). The resulting coefficient is very similar to our baseline estimate, but slightly smaller (0.409).

Second, in column (11), we additionally omit observations, in which data on exports was used to replace missing import data (40,891 observations). This results in a coefficient, which is large and positive, but smaller than our baseline estimate (0.373). Apart from changes in the control group, this is likely due to the omission of the 1957 cohort (note that the number of estimated coefficients drops from 260 to 195), which was associated with large treatment effects.

Lastly, in column (12), we additionally omit observations, in which UN COMTRADE imports were used to replace missing imports in the IMF data (17,959 observations), i.e., this variant essentially corresponds to the original IMF data on imports only. The resulting coefficient is similar to the one in column (11), albeit slightly smaller. Overall, we conclude, that, while slightly reducing the magnitude of the aggregate effect, the assumptions underlying the construction of our dataset do not materially affect our main result.

Table 4: Robustness with regard to a different dataset, domestic trade flows, and the construction of the baseline dataset

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$EU_{ij,t}$	0.328***	0.592***	0.076**	0.327***	0.592***	0.091***	0.326***	0.591***	0.102***	0.409***	0.373***	0.361***
	(0.033)	(0.035)	(0.036)	(0.033)	(0.035)	(0.035)	(0.033)	(0.035)	(0.034)	(0.037)	(0.037)	(0.037)
Dataset	SGD	SGD	SGD	SGD	SGD	SGD	SGD	SGD	SGD	Baseline	Baseline	Baseline
Sample	Small	Small	Small	Medium	Medium	Medium	Large	Large	Large	Small	Small	Small
Variant										V1	V2	V3
Observations	152,642	152,642	150,435	239,955	239,955	237,343	914,854	914,854	911,209	356,156	315,265	297,306
Exporters	72	72	72	92	92	92	229	229	229	90	90	90
Importers	72	72	72	92	92	92	229	229	229	90	89	89
Years	37	37	37	37	37	37	37	37	37	70	60	60
Coefficients	90	90	90	90	90	90	90	90	90	260	195	195
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross-border × year FE	Yes			Yes			Yes					
Domestic trade	Yes	Yes		Yes	Yes		Yes	Yes				

Notes: The table presents PPML regression results from the ETWFE model (equation (1)), for which the cohort-year-specific treatment effects were aggregated to obtain an aggregate treatment effect estimate. Columns (1)-(9) are based on the WTO SGD, while columns (10)-(12) are based on variants of our baseline dataset. The dependent variable is exports, which vary over the exporter-importer-year dimension. The 'Small' sample of the SGD contains 72 countries, accounting for 98% of world exports (columns (1)-(3)). The 'Medium' sample of the SGD contains 92 countries, accounting for 99% of world exports (columns (4)-(6)). The 'Large' sample of the SGD contains the full set of countries from the SGD (columns (7)-(9)). Variant 'V1' of our baseline dataset omits observations, in which missing values were replaced with zeroes. Variant 'V2' of our baseline dataset additionally omits observations, in which data on exports was used to replace missing import data. Variant 'V3' of our baseline dataset additionally omits observations, in which UN COMTRADE imports were used to replace missing imports in the IMF data. Cross-border × year FE indicates whether border-time dummies are included, which account for common globalization trends. 'Domestic trade' indicates whether domestic trade flows are included. 'Coefficients' reports the number of estimated coefficients apart from the fixed effects. Standard errors in parentheses are clustered by country pair. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Stata code:

```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj brdr_time) // Columns (1), (4), and (7); brdr_time are cross-border x year FE
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) // All other columns
estat simple, predict(xb) // All columns
```

6.2 Degree of heterogeneity and alternative estimators

With regard to the heterogeneity of the estimator, we consider variants that are less flexible and that require the estimation of fewer coefficients by imposing restrictions on the treatment effect heterogeneity. In addition, we consider an imputation estimator that allows for arbitrary heterogeneity at the country-pair-year level. We also implement a jackknife estimator that can address potential IPPs, which may arise in non-linear models with fixed effects. Lastly, we provide results for OLS instead of PPML. Our findings are reported in Table 5.

Restrictions on heterogeneity. First, we impose restrictions on the ETWFE model by allowing the treatment effect to vary only at the event-year level (column (1)) or at the cohort level (column (2)). Note that the model in column (1) is akin to a standard event-study specification. We find that the resulting aggregate coefficient in column (1) is only slightly larger than the baseline TWFE estimate, but still substantially smaller than the corresponding ETWFE estimate. By contrast, the model that allows for only cohort heterogeneity in column (2) yields a coefficient that is very close, but still slightly smaller than our baseline ETWFE estimate. We conclude that cohort heterogeneity is more important in our setting than treatment effect heterogeneity across time.

Table 5: Robustness with regard to degree of heterogeneity of the ETWFE estimator and alternative estimators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$EU_{ij,t}$	0.353***	0.409***	0.348	0.328***	0.419***	0.422***	0.620***	0.532***
	(0.039)	(0.034)	(.)	(0.063)	(0.072)	(0.063)	(0.040)	(0.037)
OLS/PPML	PPML	PPML	PPML	PPML	PPML	PPML	OLS	OLS
Estimator	ETWFE	ETWFE	Imputation	Jackknife TWFE	Jackknife ETWFE	ETWFE	ETWFE	TWFE
Unit heterogeneity		Cohort	Pair	Cohort	Cohort	Cohort	Cohort	Cohort
Time heterogeneity	Year		Year	Year	Year	Year	Year	Year
Standard error clustering	Exp× Imp	Exp× Imp				Exp, Imp, year	Exp× Imp	Exp× Imp
Observations	417,131	417,131	417,187	417,131	417,131	417,131	356,204	356,204
Exporters	90	90	90	90	90	90	90	90
Importers	90	90	90	90	90	90	90	90
Years	70	70	70	70	70	70	70	70
Coefficients	65	8		260	260	260	260	1
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents PPML regression results (except columns (7) and (8) which are based on OLS) using variants of the ETWFE model (equation (1)), for which the cohort-year-specific treatment effects were aggregated to obtain an aggregate treatment effect estimate. The dependent variable is exports which vary over the exporter-importer-year dimension. Column (1) imposes complete homogeneity along the cohort dimension and only allows for event-year-specific treatment effects. Column (2) imposes complete homogeneity along the time dimension and only allows for cohort-specific treatment effects. Column (3) shows results using an imputation estimator with the same fixed effect structure as the baseline. Columns (4) and (5) show results for a jackknife TWFE and ETWFE using 1,000 draws described in Nagengast and Yotov (2023). The regression in column (6) is the same as in column (2) of Table 3, but standard errors are clustered by exporter, importer, and year. Columns (7) and (8) show results of the TWFE and ETWFE model estimated using OLS. 'Coefficients' reports the number of estimated coefficients apart from the fixed effects. Standard errors in parentheses are clustered by country pair (except for column (6)). ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Stata code:
```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) hettype(event) // Column (1)
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) hettype(cohort) // Column (2)
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) cluster(exporter importer year) // Column (6)
jwdid lntrade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) fevar(idt_ci idt_cj) // Column (7); lntrade is defined as gen lntrade=ln(trade)
jwdid lntrade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) fevar(idt_ci idt_cj) hettype(twfe) // Column (8)
estat simple, predict(xb) // Columns (1)-(2) and (6)-(8)
```

Imputation estimator. Next, we consider an imputation estimator that provides noisy estimates of individual treatment effects (Borusyak et al., forthcoming). While Wooldridge (2023) shows that the imputation approach and the ETWFE regression are numerically equivalent for PPML, this equivalence breaks down under the more complex fixed effect structure of the gravity setting (Nagengast and Yotov, 2023). The imputation estimator applied to our dataset is also positive and significant (column (3)), but smaller than our baseline ETWFE estimate.[26] The difference between the ETWFE and the imputation estimate likely stems from the fact that the imputation estimator estimates the fixed effects only using the control group, while the ETWFE estimator uses information on both the control and the treatment group.[27]

Jackknife bias correction. Next, we consider potential incidental parameter problems, which may arise in non-linear estimators in gravity settings (Jochmans, 2017; Pfaffermayr, 2021; Weidner and Zylkin, 2021). Nagengast and Yotov (2023) show using a Monte Carlo simulation that the standard errors of the ETWFE regression can be downward biased in the gravity setting, which might be remedied using a split-panel jackknife bias correction. Therefore, here we jackknife estimates of both TWFE (column (4)) and ETWFE models (column (5)). The resulting point estimates are virtually unchanged in both cases. The standard errors, however, both increase considerably. A formal statistical test (Z-test) of the difference between the TWFE and ETWFE coefficients that takes the covariance between the estimates into account yields a p-value of 0.1039.

Alternative standard error clustering. In our next experiment, we investigate the robustness of our results to alternative clusterings of the standard errors. The results for our baseline estimates in Table 3 were clustered by country pair. Alternatively, the results in column (6) are clustered by exporter, importer, and year. While the standard error almost doubles, the significance of the coefficient estimate remains essentially unchanged. We

[26] While we do not report standard errors for the imputation estimator, they could likely be obtained using a bootstrap procedure.

[27] See Nagengast and Yotov (2023) for a discussion.

conclude that while alternative standard error clustering might affect the magnitude of the standard errors, the changes are not large and do not effect the significance of our main results.

Ordinary Least Squares. Lastly, we provide results using the OLS estimator instead of the PPML estimator. For comparison, columns (7) and (8) reports results from a TWFE model and an ETWFE model, respectively. We draw two conclusions from these additional results. First, in our setting, the OLS estimator yields larger coefficients irrespective of the underlying model under consideration. Second, the estimated effect of the ETWFE model is still larger than that of the TWFE model. Therefore, our main conclusion that the TWFE estimate is downward biased remains unchanged when considering the OLS estimator instead of the PPML estimator.

6.3 DiD- and ETWFE-specific assumptions

With regard to DiD- and ETWFE-specific assumptions, we consider the effect of methodological choices related to the definition of treatment onset, the potential effect of the presence of other treatments, the exclusion of long-term effects, a variant relaxing the parallel trends assumption, and alternative weighting schemes for obtaining the aggregate treatment effect coefficient. Our findings are reported in Table 6.

Treatment onset. Throughout the analysis, we assumed that the treatment 'onset' occurred two years before EU accession to capture anticipation effects. Here, we consider two alternative strategies. First, we follow Wooldridge (2021, 2023) who suggests to omit periods with anticipation effects such that the pre-treatment period then only consists of observations without anticipation effects. The resulting coefficient in column (1) equals 0.460, i.e., it is slightly larger than our baseline estimate from column (1) in Table 3. This is as expected given that the anticipation effects were smaller than the remaining treatment effects. Second, we consider using the year of EU accession as the treatment onset by shifting the treatment 'onset' back again by two years. The anticipation effects then become part of the pre-treatment period. The corresponding coefficient in column (2) is close to our baseline estimate, albeit slightly larger. Overall, we conclude that our main results are robust to alternative approaches to addressing anticipation effects.

Exclusion of other treatments. One potential concern when considering treatment effects over a long time period is that other policy interventions may be implemented in the meantime. As a result, one may confound the effects of these new policy interventions with the original treatment under consideration. In our case, the most salient event is the introduction of the European Monetary Union (EMU). To exclude potential effects of the Euro, we omit all post-treatment periods of trade between EMU members from the estimation sample taking the different accession dates of individual countries into the euro area into account. Column (3) shows that the resulting coefficient (0.403) is only very slightly smaller than our baseline estimate suggesting that EMU membership does not have a significant impact on our main result.

Exclusion of long-term effects. Our baseline estimate contains post-treatment effects for early cohorts that span several decades (Figure 7). While common in the literature, this requires the parallel trends assumption to hold over a very long time period. As noted in the previous paragraph, this also requires there to be no other treatments or policy interventions occurring in the meantime. If one is not willing to make these assumptions, a simple solution is to exclude observations far away from treatment onset. In column (4), we exclude treatment effects 20 years after EU accession. The corresponding regression still yields a large and positive coefficient (0.322). As one would expect, however, it is substantially smaller than our baseline estimate given that we now exclude many periods that are associated with the largest treatment effects.

Time-constant covariates. In Section 5.1, we did not find evidence for violations of the identifying assumptions apart from an anticipation effect in the year before EU accession. In other settings, the parallel trends assumption may only hold when conditioning on covariates. Therefore, next, we consider an extension of the ETWFE model, in which the cohort-time-specific treatment effects are allowed to vary by time-constant covariates (Wooldridge, 2023). This experiment is similar in spirit to the approach by Callaway and Sant'Anna (2021) who consider settings when the parallel trends assumption only holds after conditioning on observed covariates by using outcome regression, inverse probability weighting, and doubly-robust estimands. As time-constant covariate, we consider the bilateral distance between countries in column (5), which was the most salient difference between treated and never-treated country pairs (Table 2), and which is also the most robust and widely used proxy for bilateral trade costs in the gravity literature Anderson and van Wincoop (2004). The resulting coefficient is virtually unchanged. Note that this specification comes at the cost of estimating a considerably larger number of coefficients. We conclude that our main result is robust to alternative and less-demanding identifying assumptions.

Alternative weighting schemes. Lastly, we consider alternative weighting schemes of the cohort-year effects when computing the aggregate treatment effect. For our baseline, we give every post-treatment observation the same weight. For robustness, we give every cohort (column (7)), every event year (column (8)), or every cohort-year cell (column (9)) the same weight. In addition, we also consider a variant that uses trade weights (column (10)). However, all four alternatives yield aggregate treatment effects that are considerable larger than our baseline estimate since they place larger weights on cohort-year cells with larger treatment effects. Therefore, we conclude that our results are robust to using alternative weighting schemes.

Table 6: Robustness with regard to different DiD- and ETWFE-specific assumptions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$EU_{ij,t}$	0.460***	0.433***	0.403***	0.322***	0.426***	0.536***	0.840***	0.665***	0.828***
	(0.040)	(0.037)	(0.034)	(0.034)	(0.041)	(0.053)	(0.085)	(0.069)	(0.076)
Sample / treatment									
Omit anticipation periods	Yes								
Shift treatment by 2 years		Yes							
Exclude EMU			Yes						
Exclude effects > 20 years				Yes					
Covariate interactions									
ln Distance					Yes				
Weights	Obs	Obs	Obs	Obs	Obs	Cohort	Year	Cohort × year	Trade
Observations	415,743	417,131	413,203	414,099	416,977	417,131	417,131	417,131	417,131
Exporters	90	90	90	90	90	90	90	90	90
Importers	90	90	90	90	90	90	90	90	90
Years	70	70	70	70	70	70	70	70	70
Coefficients	244	244	239	142	589	260	260	260	260
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents PPML regression results using variants of the ETWFE model (equation (1)), for which the cohort-year-specific treatment effects were aggregated to obtain an aggregate treatment effect estimate. The dependent variable is exports which vary over the exporter-importer-year dimension. Column (1) omits the year immediately before EU accession to exclude periods with anticipation effects following Wooldridge (2023). Similarly, column (2) shifts EU accession forward by one year. Column (3) omits all post-treatment periods of trade between EMU members. Column (4) excludes treatment effects 20 years after EU accession. Column (5) includes interactions between ln Distance and cohort-year effects thereby relaxing the parallel trend assumption (Callaway and Sant'Anna, 2021; Wooldridge, 2023). Columns (6)-(9) report results for alternative weighting schemes for the cohort-year effects of column (2) of Table 3 that differ from the approach in all other specifications that give every post-treatment observation ('Obs') the same weight. Instead, the robustness checks give every cohort (column (6)), every event year (column (7)), every cohort-year (column (8)) the same weight or use trade weights (column (9)). 'Coefficients' reports the number of estimated coefficients apart from the fixed effects. Standard errors in parentheses are clustered by country pair. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Stata code:

```
jwdid trade if exclude!=1, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) // Column (1); exclude equal to one in year 1 and 2
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU_2) method(ppmlhdfe) fevar(idt_ci idt_cj) // Column (2); FT_EU_2 = FT_EU_2+2 for treated
jwdid trade if EMU==0, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) // Column (3); EMU is a dummy indicating EMU membership
jwdid trade if FT_EU==0 | year<FT_EU+20, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) // Column (4)
jwdid trade DIST, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) // Column (5)
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) // Columns (6)-(9)
estat simple, predict(xb) // Columns (1)-(5)
estat simple [pweight=w_g], predict(xb) // Column (6); w_g are cohort weights
estat simple [pweight=w_t], predict(xb) // Column (7); w_t are event year weights
estat simple [pweight=w_gt], predict(xb) // Column (8); w_gt are cohort-year weights
estat simple [pweight=w_trade], predict(xb) // Column (9); w_trade are trade weights
```

7 Conclusion

The European Union is the world's largest single market area and it is considered the most successful international integration effort to date. Yet, there are surprisingly few studies that have attempted to quantify the effects of EU membership on the international trade among its members. Moreover, recent developments in the econometrics literature suggest that the estimates from the relatively few existing papers that estimate the impact of the EU Single Market on trade with standard TWFE techniques may be biased due to 'forbidden comparisons' that (mis)use already-treated units in the control group.

Against this backdrop, we made two contributions to the existing literature. First, we combined established methods from the structural gravity literature with recent heterogeneity-robust difference-in-differences methods and we deployed several data sources to re-evaluate the impact of membership in the EU Single Market on international trade. Our results reveal that the effects of the EU have been very strong on average and that the Single Market has benefited all of its members. In addition, we document certain asymmetries in the effects of the EU in favor of the exports from 'old' to 'new' members than in the opposite direction. These findings point to possible areas for policy intervention. We also perform a battery of robustness experiments and sensitivity checks, which broadly confirm and reinforce our main conclusions.

On the methods front, we introduce a new, fast, and flexible estimation command that combines leading estimation techniques from the gravity literature with recent methods from the heterogeneity-robust DiD literature, and fast computation approaches to handle high-dimensional fixed effects in linear and non-linear settings. While our current analysis focused on the effects of EU membership on trade, we expect that the impact and use of our command will be much broader. Naturally, it can be used to study the impact of other determinants of trade flows e.g., WTO membership, Currency unions, etc.) with the gravity model of trade. In addition, it can be used to estimate gravity equations for other bilateral flows and relationships (e.g., migration, FDI, patents, international M&As, etc). Finally, it can be also be used in more general (non-gravity) panel and cross-sectional settings.

References

Agnosteva, Delina E., James E. Anderson, and Yoto V. Yotov, "Intra-national Trade Costs: Assaying Regional Frictions," *European Economic Review*, 2019, *112* (C), 32–50.

Ahmadi, Iman, Nadia Abou Nabout, Bernd Skiera, Elham Maleki, and Johannes Fladenhofer, "Overwhelming targeting options: Selecting audience segments for online advertising," *International Journal of Research in Marketing*, 2024, *41* (1), 24–40.

Anderson, James E. and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 2003, *93* (1), 170–192.

Anderson, James E. and Eric van Wincoop, "Trade Costs," *Journal of Economic Literature*, 2004, *42* (3), 691–751.

Anderson, James E. and Yoto V. Yotov, "Quantifying the Extensive Margin(s) of Trade: The Case of Uneven European Integration," CESifo Working Paper Series 9822, CESifo 2022.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodriguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Baier, Scott L. and Jeffrey H. Bergstrand, "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 2007, *71* (1), 72–95.

_ , Yoto V. Yotov, and Thomas Zylkin, "On the Widely Differing Effects of Free Trade Agreements: Lessons from Twenty Years of Trade Integration," *Journal of International Economics*, 2019, *116*, 206–226.

Baldwin, Richard E. and Daria Taglioni, "Gravity for Dummies and Dummies for Gravity Equations," *NBER Working Paper No. 12516*, 2006.

Bergstrand, Jeffrey H., Mario Larch, and Yoto V. Yotov, "Economic Integration Agreements, Border Effects, and Distance Elasticities in the Gravity Equation," *European Economic Review*, 2015, *78*, 307–327.

Borusyak, Kirill and Xavier Jaravel, "Revisiting Event Study Design," *Harvard University Working Paper*, 2017.

_ , _ , and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," *The Review of Economic Studies*, forthcoming.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230. Themed Issue: Treatment Effect 1.

_ , Andrew Goodman-Bacon, and Pedro H. C Sant'Anna, "Difference-in-differences with a Continuous Treatment," Working Paper 32117, National Bureau of Economic Research February 2024.

_ , _ , and _ , "Event-Studies with a Continuous Treatment," Working Paper 32118, National Bureau of Economic Research February 2024.

Cheng, I-Hui and Howard J. Wall, "Controlling for Heterogeneity in Gravity Models of Trade and Integration," *Federal Reserve Bank of St. Louis Review*, 2005, *87* (1), 49–63.

Correia, Sergio, "A Feasible Estimator for Linear Models with Multi-Way Fixed Effects," 2016.

— , "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator," *working paper*, 2016.

— **, Paulo Guimaraes, and Tom Zylkin**, "Fast Poisson Estimation with High-dimensional Fixed Effects," *Stata Journal*, March 2020, *20* (1), 95–115.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer, "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 2012, *79* (2), 581–608.

Dai, Mian, Yoto V. Yotov, and Thomas Zylkin, "On the Trade-diversion Effects of Free Trade Agreements," *Economics Letters*, 2014, *122* (2), 321–325.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, September 2020, *110* (9), 2964–2996.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-way fixed effects and differences-in-differences estimators with several treatments," *Journal of Econometrics*, 2023, *236* (2), 105480.

de Chaisemartin, Clément, Xavier D'Haultfoeuille, Félix Pasquier, and Gonzalo Vazquez-Bare, "Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period," 2023.

Dhaene, Geert and Koen Jochmans, "Split-panel Jackknife Estimation of Fixed-effect Models," *The Review of Economic Studies*, 02 2015, *82* (3), 991–1030.

Dhingra, Swati, Hanwei Huang, Gianmarco Ottaviano, João Paulo Pessoa, Thomas Sampson, and John Van Reenen, "The costs and benefits of leaving the EU: trade effects," *Economic Policy*, 10 2017, *32* (92), 651–705.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Egger, Peter H. and Filip Tarlea, "Multi-way Clustering Estimation of Standard Errors in Gravity Models," *Economics Letters*, 2015, *134* (C), 144–147.

— **and Sergey Nigai**, "Structural Gravity with Dummies Only: Constrained ANOVA-Type Estimation of Gravity Models," *Journal of International Economics*, 2015, *97*, 86–99.

— **, Mario Larch, and Yoto V. Yotov**, "Gravity Estimations with Interval Data: Revisiting the Impact of Free Trade Agreements," *Economica*, 2022, *89* (353), 44–61.

Eurostat, "Statistics Explained," *Intra-EU trade in goods main features*, 2023.

Felbermayr, Gabriel, Mario Larch, Erdal Yalcin, and Yoto V. Yotov, "On the Heterogeneous Trade and Welfare Effects of GATT/WTO Membership," *CESifo Working Paper No. 8555*, 2020.

Fontagne, Lionel and Yoto Yotov, "Reassessing the impact of the Single Market and its ability to help build strategic autonomy," Technical Report, Unpublished Manuscript 2024.

Fontagné, Lionel, Michael Freudenberg, and Nicolas Péridy, "Intra-Industry Trade and the Single Market: Quality matters," *CEPR Discussion Paper*, 1998, (1959).

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár, "Contamination Bias in Linear Regressions," Working Paper 30108, National Bureau of Economic Research June 2022.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277. Themed Issue: Treatment Effect 1.

Gurevich, Tamara and Peter Herman, "The Dynamic Gravity Dataset: 1948-2016," 2018. USITC Working Paper 2018-02-A.

Hamilton, C.B. and L.A. Winters, "Opening Up International Trade in Eastern Europe," Papers 511, Stockholm - International Economic Studies 1992.

Head, Keith and Thierry Mayer, "Non-Europe: The magnitude and causes of market fragmentation in the EU," *Review of World Economics (Weltwirtschaftliches Archiv)*, June 2000, *136* (2), 284–314.

Hull, Peter, "Estimating Treatment Effects in Mover Designs," 2018.

Iliev, Dragomir, Galin Stefanov, and Yoto Yotov, "Estimating Bulgaria'S Trade Borders With The Eu An Application Of The Empirical Gravity Model Of Trade," *Business Management*, 2016, *4*, 1–3.

Jochmans, Koen, "Two-Way Models for Gravity," *The Review of Economics and Statistics*, 2017, *99* (3), 478–485.

Larch, Mario and Yoto V. Yotov, "Estimating the Effects of Trade Agreements: Lessons From 60 Years of Methods and Data," Unpublished Manuscript, Bayreuth and Drexel 2022.

_ **, José-Antonio Monteiro, Roberta Piermartini, and Yoto Yotov**, "On the Effects of GATT/WTO Membership on Trade: They are Positive and Large After All," School of Economics Working Paper Series 2019-4, LeBow College of Business, Drexel University May 2019.

Leppert, Rebecca, "How exactly do countries join the EU?," Technical Report, Short Reads, The Research Center 2022.

Liebersohn, Jack, "How does competition affect retail banking? Quasi-experimental evidence from bank mergers," *Journal of Financial Economics*, 2024, *154*, 103797.

Mou, Hongyu and Yiqing Xu, "PANELVIEW: Stata module to visualize panel data," Statistical Software Components, Boston College Department of Economics January 2022.

Nagengast, Arne and Yoto V. Yotov, "Staggered Difference-in-Differences in Gravity Settings: Revisiting the Effects of Trade Agreements," School of Economics Working Paper Series 2023-6, LeBow College of Business, Drexel University 2023.

Olivero, María Pía and Yoto V. Yotov, "Dynamic Gravity: Endogenous Country Size and Asset Accumulation," *Canadian Journal of Economics*, 2012, *45* (1), 64–92.

Papazoglou, Christos, Eric J. Pentecost, and Helena Marques, "A Gravity Model Forecast of the Potential Trade Effects of EU Enlargement: Lessons from 2004 and Path-dependency in Integration," *The World Economy*, 2006, *29* (8), 1077–1089.

Pfaffermayr, Michael, "Gravity models, PPML estimation and the bias of the robust standard errors," *Applied Economics Letters*, 2019, *26* (18), 1467–1471.

_ , "Confidence Intervals for the Trade Cost Parameters of Cross-section Gravity Models," *Economics Letters*, apr 2021, *201*, 109787.

_ , "Cross-section Gravity Models, PPML Estimation and the Bias Correction of the Two-way Cluster-robust Standard Error," *unpublished manuscript*, 2022.

Roth, Jonathan, "Interpreting Event-Studies from Recent Difference-in-Differences Methods," 2024.

_ **and Pedro H. C. Sant'Anna**, "When Is Parallel Trends Sensitive to Functional Form?," *Econometrica*, 2023, *91* (2), 737–747.

Santos Silva, João M.C. and Silvana Tenreyro, "The Log of Gravity," *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199. Themed Issue: Treatment Effect 1.

Weidner, Martin and Thomas Zylkin, "Bias and Consistency in Three-way Gravity Models," *Journal of International Economics*, 2021, *132*, 103513.

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge, Massachusetts: The MIT Press, 2010.

Wooldridge, Jeffrey M, "Two-way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators," *Available at SSRN 3906345*, 2021.

_ , "Simple Approaches to Nonlinear Difference-in-Differences with Panel Data," *The Econometrics Journal*, September 2023, *26* (3), C31–C66.

Yotov, Yoto V., "On the Role of Domestic Trade Flows for Estimating the Gravity Model of Trade," *Contemporary Economic Policy*, 2022.

_ **, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch**, *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, Geneva: UNCTAD and WTO, 2016.

Supplementary Appendix

A.1 Details on the `jwdid` command

This section of the appendix provides further details on our estimation command and its options. Our departing point is the benchmark ETWFE estimator from the main text:

$$Y_{i,t} = \alpha + \sum_{g \in G} \sum_{s=g}^{T} \theta_{g,s} D_{g,s} + \xi_i + \xi_t + \varepsilon_{i,t}, \tag{A1}$$

where $Y_{i,t}$ is the dependent variable, $D_{g,t}$ is a dummy that takes the value of 1 if the observation is in the treatment group g, on period t and 0 otherwise. G is a set that indicates at what time treatment started for all observations, and T is the last period of the analysis. ξ_i and ξ_t are sets of fixed effects for the individual and time dimensions, respectively.[A1] In this setup, the $\theta_{g,t}$ coefficients represent the average treatment effect that the treatment group g experiences at time t ($ATT(g,t)$). As described in Wooldridge (2021), allowing for a flexible specification of the $\theta_{g,t}$ avoids the problem of bad controls and negative weights that have been identified in the literature as potential problems in the estimation of DID models using traditional TWFE estimators. In addition, as demonstrated by Wooldridge (2023), the ETWFE estimator can be easily modified and implemented to allow for non-linear models, e.g., for cases where the dependent variable is binary (logit) or count data (poisson). The latter is particularly important for our purposes since, as argued earlier, the PPML estimator is the leading gravity estimator.

Against this backdrop, the benchmark syntax for the proposed command is as follows:

```
jwdid y, ivar(i) tvar(t) gvar(g)
```

Here `y` is the dependent variable, `ivar(i)` is used to identify the individual panel data dimension or unit, `tvar(t)` identifies the time dimension, and `gvar(g)` identifies the treatment group. Specifically, for observation i, g would take the value of zero if the unit is never treated (within the window of the analysis), and would take a value different from zero to indicate the year that treatment started

[A1] Often, one can use group fixed effects instead of individual fixed effects, and would still obtain numerically identical results in the linear model case.

for unit i. Following standard assumptions, this specification assumes that the treatment is an absorbing status, meaning that once a unit is treated, it remains treated for the rest of the analysis.

By default, `jwdid` will estimate the baseline model Equation A1 using the `reghdfe` command (Correia, 2016a), assuming clustered standard errors at the `i` level. If a different level is desired, the user can specify the `cluster(cvar)` option. While the command does not impose the assumption that the data is a panel, the methodology is designed to work with panel data. In case of repeated cross-sectional data, one should instead use the following syntax:

```
jwdid y, tvar(t) gvar(g) [cluster(cvar)]
```

By excluding `ivar(i)`, the command assumes data is a repeated cross-section, proceeding to include group fixed effects only. The `cluster(cvar)` option is not required, but can be used to request the standard errors to be clustered at the level `cvar` when using `reghdfe` or `ppmlhdfe`. Specifically, this command will estimate the following model:

$$Y_{i,t} = \alpha + \sum_{g \in G} \sum_{s=g}^{T} \theta_{g,s} D_{g,s} + \xi_g + \xi_t + \varepsilon_{i,t} \tag{A2}$$

The model specifications in equations (A1) and (A2) make the implicit assumption that parallel trends are satisfied, using all never treated and not-yet treated observations as controls (not included category) for the identification of treatment effects. If one instead wants to relax this assumption, the user can specify the option `never`:

```
jwdid y, ivar(i) tvar(t) gvar(g) never
```

which will estimate the following model:

$$Y_{i,t} = \alpha + \sum_{g \in G} \sum_{s=t_0}^{g-2} \theta_{g,s}^{pre} D_{g,s} + \sum_{g \in G} \sum_{s=g}^{T} \theta_{g,s}^{post} D_{g,s} + \xi_i + \xi_t + \varepsilon_{i,t}. \tag{A3}$$

In principle, this is the same as the strategy proposed by Sun and Abraham (2021), allowing for full heterogeneity across all groups and all event time periods. This specification is also numerically identical to the one proposed by Callaway and Sant'Anna (2021), for the case where there are no covariates. In this case, the only observations that are used as controls are the ones that were never treated. In this specification, all $\theta^{pre}_{g,t}$ can be used to test for the parallel trends assumption, and all $\theta^{post}_{g,t}$ can be used to estimate the treatment effects.

A.1.1 Extensions and Additional Command Options

In this section, we extend the command by introducing options that would enable us to capture important methodological aspects from the gravity literature and beyond.

Nonlinear models. As described in Wooldridge (2023), the standard ETWFE model described in Equation A1 or Equation A2 identifies the average treatment effect by imposing a linear parallel trends assumption. However, such assumption may not be valid in cases, such as when the dependent variable follows some limited distribution. Roth and Sant'Anna (2023) discuss a similar problem, stating that the choice of transformation of the dependent variable is crucial for the identification of the average treatment effect, and only under certain conditions would the ATT be identified for any transformation.

In this regard, Wooldridge (2023) proposes that the linear ETWFE models can be adapted to allow for non-linear models, by simply imposing the linear parallel trends assumption only on the latent variable of the model, but not the outcome itself. Consider the following transformation of the model defined by Equation A3:

$$E(Y_{i,t}|X,\xi_i,\xi_t) = H\left(\alpha + \sum_{g\in G}\sum_{t=t_0}^{g-2}\theta^{pre}_{g,s}D_{g,s} + \sum_{g\in G}\sum_{s=g}^{T}\theta^{post}_{g,s}D_{g,s} + \xi_i + \xi_t\right). \tag{A4}$$

This specification focuses on identifying the conditional expected value of the outcome of interest as a function of the treatment status, and the individual and time fixed effects. If we assume the $H()$ is the identify function, we would be back to the linear model described by Equation A3. However, if we assume that $H()$ is a non-linear function, like in the linear exponential family for poisson, or logistic for logit models, we could estimate the average treatment effect under different assumptions, imposing only a linear parallel trends assumption on the latent variable of the model.

The `jwdid` command allows the user to specify the `method()` option to estimate models described by Equation A4, where one would specificy the regression model to be estimated, followed by the options associated with that model. For example, if we were interested in estimating a poisson model, we would use the following syntax:

```
jwdid y, ivar(i) tvar(t) gvar(g) never method(poisson)
```

There are no restrictions on the type of `method` one can use with the `jwdid` command, but it has not been tested with all possible models. The user should be aware that the `method()` option is passed directly to the model estimation step, and the user should be familiar with the syntax of the model being estimated. Other 'methods' options can be used as well following the syntax method(cmd, cmdoptions).

It should be noted that when estimating non-linear models with a large number of fixed effects, one may face an incidental parameters problem. This is not generally a problem for the linear case, because the parameters of interest can be identified without explicitly estimating the fixed effects, using, for example the within transformation. However, with the exception of poisson models, fixed effects are generally estimated subject to the incidental parameter problem. To reduce the impact of this problem, whenever `method()` is specified `jwdid` will incorporate *group* fixed effects, instead of *individual* fixed effects.

For the linear case with balanced data, using *group* instead of *individual* fixed effects provides numerically identical results. If the panel is unbalanced, the results will not be identical. In such cases, the option `corr` will create additional variables that address the difference. In the case of non-linear models, the best solution is to use *group* fixed effects. However, if one is interested in poisson models, the alternative to group fixed effects is to use `ppmlhdfe` (Correia et al., 2020).[A2] This is the state-of-the-art estimator for poisson models with fixed effects, and it is the recommended estimator for trade analysis.

Additional covariates. As described in Wooldridge (2021), it is possible to include covariates in the model, by simply adding corrections that enable to easily identify the average treatment effect. However, following the literature on DID models, the implicit assumption is that covariates are

[A2] The correction implemented with `corr` is not useful to recover the coefficients from `ppmlhdfe` using `poisson` command time-invariant. `jwdid` does not impose any assumption on the covariates, but the user should be aware of the implications. In general, when covariates are considered, the model of interest is similar to Equation A3, but adjusted for covariates:

$$
\begin{aligned}
Y_{i,t} = \alpha &+ \sum_{g\in G}\sum_{s=t_0}^{g-1}\theta^{pre}_{g,s}D_{g,s} + \sum_{g\in G}\sum_{s=g}^{T}\theta^{post}_{g,s}D_{g,s} + \sum_{g\in G}\sum_{s=t_0}^{g-1}D_{g,s}\tilde{x}_i'\beta^{pre}_{g,t} + \sum_{g\in G}\sum_{s=g}^{T}D_{g,s}\tilde{x}_i'\beta^{post}_{g,t} \\
&+ x_i'\beta + \sum_{t=t_0}^{T}D_{i,t}x_i'\beta_t + \sum_{g\in G}D_{i,g}x_i'\beta_g + \xi_i + \xi_t + \varepsilon_{i,t},
\end{aligned}
\tag{A5}
$$

where $D_{i,t}$ is a dummy variable that is equal to 1 if period is equal to t, and $D_{i,g}$ is a dummmy variable that is equal to 1 if the group membership is equal to g, and zero otherwise. $\tilde{x}_i$ are the within cohort and period demeaned variables. When using `reghdfe` or `ppmlhdfe` the term $\sum_{g\in G} D_{i,g}x_i'\beta_g$ is omitted if all covariates are time constant.

When one uses the default specification, the parameters $\theta^{pre}_{g,t}$ and $\theta^{post}_{g,t}$ still identify the average treatment effect for group g at time t. In addition, the β parameters identify covariate heterogeneity effect. One could also use the untransformed covariates in Equation A5, and still be able to obtain the same group/time specific treatment effects with the post estimation commands. However, Theta parameters can no longer be interpreted as treatment effects. From the user perspective, `jwdid` would simply need to be called as follows:

```
jwdid y x, ivar(i) tvar(t) gvar(g) never [xasis]
```

where `x` are all covariates of interest. If one uses the option `xasis`, the command will use the covariates without demeaning them, which may save some computation time.

Treatment heterogeneity. As may be apparent from Equation A5, the number of estimated parameters can grow quickly with the number of groups/cohorts, time periods, and covariates. This could lead to increasing computational burden of the estimation. An alternative, which is already implemented via `xthdidregress` and `hdidregress` in Stata 18, is to estimate a model that reduces the heterogeneity of the treatment effects. Specifically, it allows treatment effects to vary across cohorts, across absolute time, or across relative time. For the case without covariates, the specification of Equation A3 can be modified to impose the heterogeneity restrictions as follows.

- Time heterogeneity:

$$Y_{i,t} = \alpha + \sum_{t=t_0}^{T} \theta_t^{pre} D_{i,t,pre} + \sum_{t=t_0}^{T} \theta_t^{post} D_{i,t,post} + \xi_i + \xi_t + \varepsilon_{i,t}. \tag{A6}$$

- Cohort heterogeneity:

$$Y_{i,t} = \alpha + \sum_{g \in G} \theta_g^{pre} D_{i,g,pre} + \sum_{g \in G} \theta_g^{post} D_{i,g,post} + \xi_i + \xi_t + \varepsilon_{i,t}. \tag{A7}$$

- Event (Relative Time) heterogeneity:

$$Y_{i,t} = \alpha + \sum_{e=E_{min}}^{-2} \theta_e D_{i,e} + \sum_{e=0}^{E_{max}} \theta_e D_{i,e} + \xi_i + \xi_t + \varepsilon_{i,t}, \tag{A8}$$

where $D_{i,t,pre}$ and $D_{i,t,post}$ are dummies that take the value of 1 if observation i, which is part of an eventually treated group, is not yet treated or is already treated at time t, respectively. $D_{i,g,pre}$ and $D_{i,g,post}$ are dummies that take the value of 1 if observation i belongs to group g and is not yet treated or is already treated at time t, respectively. $D_{i,e}$ is a dummy that takes the value of 1 if observation i is e periods relative to when treatment started. E_{min} and E_{max} are the minimum and maximum event periods, possible. The *pre* coefficients are only considered when the `never` option is used.

The `jwdid` command allows the user to specify each one of these restrictions using the `hettype()` option.

```
jwdid y, ivar(i) tvar(t) gvar(g) hettype(option)
```

where `option` can be `time`, `cohort`, or `event`. If no option is selected, the command will estimate the model described by Equation A3, which is the equivalent to allowing for full cohort-time heterogeneity.

Other options. As described earlier, when covariates are considered in the model, the default option is to interact all covariates (or the demeaned transformations) with the same level of covariate heterogeneity. Sometimes, however, one may not be interested in estimating the same level of heterogeneity for all covariates. It may be possible, for example, to consider separate sets of covariates that could be interacted only with the time or the group dimensions. Specifically, assume there are no variables we wish to consider for the treatment heterogeneity, but instead consider three sets of covariates: x^{EX} or variables we wish to incorporate without further interactions, x^{T} or variables that would be interacted with the time variables only, and x^{G} or variables that would be interacted with group indicators only. In this case, the setup would be:

$$
\begin{aligned}
Y_{i,t} = \alpha &+ \sum_{g \in G} \sum_{s=t_0}^{g-1} \theta_{g,s}^{pre} D_{g,s} + \sum_{g \in G} \sum_{s=g}^{T} \theta_{g,s}^{post} D_{g,s} \\
&+ x_i'^{EX} \beta + \sum_{t=t_0}^{T} D_{i,t} x_i'^{T} \beta_t + \sum_{g \in G} D_{i,g} x_i'^{G} \beta_g + \xi_i + \xi_t + \varepsilon_{i,t},
\end{aligned}
\tag{A9}
$$

which can be estimated using the following syntax:

```
jwdid y , ivar(i) tvar(t) gvar(g) never exogar(x_ex) xtvar(x_t) xgvar(x_g)
```

If covariates are included after `y`, they would still be treated following the specification of Equation A5, or following any of the heterogeneity restrictions described in earlier. An advanced version of this option is the inclusion of high-dimensional fixed effects (and interactions with fixed effects) that are different from the individual and time fixed effects. It is possible to request the inclusion of those types of fixed effects using the option `fevar()`, which is only valid if one is using the default estimator method `reghdfe` or `ppmlhdfe`. In both cases, the additional fixed effects (or interactions) are included in the estimation of the model without further interactions.

A.1.2 Post-estimation Options and Analysis

In this section we describe a series of post-estimation options, which may be useful for various purposes.

Aggregated treatment effects. After the estimation of the model, under the default options, one can use the coefficients $\theta^{post}_{g,t}$ as direct estimates of the group and time specific average treatment effects on the treated. However, one may also be interested in estimating aggregated ATTs for the aggregate, across groups or periods, or dynamic effects. Furthermore, when the underlying method is a non-linear model, these coefficients cannot be directly interpreted as the average treatment effect on the outcome, but only on the latent variable.

To accommodate such needs, the `jwdid` command comes along with the post estimation command `jwdid_estat` that can be used for that purpose. Internally, it uses the `margins` command to identify average treatment effects under the following algorithm:

1. Using the model estimates, predict the outcome of interest for all observations given the observed covariates and fixed effects. Call this $\hat{Y}_i(obs)$ or predicted outcome under the observed covariates. The model prediction could be the linear prediction, or the predicted probability in the case of logit models, or the predicted count in the case of poisson models.
2. Consider the specification Equation A5, and assume that all $\theta^{post}_{g,t}$ (and $\theta^{pre}_{g,t}$ if `never` is used), as well as all β^{post} and β^{pre} are zero, and predict the outcome of interest. Call this $\hat{Y}_i(0)$ or predicted outcome under the counterfactual scenario of no treatment.

In this case, the predicted Average Treatment Effect on the Treated for observation i is given by:

$$\widehat{ATT}_i = \hat{Y}_i(obs) - \hat{Y}_i(0)$$

This is zero for observations that were never treated, and nonzero for post-treatment treated observations.[A3]

From here, any aggregated average treatment effects can be calculated as follows:

$$AGGTE_r = \frac{\sum_i^N \widehat{ATT}_i \times w_{i,t} \times R_{i,t}}{\sum_i^N w_{i,t} \times R_{i,t}},$$

[A3] The case of pre-treatment treated observations may be assumed to be zero if the `never` option is not used. When `never` is used, pre-treatment ATTs can be used for testing the parallel trends assumption.

where $R_{i,t}$ takes the value of one whenever observation i fullfills the required conditions, and $w_{i,t}$ is the weight of the observation i at time t used in for the estimation model.[A4] $AGGTE$ is the aggregated average treatment effect on the treated given the conditions R. In general, there are four types of aggregations that are implemented in the `jwdid_estat/estat` command:

1. `estat simple`: This option calculates the average treatment effect on the treated for all observations that were treated at some point in time. The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t \geq g \text{ for observation } i \in g$$
$$R_{i,t} = 0 \text{ otherwise.}$$

2. `estat group`: This option calculates the average treatment effect for observations that were treated at time g. The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t \geq g_c \text{ and } i \in g_c \text{ for observation } i$$
$$R_{i,t} = 0 \text{ otherwise.}$$

   where g_c is a particular group/cohort of interest. `estat group` estimates this for all groups g in G.

3. `estat calendar`: This option calculates the average treatment effect at time t for all observations that were effectively treated at that point. The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t_c \geq g \text{ and } t = t_c \text{ for observation } i$$
$$R_{i,t} = 0 \text{ otherwise.}$$

   where t_c is a particular time of interest. `estat calendar` estimates this for all times t in T that has at least one unit that was treated.

4. `estat event`: This option calculates dynamic treatment effects, also known as event studies, using the period before treatment as the reference. When the condition 'never' is used, this approach can be used to estimate pre-treatment ATT's, which could be used for testing the

[A4] While this is the default option, `estat` command also allow you to provide other weights for aggregation parallel trends assumption. The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t - g = e_c \text{ and } e \neq -1 \text{ for observation } i$$

$$R_{i,t} = 0 \text{ otherwise}$$

where e_c is a particular event of interest. In contrast with the previous aggregations, if option `never` was used for estimation, one could also add the option `pretrend` to run a simple parallel trend test with the following null hypothesis:

$$H_0 : AGGTE_e = 0 \text{ for all } e < -1 \text{ vs } H_1 : AGGTE_e \neq 0 \text{ for some } e < -1$$

Failure to reject this hypothesis is evidence in support of parallel trends assumption. Otherwise, one can use `test` command to test for the significance of specific pre-treatment ATT's.[A5]

Weights. The default option for the estimation of the aggregated ATTs is to use the weights $w_{i,t}$ that were used in the estimation of the model. However, if the user wants to use different weights, it is possible to do using the following syntax:

```
estat [aggregation] [pw = weight]
```

Standard Errors. Wooldridge (2021) suggests than when one estimates standard errors for the aggregated ATTs, one should use `vce(unconditional)` option in `Stata`, to allow for uncertainty in the explanatory variables. `jwdid_estat`/`estat` does not use this approach by default, because it requires that the underlying command is able to produce scores for the estimated model. For example, if the model was estimated using `method(regress)`, the scores will be available, and unconditional standard errors for the aggregated ATTs can be estimated as follows:

```
estat [aggregation], [vce(unconditional)]
```

This is not possible if one uses `reghdfe` or `ppmlhdfe`.

[A5] It should be noticed that this test is different from the test proposed by Callaway and Sant'Anna (2021), which is based on testing *all* group/time specific ATT's, instead of the ones aggregated by event time.

Other aggregation restrictions. As described in earlier, the default aggregation considers all observations treated observations, imposing restrictions only in terms of time, group or event dimensions. However, one may be interested in imposing further restrictions that could leverage on the use of covariates in the model specification. For example, say that one estimates a DID model with covariates using the following syntax:

```
jwdid y i.dx, ivar(i) tvar(t) gvar(g) never
```

As usual, one could request the estimation of the aggregated ATTs for the whole sample:

```
estat [aggregation]
```

However, one would also be able to make the same estimation imposing the added restriction that the covariate `dx` is zero or one, using the option `orestriction()`:

```
estat [aggregation], orestriction(dx==0)
estat [aggregation], orestriction(dx==1)
```

The expression inside the parenthesis should be a valid Stata expression that is used when calculating the aggregated ATTs.

Storing, and saving results. After aggregate effects have been estimated, the user may want to store the results for further analysis or reporting. Because `estat` uses `margins` in the background, the default option is to store the results in memory as `r()` elements. Alternatively, `jwdid_estat`/`estat` allows the user to store the output of the command using three different options:

1. `estat [aggregation], post`: As with margins, option `post` will "post" the results of the command to be the current estimations in memory `e()`, which can be saved as usual for further analysis.

2. `estat [aggregation], estore(name)`: This option stores the results from the aggregation in memory as `name`. This is similar to using `estimation store name` after a regression command. The previously estimated results from `jwdid` are not overwritten.

3. `estat [aggregation], esave(filename)`: This option saves the results from the aggregation in a file `filename`, as a `ster` file, which can be used at a later point.

Plotting. After **time**, **group** or **event** aggregations are estimated, it is possible to request plotting those results using `estat plot`. The basic syntax is to type it after the aggregation command, with only minimal command specific options:

1. `estat plot, style(style)`: The option `style` allows the user to select the style of the plot. The default is using a `rspike` style, but `rarea`, `rcap` and `rbar` are also available. See `help twoway` for more information on the styles.

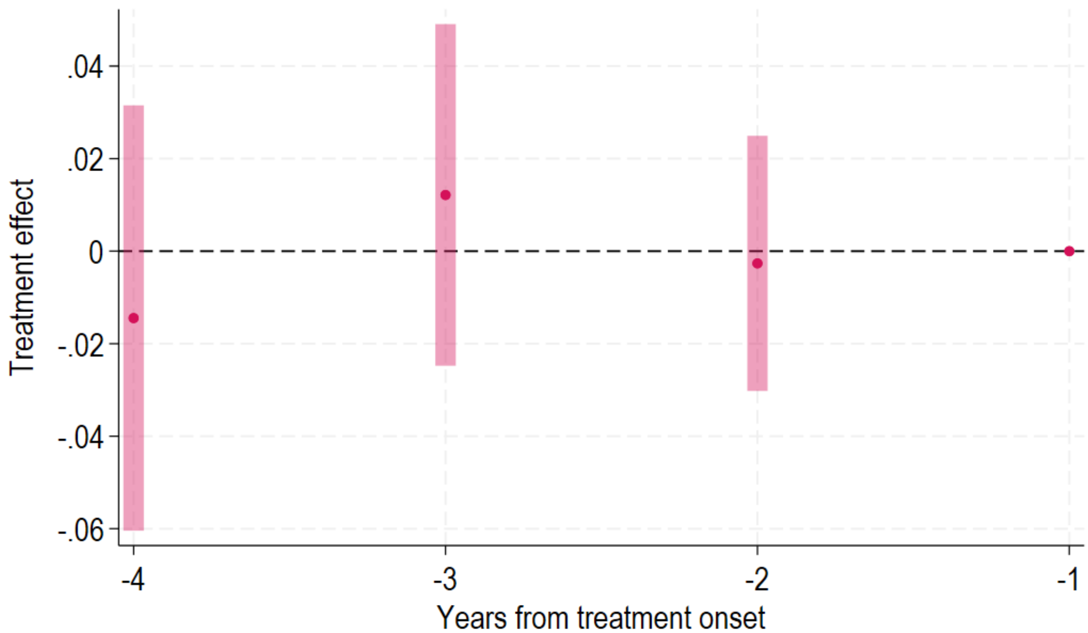
2. `estat plot,  pstyle1(str) color1(str) pstyle2(str) color2(str) lwidth1(str) lwidth2(str) barwidth1(str) barwidth2(str)`: The options `pstyle#`, `color#`, `lwidth#` and `barwidth#` can be used to alter the color and style of the lines in the plot for the pre and post treatment periods. Only `event` aggregation allows for `#2` options.

3. `estat plot, twoway_options`: Most other `twoway` graph option can be used after the `estat plot` command.

A.2 Additional results

A.2.1 Alternative pre-treatment test

Figure A1: Alternative pre-treatment test

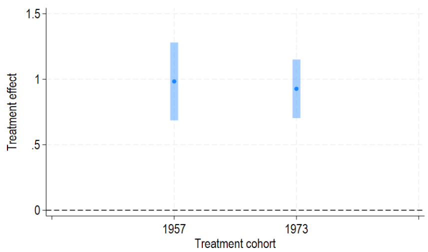


Notes: The figure reports pre-trend estimates from a PPML estimation of equation (1) augmented with cohort-year-specific placebo effects in the four years before treatment onset. To avoid misleading visual inferences, we follow the recommendation by Roth (2024) and put our pre-treatment estimates in a different plot from the post-treatment estimates. The regression is estimated with untreated observations only in the spirit of Borusyak et al. (forthcoming). The cohort-year-specific treatment effects were aggregated across cohorts to obtain event-time-specific treatment effect estimates. The coefficients before EU accession are jointly insignificant (p-value of 0.1490). 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

```
jwdid trade if FT_EU==0 | year<FT_EU, ivar(id_ci_cj) tvar(year) gvar(FT_EU) ///
           method(ppmlhdfe) fevar(idt_ci idt_cj) never
estat event, predict(xb) pretrend window(-4 -1)
estat plot, ytitle("Treatment effect") xtitle("Years from treatment onset") pstyle1(p2)
```

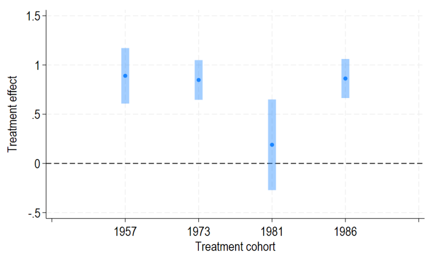
A.2.2 Effects by cohorts conditioning on same number of post-treatment years

Figure A2: Effects by cohorts conditioning on same number of post-treatment years

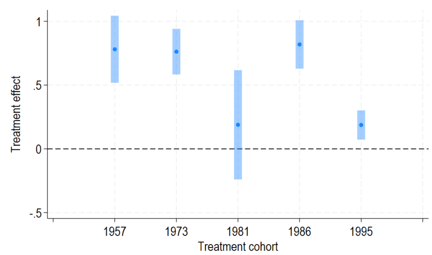


(a) 1973 cohort

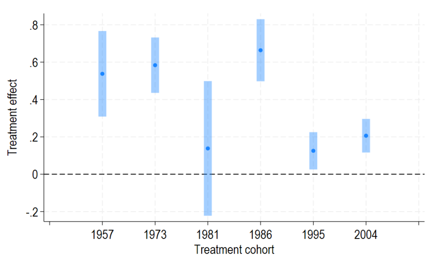
(b) 1981 cohort



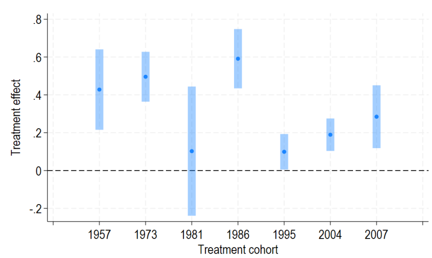
(c) 1986 cohort



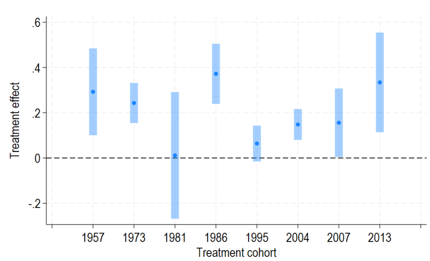
(d) 1995 cohort



(e) 2004 cohort



(f) 2007 cohort



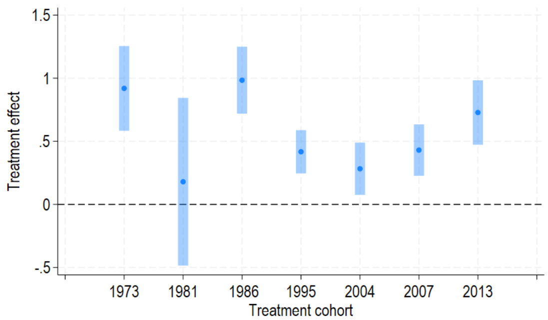
(g) 2013 cohort

Notes: The figure reports cohort-specific treatment effects conditioning on the same number of post-treatment years, for which the cohort-year effects (corresponding to the aggregate EU effect in column (2) of Table 3) were aggregated across event time. More specifically, every subfigure conditions on the number of post-treatment years of the last cohort to exclude composition effects: (a) 49 years. (b) 41 years. (c) 36 years. (d) 27 years. (e) 18 years. (f) 15 years. (g) 9 years. 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

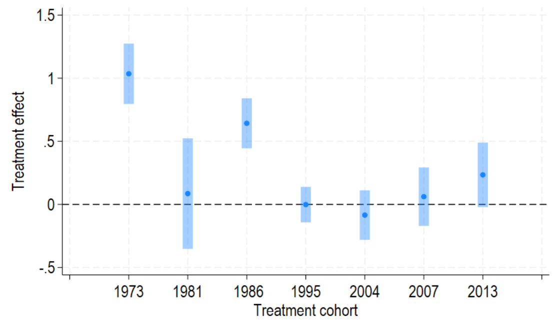
```
jwdid trade, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj)
local max_years = 2019-‘year’
estat group, predict(xb) ores(year<=FT_EU+‘max_years’  FT_EU<=‘year’)
estat plot, ytitle("Treatment effect",size(large)) xtitle("Treatment cohort",size(large))
```

A.2.3 Effects by cohort for each recency of EU membership

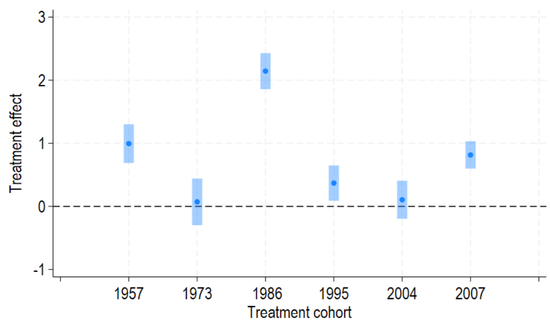
Figure A3: Effects by cohort for each recency of EU membership



(a) By cohort: Old-to-new



(b) By cohort: New-to-old



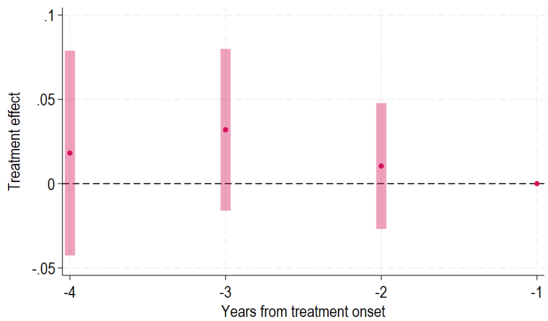
(c) By cohort: New-to-new

Notes: The figure reports cohort-specific treatment effects by recency of membership from a specification allowing for additional heterogeneity, for which the cohort-recency-year effects were aggregated across event time by cohort and recency of membership. 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

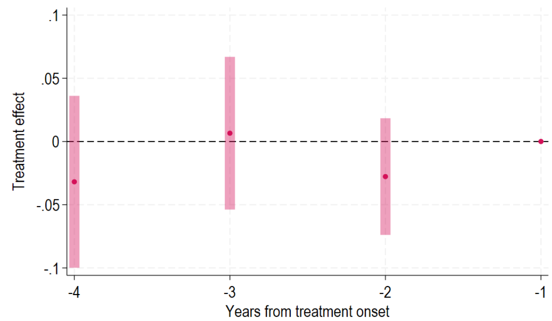
```
jwdid trade i.newold, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj)
estat group, predict(xb) ores(newold==‘i’)
estat plot, ytitle("Treatment effect",size(large)) xtitle("Treatment cohort",size(large))
```

A.2.4 Pre-treatment effects by recency of EU membership

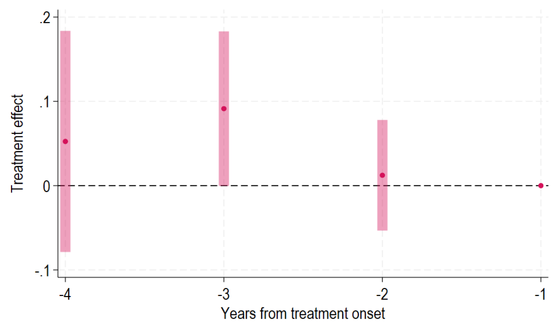
Figure A4: Pre-treatment effects by recency of EU membership



(a) Old-to-new



(b) New-to-old



(c) New-to-new

Notes: The figure reports cohort-specific treatment effects by recency of membership from a specification allowing for additional heterogeneity, for which the cohort-recency-year effects were aggregated across event time by cohort and recency of membership. 95% confidence intervals are shown using standard errors clustered by country pair.
Stata code:

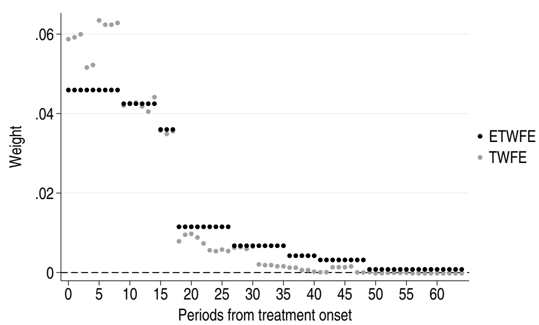
```
jwdid trade i.newold, ivar(id_ci_cj) tvar(year) gvar(FT_EU) method(ppmlhdfe) fevar(idt_ci idt_cj) never
estat event, predict(xb) pretrend window(-4 -1) ores(newold==‘i’)
estat plot, ytitle("Treatment effect") xtitle("Years from treatment onset") pstyle1(p2)
```

A.2.5 Implicit weights attached by the OLS TWFE estimator

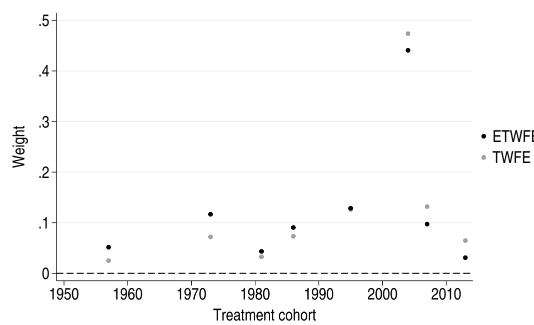
To provide evidence on the source of the difference between the ETWFE and the TWFE estimates, we consider the decomposition by de Chaisemartin and D'Haultfœuille (2020), which decomposes the OLS TWFE estimate into implicit weights attached by the estimator to cohort-year effects from the ETWFE estimator. First, we rerun the regression using OLS and we find a very similar pattern of results to the case of PPML (columns (7) and (8) in Table 5: The OLS TWFE estimate (0.579) is considerably smaller than the OLS ETWFE estimate (0.659). Given that the pattern of differences between the TWFE and ETWFE results is relatively similar for OLS and PPML, we hypothesize that the OLS decomposition results are likely informative about the PPML estimates as well.

Next, we compute the implicit weights attached by the OLS TWFE estimator to individual cohort-year effects of the OLS ETWFE estimator. The results are presented in Table A5 aggregated by event time (Figure A5a) and by (Figure A5b). This shows that the smaller coefficient of the OLS TWFE model results from two sources. First, the TWFE model places larger weights than the ETWFE model on short-term effects, which are associated with smaller treatment effects given that the EU effects gradually increased over time (Figure 5). Second, the TWFE also places larger weights than the ETWFE model on later cohorts, which are also associated with smaller treatment effects (Figure 6). This pattern of results is similar to what is found in the related literature on the trade effects of regional trade agreements (Nagengast and Yotov, 2023). Overall, we therefore conclude that in our case the TWFE is likely downward biased since it places larger weights on short-term effects and on the effect of late cohorts, which are both associated with smaller treatment effects.

Figure A5: Weights of OLS ETWFE and TWFE estimator



(a) By event time



(b) By cohort

Notes: The figure reports the weights used in the computation of the aggregate treatment effects of the ETWFE from column (7) of Table 5 in dark color ('ETWFE') along with the implicit weights attached by the OLS TWFE estimator from column (8) of Table 5 to cohort-year cells computed following de Chaisemartin and D'Haultfœuille (2020) in light color ('TWFE'). Panel (a) reports weights aggregated by event time. Panel (b) reports weights aggregated by cohort.