DEUTSCHE
BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 06/2026

From imposition to lifting: Estimating the effects of sanctions over their lifecycle

Ohyun Kwon
(Drexel University)

Arne J. Nagengast
(Deutsche Bundesbank)

Jangsu Yoon
(University of Kentucky)

Yoto V. Yotov
(Drexel University and CGPA)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank, Press and Public Relations Division, at the above address or via email: www.bundesbank.de/contact

Internet http://www.bundesbank.de

DOI https://doi.org/10.71734/DP-2026-6
ISBN 978-3-98848-063-7
ISSN 2941-7503

Deutsche Bundesbank Discussion Paper No 06/2026

From Imposition to Lifting: Estimating the Effects of Sanctions Over Their Lifecycle*

Ohyun Kwon
Drexel University

Arne J. Nagengast
Deutsche Bundesbank

Jangsu Yoon
University of Kentucky

Yoto V. Yotov
Drexel University
CGPA

Abstract

We combine the latest difference-in-differences estimators for treatments with exit and structural gravity literature to evaluate the effects of sanctions on trade, when they are in place and when they are lifted. Our analysis shows that sanctions reduce trade between senders and targets by 58%, with estimates from our preferred model 50% larger than those from traditional two-way fixed effects (TWFE) models. A bias decomposition highlights arbitrary weighting and contamination bias in TWFE estimates. Sensitivity checks confirm the robustness of our findings, emphasizing the relevance of these methods for gravity estimations, including trade, migration, foreign investment, and other bilateral flows.

Keywords: Difference-in-differences, Multiple treatments, Gravity, Sanctions, Trade.
JEL classification: C13, C23, F10, F13, F14, F51, H5, N4.

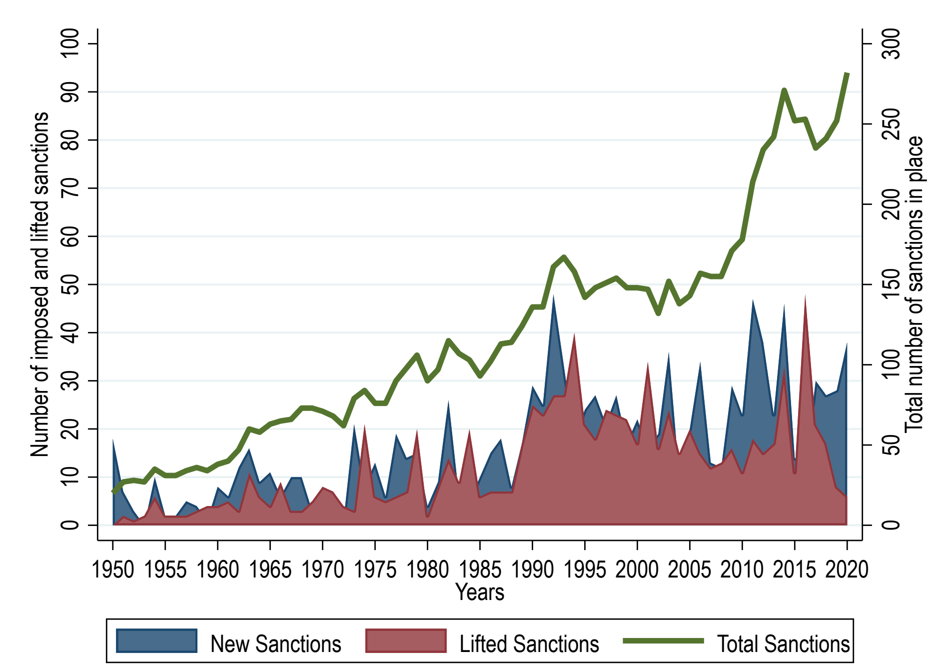
---

*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: Kwon (ok85@drexel.edu); Nagengast (arne.nagengast@bundesbank.de); Yoon (jangsu.yoon@uky.edu); Yotov (yotov@drexel.edu).

1 Introduction

Economic sanctions have become an integral part of contemporary statecraft. As depicted in Figure 1, which is based on the Global Sanctions Database (GSDB),[1] both the total number of sanctions that are in place as well as the number of newly imposed sanctions increased significantly between 1950 and 2020. Moreover, according to the GSDB, every country in the world has either been a sender or a target of sanctions. The increased use of sanctions has been paralleled by numerous efforts, from policymakers and academics alike, to evaluate their effectiveness and the costs of sanctions in terms of economic damage by focusing on various economic outcomes, e.g., growth, trade, health, mortality, etc.[2]

Figure 1: New, Lifted, and Existing Sanctions, 1950-2020



Notes: This figure is based on the third release of the Global Sanctions Database (GSDB), Syropoulos et al. (2024), and it depicts the evolution in the number of sanctions in place (the thick green line, which is based on the right Y-axis), the number of new sanction cases that are imposed (the blue area, measured on the left Y-axis), and the number of sanction cases that are lifted (the red area, measured on the left Y-axis) in the world in each year between 1950 and 2020.

[1]According to the New York Times, the GSDB is the most comprehensive dataset on economic sanctions (Koeze, 2022). The first edition of the data is due to Felbermayr et al. (2020), while the third release was constructed by Syropoulos et al. (2024). We offer further details on the GSDB in Section 3.

[2]A review of the large sanctions literature is beyond our objectives, however, a *Google Scholar* search on "sanctions" from October 19, 2025, delivered 41,700 results since 2021. This is more than twice as many results as a search on "trade agreements" over the same period, and it is quite different than the search over all years, which delivers 2 million results for sanctions and 4.2 million results for trade agreements.

Perhaps not surprisingly, most of the existing literature has focused on estimating the direct/partial effects of sanctions. Obtaining sound partial estimates is also crucial for quantifying the full, general equilibrium impact and effectiveness of sanctions.[3] However, as forcefully argued by the recent difference-in-differences (DiD) literature on staggered treatment adoption (de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2024), the established methods to estimate the effects of sanctions, which fall under the broad category of *"two-way fixed effects (TWFE) models"* (Borusyak et al., 2024; Wooldridge, 2025) may deliver biased estimates in the presence of treatment effect heterogeneity, which is very likely present in the case of sanctions.[4] Furthermore, unlike most other international policies that can be described and evaluated as treatments with entry only, sanctions are not only imposed but also very often lifted. In fact, as visualized in Figure 1, the number of lifted sanction cases is comparable to the number of new sanction cases and, according to the GSDB, the average duration of sanctions is about 5.5 years.

Thus, from an econometric perspective, sanctions should be viewed as treatments with entry *and exit*, which, in turn, poses new and potentially important challenges for the identification of their effects. Specifically, as argued by a relatively recent, but already influential econometrics literature on multiple treatments, (e.g., Hull (2018); Sun and Abraham (2021); de Chaisemartin and D'Haultfœuille (2023); Goldsmith-Pinkham et al. (2024)), TWFE regressions with several treatments or treatments with entry and exit, such as sanctions, may be biased not only due to negative (or arbitrary) weighting of treatment effects of the same category (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2024; Wooldridge, 2025), but also due to *contamination bias* triggered by the effects of other treatments.

[3] For example, without proper partial estimates of the impact of the current sanctions on Russia's trade with the sanctioning states, one cannot evaluate the trade diversion effects of the sanctions on Russia's trade with China, India, and Turkey, which have been crucial for the (in)effectiveness of these sanctions.

[4] Throughout the paper, we use TWFE (ETWFE) to refer to (extended) two-way fixed effects, following the econometrics literature. In the gravity literature, a two-way fixed effects model typically includes only exporter-year and importer-year fixed effects. However, our empirical analysis employs a three-way gravity model with country-pair, exporter-year, and importer-year fixed effects.

What makes trade sanctions even more interesting from a policy perspective is that bilateral trade flows may respond differently after the lifting of sanctions. For example, trade flows may recover fully and instantaneously (i.e., within a year), it may take time for trade flows to recover back to their pre-sanction levels, trade flows may not recover at all, or they may increase to new levels above their pre-sanction levels. All these possibilities are potentially important from a policy perspective, however, they cannot be evaluated if the econometric treatment of sanctions does not properly account for their lifting/exit.

Against this backdrop, our contribution is fourfold. First, from a policy perspective, we re-evaluate the effects of sanctions on trade when they are in place. We confirm that sanctions hurt trade. In addition, and most importantly, we demonstrate that existing estimates based on established methods may be significantly biased.[5] Second, we also obtain novel estimates of the effects of sanctions not only when they are in place, but also during the period after they are lifted, thus offering further evidence on the impact and effectiveness of trade sanctions.[6]

Third, from a methodological perspective, we capitalize on the most recent developments in the DiD literature with multiple treatments (e.g., Hull (2018); Sun and Abraham (2021); de Chaisemartin and D'Haultfœuille (2023); Goldsmith-Pinkham et al. (2024)) to demonstrate that sanctions estimates based on established approaches from the TWFE trade gravity literature may be significantly biased, both due to negative (or arbitrary) weighting and contamination bias from multiple treatments. Specifically, we use an extended TWFE (ETWFE) model that allows for treatment effect heterogeneity in the

[5] Without attempting to offer an exhaustive list of the papers that have studied the impact of sanctions on trade, some examples include Hufbauer and Oegg (2003), Caruso (2003), Hufbauer et al. (2007), Afesorgbor (2019), Felbermayr et al. (2020), Crozet and Hinz (2020), Dai et al. (2021), Gutmann et al. (2023), and Felbermayr et al. (2025b). Felbermayr et al. (2021), Morgan et al. (2023), and Felbermayr et al. (2025a) offer broader perspectives on the evolution of the sanctions literature. Dai et al. (2021), who estimate the effects of sanctions on trade during their imposition and after their lifting, is the closest to us in terms of objectives. The key difference between our analysis and Dai et al. (2021), as well as all previous investigations of the effects of sanctions, is that, to our knowledge, they all rely on TWFE methods, while we use ETWFE methods with exit, and we show that they have significant quantitative implications for estimating the effects of economic sanctions.

[6] For example, increased trade between the sender and the target post-lifting can be interpreted (i) as evidence of increased leverage by the sending country over the receiving country, or alternatively, (ii) as evidence that the sanctions have achieved their political objectives and, as a result of that, the economic relationship between the sender and the target has reversed its direction.

presence of entry and exit (Wooldridge, 2023, 2025), which we adapt to a structural gravity setting.[7] Even though our focus is on trade, we expect that the methods we implement here will have broader implications for estimates of the effects of sanctions on other economic outcomes too.

Fourth, we move beyond average effects and provide a systematic ex-post heterogeneity analysis of sanctions' impact on trade. Specifically, using cohort-event-time ETWFE estimates as inputs, we relate the imposition and lifting effects of sanctions to geography, pre-sanction trade integration, sanction-case attributes, concurrent financial sanctions and sanction history, and country income levels, and we document distinct recovery profiles for short- versus long-lived episodes. This delivers a granular characterization of where sanctions bite hardest and where trade scars are most persistent, complementing existing work that studies heterogeneity in sanctions' trade effects using interactions in a TWFE framework, such as Felbermayr et al. (2025b).

To perform the analysis, we rely on standard datasets,[8] and our findings have several implications, both from a policy and a method perspective. Our main policy findings can be summarized as follows: First, we obtain large, negative, and statistically significant estimates of the effects of sanctions on trade, which imply that, on average, complete trade sanctions have eliminated about 58% of bilateral trade between senders and targets. Second, the average impact of sanctions on trade between senders and targets following their lifting is economically insignificant and lacks statistical significance. Third, allowing for heterogeneous effects of short versus long sanctions, we find that short sanctions (lasting 6 years or less) exhibit a clear post-lifting rebound, with trade rising relative to its pre-sanction level. In contrast, long sanctions show little evidence of returning to pre-sanction levels after they are lifted. Fourth, we show that the trade effects of

[7] Nagengast and Yotov (2025) also combine gravity and staggered DiD methods to evaluate the effects of trade agreements as a treatment with entry only. A crucial difference between the current analysis and Nagengast and Yotov (2025) is that our focus is on a treatment with entry *and exit*.

[8] The data on sanctions come from the GSDB, while the original sources for the trade data are the IMF's Direction of Trade Statistics (DoTS) database and the United Nations Commodity Trade Statistics (COMTRADE) database. The combination of these datasets results in an unbalanced panel estimation sample of aggregate bilateral trade flows and sanctions, which covers 260 countries over the period 1950-2019. We provide additional information on the data, sources, and summary statistics in Section 3.

sanctions are highly heterogeneous across dyads and episodes: dyads involving distant or contiguous partners, war-related objectives, unsuccessful outcomes, and a history of other (non-trade) sanctions tend to exhibit larger, and in several cases more persistent, trade contractions, whereas pre-sanction rules-based trade integration (WTO co-membership and RTAs) and higher income levels, especially on the importer side, are associated with attenuated contemporaneous losses and, in several cases, stronger post-lifting recoveries. A series of sensitivity experiments confirm the robustness of our results.

More importantly, from a method perspective, our analysis implies that existing TWFE gravity methods that do not properly account for the lifting of sanctions (i.e., exit) may deliver significantly biased estimates. Specifically, our preferred ETWFE estimates of the effects of sanctions during the period when they are in place are about 50% larger than those from a traditional TWFE regression. Importantly, we find that the bias in TWFE estimates arises not only from the negative weighting of treatment effects but also from substantial contamination across different treatments. Moreover, we do not obtain significant estimates of the effects of sanctions during the period after they are lifted, while the corresponding TWFE estimates are positive and statistically significant.

In sum, we demonstrate that the new heterogeneity-robust methods, which are specifically designed to account for the challenges of estimating the effects of treatments with exit, lead to very different estimates of the effects of sanctions on trade within the gravity model. Our analysis has two broader implications: (i) The estimates of other treatment effects with entry and exit on international trade (e.g., various standards, temporary trade promotion and trade bans, etc.) may also suffer significant biases (which we illustrate for currency unions in Section 5), and (ii) The new DiD methods with entry and exit may have important implications for gravity estimations beyond trade, e.g., on migration, foreign investment, or any other bilateral flows.

The rest of the paper is structured as follows. Section 2 presents our estimation methods. Section 3 describes our dataset, the data sources, and offers summary statistics for the key variables. Section 4 presents our main findings including bias decomposition and an analysis of how average sanction effects vary across observable characteristics. Section 5 discusses additional results and Section 6 reports a series of validity checks and sensitivity analyses. Section 7 concludes the paper with a brief summary and discussion of potential directions for future work.

2 Methods

Considering the most recent developments in the literature on staggered DiD with entry and exit, in combination with established methods from the trade gravity literature, delivers our main estimating equation:

$$\ln Y_{ij,t} = \sum_{g\in\mathcal{G}} \sum_{h\in\mathcal{H}_g} \sum_{s=g}^{T} \delta_{gh,s-g} \mathbf{1}\{\, G_{ij} = g,\ H_{ij} = h,\ t = s\,\} + \pi_{i,t} + \chi_{j,t} + \tau_{i,j} + \epsilon_{ij,t}. \qquad (1)$$

Here, $Y_{ij,t}$ represents the nominal trade flows from country i to country j at time t.[9] The first term on the right-hand side of (1) is the key term in our analysis. Specifically, following the heterogeneity-robust ETWFE approach by Wooldridge (2023, 2025), the triple sum, $\delta_{gh,s-g}\mathbf{1}\{\, G_{ij} = g,\ H_{ij} = h,\ t = s\,\}$, can also accommodate staggered exits from treatment and replace the indicator variables that capture the effects of sanctions in the existing literature relying on TWFE specifications. Defining event time by $\ell \equiv s - g$, $\delta_{gh,\ell}$ indexes cohort-event-year effects (with $\ell = 0$ the first year a sanction was imposed, $\ell \geq h - g$ post-exit). Here, country pair ij belongs to treatment cohort gh if a sanction was imposed in year g and lifted in year h, $\mathcal{G}$ is the set of observed start years in the sample, and for each $g \in \mathcal{G}$, $\mathcal{H}_g$ is the set of exit years for sanctions starting in year g,[10] G_{ij} and H_{ij} denote the start and exit years for pair (i, j), T is the last year of the panel,

[9]Following recent advancements in the DiD literature, we obtain our results in a log-linear setting instead of using the Poisson Pseudo Maximum Likelihood (PPML) estimator. This choice allows us to apply the insightful bias decompositions by Sun and Abraham (2021), de Chaisemartin and D'Haultfœuille (2023), and Goldsmith-Pinkham et al. (2024), which are crucial for breaking down estimates into own and contamination bias terms. These decompositions are not directly applicable in non-linear models like PPML.

[10]For example, $\mathcal{H}_{1990} = \{1991, 1994\}$ would be the set of exit years for sanction episodes starting in 1990, meaning that, in this hypothetical example, the sanctions were lifted in 1991 for some country pairs and in 1994 for others.

the indicator $\mathbf{1}\{G_{ij} = g,\ H_{ij} = h,\ t = s\}$ equals one for members of cohort (g, h) in calendar year s (i.e., event time $\ell = s - g$) and zero otherwise, and $\delta_{gh,s-g}$ captures the cohort-event-time-specific treatment effects.[11]

To highlight the differences between the TWFE and the ETWFE specification explicitly, and using the terminology of de Chaisemartin and D'Haultfœuille (2023), a *'long' TWFE* specification contains an indicator variable for sanctions between i and j at time t, $ON_{ij,t}$, and an additional indicator variable, $OFF_{ij,t}$, which is equal to one in all periods after the sanction was lifted (e.g., Besedeš et al., 2017; Gutmann et al., 2023):[12]

$$\ln Y_{ij,t} = \delta^{ON}_{TWFE} ON_{ij,t} + \delta^{OFF}_{TWFE} OFF_{ij,t} + \pi_{i,t} + \chi_{j,t} + \tau_{i,j} + \epsilon_{ij,t}. \tag{2}$$

Alternatively, the commonly used *'short' TWFE* specification (e.g., Hufbauer et al., 2007; Felbermayr et al., 2020) only contains a single indicator variable for sanctions, $ON_{ij,t}$, which treats time periods after the sanction was lifted as untreated:

$$\ln Y_{ij,t} = \delta^{ON}_{TWFE} ON_{ij,t} + \pi_{i,t} + \chi_{j,t} + \tau_{i,j} + \epsilon_{ij,t}. \tag{3}$$

Importantly, our specification takes into account the most recent developments in the literature on DiD with multiple treatments (Hull, 2018; Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2023; Goldsmith-Pinkham et al., 2024). The literature shows that TWFE regressions with several treatments are not guaranteed to identify convex combinations of treatment effects, with estimates potentially contaminated by the effects of other treatments. Decomposition results demonstrate that the bias of the TWFE regression in the presence of heterogeneous treatment effects can be derived from two different sources: (i) negative (or arbitrary) weighting of treatment effects in the same

[11]For example, $\delta_{1990,1991,0}$ captures the treatment effect of all sanctions imposed in 1990 and lifted in 1991 on trade in the first year of the sanction, $\delta_{1990,1991,1}$ captures the treatment effect of the same sanction in the first year after lifting, and $\delta_{1990,1991,2}$ captures the treatment effect in the second year after lifting, and so on.

[12]Similarly, some recent studies on sanctions rely on event-study TWFE specifications (e.g., Dai et al., 2021; Gutmann et al., 2023). To benchmark our analysis against these studies, we also provide aggregations of our ETWFE estimates in event time.

category and (ii) contamination bias from other treatment effects. The first is identical to the well-documented bias of the TWFE regression with a single treatment (e.g., de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2024; Wooldridge, 2025).[13] The contamination bias term is the weighted sum of the effect of other treatments, with the weights summing to zero for mutually exclusive treatments (e.g., de Chaisemartin and D'Haultfœuille, 2023; Goldsmith-Pinkham et al., 2024).[14] Therefore, in the presence of heterogeneous treatment effects, this term is not guaranteed to sum to zero. Conceptually, the source of the bias in the TWFE regression lies in two types of 'forbidden comparisons' (Borusyak et al., 2024), which include comparing units treated in later periods to units treated with (i) the same treatment in earlier periods or (ii) a different treatment in earlier periods.

Several alternative estimation approaches with different emphases have been proposed in the literature, which are heterogeneity-robust and immune to contamination bias in settings with several treatments. For example, Sun and Abraham (2021) offer an alternative estimation method for event-study designs. de Chaisemartin and D'Haultfœuille (2023) propose an alternative model-based estimator that also extends to the case of non-mutually exclusive treatments. Goldsmith-Pinkham et al. (2024) discuss three design-based estimation approaches that avoid contamination bias for mutually exclusive treatments. Wooldridge (2023, 2025) propose a simple extension of the TWFE regression that also nests an estimation strategy for the case, where an intervention switches off, i.e., for exit.[15] In this paper, we follow the approach suggested by Wooldridge (2025) mainly due to its ease to accommodate the fixed effect structure of the standard gravity specification.

The key parameters $\delta_{gh,s-g}$ in equation (1) identify the treatment effects under the following three assumptions (Wooldridge, 2025). First, a no anticipation assumption requires that, on average and conditional on the fixed effects, there is no effect of the

[13] de Chaisemartin and D'Haultfœuille (2022) and Roth et al. (2023) offer surveys of this literature.

[14] For the case of non-mutually exclusive treatments, the contamination weights do not sum to zero (de Chaisemartin and D'Haultfœuille, 2023).

[15] The imputation approach by Borusyak et al. (2024) also allows for non-absorbing treatments, while this case is not explicitly discussed in their paper. In the same vein, Dube et al. (2025) focus on a local projections approach that can also accommodate entry and exit.

treatment before its onset. Second, a parallel trends (in differences) assumption requires that, conditional on the fixed effects, the average counterfactual change in outcomes would have been the same across treated and control cohorts in pre-treatment periods. Third, and specific to the setting with exit studied here, shocks to potential outcomes are assumed not to be correlated with exit in future periods (see Wooldridge (2023) for a discussion). In Section 4.1, we follow the recommendations by Wooldridge (2023, 2025) and test these assumptions explicitly.

In this paper, we are interested in evaluating by how much international trade has changed, on average, in all country pairs and post-intervention years as a result of the imposition and lifting of sanctions. Consistent with the gravity literature and much of the recent heterogeneity-robust DiD literature, we take two static averages as our baseline estimands to provide comparable, single-number summaries: the average effect while sanctions are in force (ON) and the average effect after sanctions are lifted (OFF). These static estimands are policy-relevant summary measures (the average log change in trade during and after sanctions) and facilitate direct comparisons with TWFE-based studies reporting single coefficients. Therefore, our main estimands are two sanction averages, which are defined as sample-share-weighted averages of cohort-event-time-specific treatment effects:

$$\theta^{ON} = \sum_{g \in \mathcal{G}} \sum_{h \in \mathcal{H}_g} \sum_{s=g}^{h-1} \frac{N_{ghs}}{\mathcal{N}_{ON}} \delta_{gh,s-g}, \tag{4}$$

and

$$\theta^{OFF} = \sum_{g \in \mathcal{G}} \sum_{h \in \mathcal{H}_g} \sum_{s=h}^{T} \frac{N_{ghs}}{\mathcal{N}_{OFF}} \delta_{gh,s-g}. \tag{5}$$

Here, θ^{ON} summarizes the average log change during sanctions and θ^{OFF} the average log change after lifting (both relative to the pre-sanction period). In practice, we obtain plug-in estimates by replacing $\delta_{gh,\ell}$ with the estimates $\hat{\delta}_{gh,\ell}$ of the ETWFE model. In both cases, we assign equal weights to all relevant post-treatment observations, which correspond to the number of observations of cohort gh in period s, N_{ghs}, relative to the total number of treated observations, $\mathcal{N}_{ON} = \sum_{g \in \mathcal{G}} \sum_{h \in \mathcal{H}_g} \sum_{s=g}^{h-1} N_{ghs}$ and $\mathcal{N}_{OFF} = \sum_{g \in \mathcal{G}} \sum_{h \in \mathcal{H}_g} \sum_{s=h}^{T} N_{ghs}$, respectively (de Chaisemartin and D'Haultfœuille, 2023). Similarly, the standard errors are computed from the estimated covariance matrix using the same weights. While our primary estimands are the static ON and OFF averages, we also place emphasis on event-time (horizon-specific) treatment effects to characterize the dynamics of sanction impacts. These event-study estimates are constructed by averaging cohort-event-time effects across cohorts and tracing the adjustment path, persistence, and post-exit behavior. For completeness, we also compute cohort-specific treatment effects by averaging over the event-time dimension.

The remaining terms in equation (1) are motivated by the empirical gravity literature on trade. Specifically, $\pi_{i,t}$ are exporter-year fixed effects and $\chi_{j,t}$ are importer-year fixed effects. From a trade theory perspective, the motivation for the inclusion of these two sets of fixed effects is to account for country size and the structural multilateral resistance terms of Anderson and van Wincoop (2003). Additionally, from an econometric perspective, these fixed effects fully absorb and account for any time-varying country characteristics, such as trade diversion (e.g., Tyazhelnikov and Romalis, 2024), that may impact trade from either the exporter or importer side. $\tau_{i,j}$ denotes a set of directional pair fixed effects that directly follow the DiD specification under consideration. In addition, from a trade theory perspective, they control for any symmetric or asymmetric time-invariant bilateral costs and determinants of trade flows (e.g., distance, contiguity, etc.).

$\epsilon_{ij,t}$ is an error term and we cluster the standard errors by (directional) country-pair. We rely on clustering by country-pair because this approach is standard in the trade gravity literature. While there is no definitive analytical guidance on the appropriate level of clustering, pair-level clustering is typically used to account for serial correlation of shocks over time within a specific trading relationship. Conceptually, many trade shocks, such as policy changes, geopolitical events, or supply disruptions, are likely to be persistent within exporter-importer pairs rather than across all exporters or importers. Clustering separately by exporter and importer assumes a much broader correlation structure, allowing the shocks of two unrelated country pairs to be correlated simply because they share an exporter or an importer. This may overstate the degree of dependence in the data. Therefore, for transparency, we report the results under both clustering schemes (see Section 6). Importantly, our main results remain robust regardless of the clustering method, which reinforces the stability of our findings.

Finally, similar to the TWFE model, the ETWFE model identifies only partial effects of sanctions on trade, since the inclusion of exporter-year and importer-year fixed effects absorbs multilateral resistance terms, which are expected to adjust following the imposition or removal of sanctions.

3 Data

To perform the analysis, we combine several datasets, including data on bilateral trade flows, data on economic sanctions, and 'other' data that is used for the additional control variables in our regressions. The result is an unbalanced panel dataset of aggregate bilateral trade flows and economic sanctions at the country-pair level, which covers 260 importers and 260 exporters over the period 1950-2019.

3.1 Data sources

Bilateral trade flows data. The data for our dependent variable, nominal aggregate bilateral trade flows, was constructed by Felbermayr et al. (2025b), 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. (2025b) 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)), cost, insurance, and freight (CIF) imports are used as the baseline data. Then, missing values are replaced with data on free on board (FOB) exports. The trade data of Felbermayr et al. (2025b) covers the period 1950-2019, and it has two advantages for our purposes: it covers (i) a long period of time and (ii) a large number of countries.

Sanctions data. The data on sanctions comes from the third edition of the Global Sanctions Database, GSDB-Release-3, (Syropoulos et al., 2024). The first version of the GSDB was constructed by Felbermayr et al. (2020) from publicly available sources and, to ensure comprehensive coverage, it was cross-checked against existing databases, e.g., the Hufbauer-Schott-Elliott sanctions data of Hufbauer et al. (2007) and the Threat and Imposition of Economic Sanctions (TIES) database of Morgan et al. (2014).[17] GSDB-Release-3 defines a sanction case as an official sanction policy that is initiated by an individual country, a group of countries, or an international organization against another nation and that is identified on the basis of an official decision (e.g., an executive order, a parliamentary decision, or a resolution), and it covers 1,325 sanction cases over the period 1950-2022.[18]

The GSDB classifies sanctions into six categories by type: trade, financial activity, arms, military assistance, travel, and other sanctions. The GSDB distinguishes between partial and complete trade sanctions. Following the existing literature, the focus of our main analysis is on *complete trade sanctions*, however, in the robustness analysis (Sec-

[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 GSDB is freely available to researchers, although access must be requested by email at GSDB@drexel.edu. Details are available at http://www.globalsanctionsdatabase.com/.

[18]The time dimension of the GSDB is yearly. Thus, sanctions may be imposed or lifted at any time of the year, which may impact our estimates. For example, if a sanction is imposed toward the end of a given year, our estimates of the initial sanction effects may be biased downward (in absolute value).

tion 6), we also obtain estimates for partial trade sanctions. We also evaluate the effects of financial sanctions and other types of sanctions on trade.

Since we are interested in both the effect of sanctions and their lifting, we restrict the sample to include only those sanctions that were lifted before 2019, the last year in our trade dataset. To ensure clean identification of effects, we focus on sanctions with data available for the year the sanction was imposed, the year it was lifted, and the immediate pre-entry and pre-exit years.[19] To focus on the most relevant period and reduce the potential influence of other events on our results, we restrict the estimating sample to the 10 years following the lifting of sanctions for treated country pairs. For country pairs that were never sanctioned, the sample by definition includes all available observations. We exclude post-lifting observations after which new complete trade sanctions were re-imposed. Additionally, we restrict the baseline sample to sanctions with durations of 14 years or less, covering approximately 97% of all treated observations. For robustness, in Section 6, we consider a sample that contains all post-exit observations and a sample that includes sanctions of all durations. Overall, this leaves us with 37 unique sanction cohorts defined by different combinations of imposition and lifting years, covering 795 country pairs.

3.2 Descriptive statistics

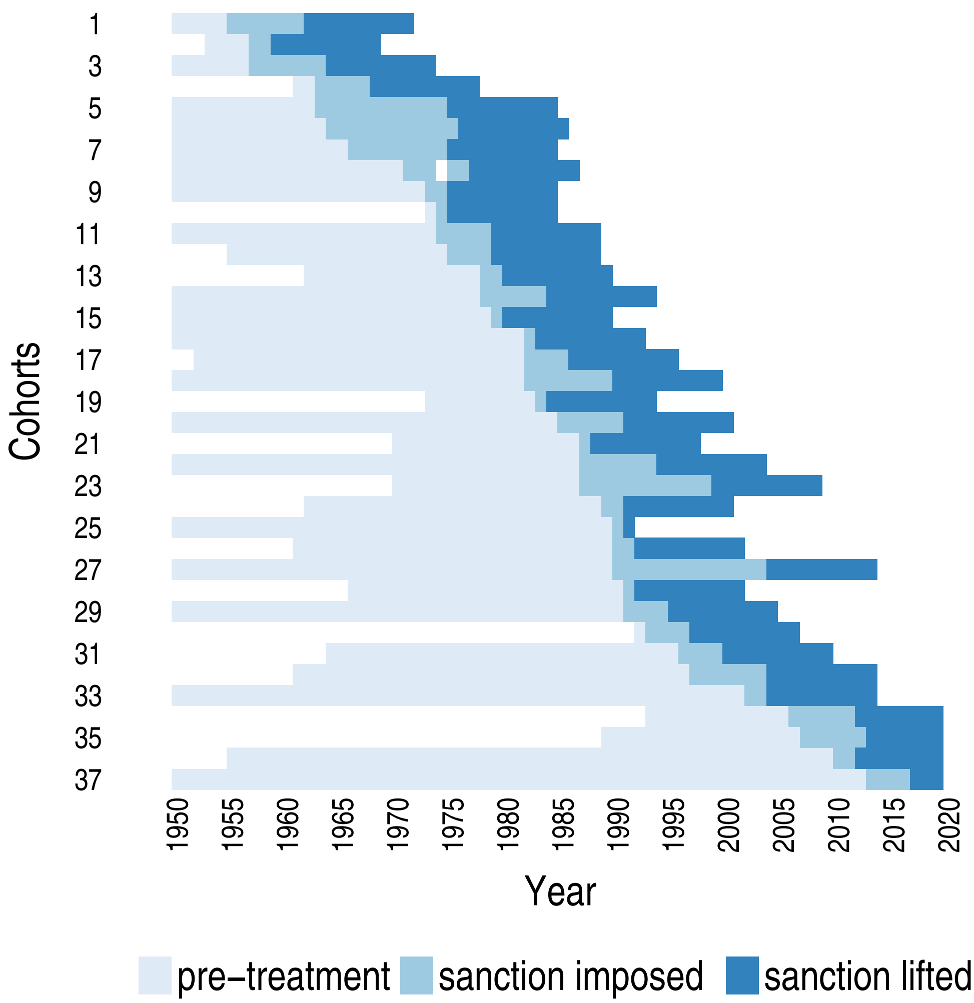
Figure 2 illustrates the treatment status of cohorts over time. Pre-treatment years are shown in light blue, years with sanctions imposed are shown in medium blue, and years after sanctions were lifted are shown in dark blue. Years not included in the estimation sample, for example, due to missing or zero trade data, are shown in white.[20] The figure demonstrates that sanction episodes vary in length and are well distributed across the timeline from 1950 to 2019. This distribution underscores the importance of considering

[19] Note that missing-at-random data is unlikely to affect our results under our identification assumptions. As discussed in Section 6, our estimation results remain robust for different estimation samples.

[20] Note that for cohort 8 in the fourth year under sanctions, reported trade is zero, resulting in a gap. This hints at the potential importance of the extensive margin, which we analyze in Section 5.

exits from sanctions when analyzing their effects. The figure complements the findings presented in Figure 1, reinforcing the need to account for the dynamics of both the imposition and lifting of sanctions in the evaluation of their overall impact.

Figure 2: Cohorts and treatment status over time



Notes: This figure illustrates the treatment status of the treated cohorts over time for the baseline sample used to obtain our main results in Table 2. Pre-treatment years are represented in light blue. Post-treatment years during which sanctions were imposed are represented in medium blue. Post-treatment years after sanctions were lifted are represented in dark blue. Years not included in the estimation sample, for example, due to missing or zero trade data, are shown in white. To focus on the most relevant period and minimize the potential impact of other events on our estimates, the estimating sample was restricted to 10 years after a sanction was lifted for treated country pairs. For country pairs that were never sanctioned, the sample by definition includes all available observations. The figure was generated using the Stata module `panelview` (Mou and Xu, 2022).

Table 1 provides descriptive statistics on the duration of sanction episodes. The table lists the number of cohorts, the number of country-pair-year observations during which sanctions were imposed (i.e., ON-treatment observations), share, and cumulative total for each duration in the baseline dataset. The data show that a significant portion of sanction episodes are relatively short, with around a quarter lasting between 1 and 4 years. Longer durations, such as 9 and 14 years, also have notable shares of approximately 18% and 39%, respectively. The cumulative total column highlights that the majority of sanction episodes fall within the first 9 years, covering 51% of all observations. The table emphasizes the varying lengths of sanctions and the potential importance of accounting for these differences in analyses.

Table 1: Descriptive statistics on sanction episode durations

Duration	Cohorts	ON observations	Share	Cumulative total
1	7	76	2.14	2.14
2	7	400	11.27	13.41
4	6	358	10.08	23.49
5	2	20	0.56	24.06
6	5	153	4.31	28.37
7	4	184	5.18	33.55
8	1	8	0.23	33.77
9	1	626	17.63	51.41
12	3	328	9.24	60.65
14	1	1,397	39.35	100.00
Total	37	3,550	100.00	

Notes: This table provides descriptive statistics on the duration of sanction episodes for the baseline sample that is used to obtain our main results in Table 2. The duration is measured in years, and the table lists the number of cohorts, the number of country-pair-year observations during which sanctions were imposed (i.e., ON-treatment observations), share, and cumulative total for each duration.

Lastly, Table A1 in the Appendix shows the number of observations by cohort, differentiating between observations while sanctions were imposed and after sanctions were lifted. The table indicates that the baseline sample contains 10,855 post-treatment observations, of which 3,550 are observations while sanctions were imposed and 7,305 are observations after sanctions were lifted. A total of 795 country pairs were treated, composed of 142 exporters and 149 importers. In comparison, the majority of country pairs in the dataset (46,495) were never treated. Second, the majority of cohorts contain a reasonable number of observations per year. For robustness, in Section 6, we also consider an alternative sample with a minimum number of observations per cohort-year cell.

4 Main results and analysis

This section presents our main findings in four steps. First, we assess the identifying assumptions of the ETWFE model (Section 4.1). Second, we report our preferred ETWFE estimate of the average sanction effect and compare it directly with the literature's standard TWFE estimate, and we show event-study evidence on the dynamics for both models (Section 4.2). Third, we present bias decompositions, clarifying why the TWFE model yields different results (Section 4.3). Finally, we examine how average sanction effects vary across observable characteristics (Section 4.4).

4.1 Test of identifying assumptions

As discussed in Section 2, the ETWFE model relies on 'no anticipation' and 'parallel trends' (in differences) assumptions. To assess the validity of these assumptions, we implement tests following Wooldridge (2025). Specifically, we introduce placebo treatment indicators for the years preceding the imposition of sanctions and re-estimate equation (1). This amounts to testing for differential pre-trends, using periods more than ten years before sanctions ($\ell < -10$) as the omitted pre-period.[21] Figure 3a presents the placebo treatment effects, aggregated by event time. The estimates for the pre-treatment period are not statistically significant and cluster around zero, suggesting no significant anticipation effects or differential pre-trends. A joint hypothesis test of the event-time leads further supports this conclusion, yielding a p-value of 0.129. Due to the clustering of some pre-treatment effects slightly above zero, the joint significance test approaches, but does not attain, statistical significance.

[21] Additionally, a pre-treatment test based solely on observations before the treatment, as suggested by Borusyak et al. (2024) and in the vein of Roth (2022), yields similar results (Figure A1 in the Appendix). Note that the two tests are not equivalent in our context due to the presence of control variables (i.e., fixed effects) inherent in the gravity specification we employ. Furthermore, a pre-treatment in the spirit of Liu et al. (2024), obtained from an imputation-type estimator (Borusyak et al., 2024) using only never-treated and not-yet-treated observations, also yields similar results (Figure A2 in the Appendix).

An additional assumption when considering settings with exit is that shocks to potential outcomes do not precipitate exits in subsequent periods. This is essentially a strict exogeneity assumption for the time-varying treatment indicator, after accounting for unobserved confounders with fixed effects, as discussed in detail by Wooldridge (2023). To evaluate this assumption, we employ two diagnostics. First, we adapt the no anticipation test from Wooldridge (2023), allowing for treatment effect heterogeneity by entry year in a sample limited to pre-exit years, and include a dummy variable for the year immediately preceding exit.[22] The coefficient for this variable is not statistically significant (0.371, SE 0.349). Second, we examine potential pre-trends prior to exit within our baseline model.[23] Figure 3b compares the treatment effects of sanctions in the two periods before they are lifted, using a consistent sample of country pairs that are present in both periods (exit-2 and exit-1) to avoid composition effects.[24] The estimates for the periods immediately before exit (exit-2 and exit-1) are similar in magnitude and their difference is not statistically significant (p-value of 0.884), indicating no evidence of pre-trends.[25]

In summary, the analyses in this subsection do not reveal any violations of the identifying assumptions. Therefore, we proceed with the main analysis.

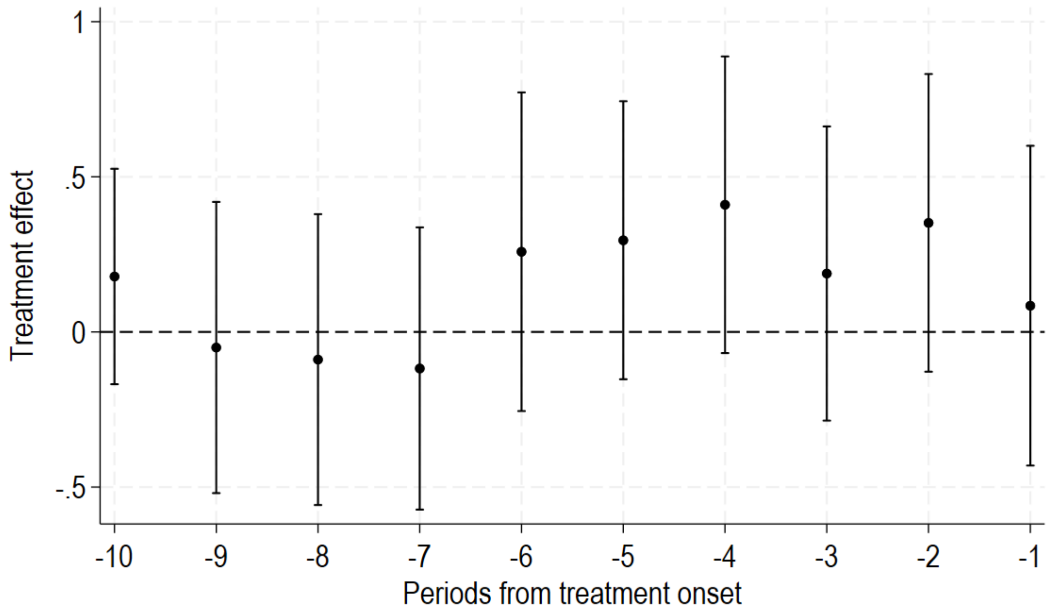
[22] Note that allowing for full treatment effect heterogeneity as in our baseline (i.e., defining cohorts by entry and exit year) is infeasible, since the dummy variable would then be collinear with the cohort-year treatment effects.

[23] Note that, in principle, the treatment effect in the two years before exit might also be influenced by the dynamics of treatment effects over time, specifically the increasing impact of sanctions as they accumulate.

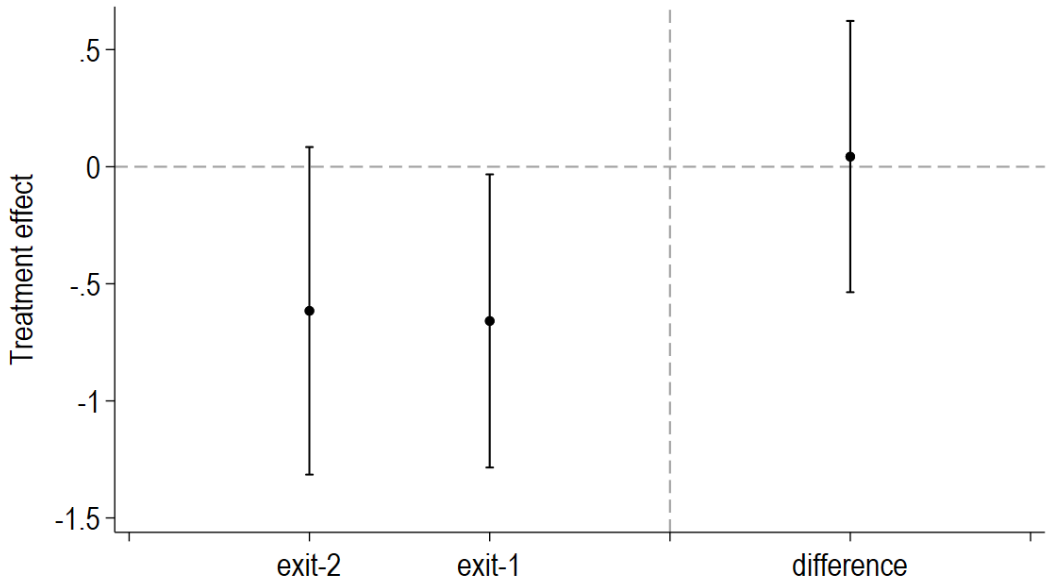
[24] For some country pairs treatment effects are not observed in both periods. For example, if a sanction lasts only one year, there is by definition no treatment effect two years before exit. In other cases, trade observations may be missing or zero.

[25] Note that the estimates for exit-2 and exit-1 are treatment effects defined relative to the exit year, not pre-treatment estimates as in event-study analyses defined relative to entry. As these points reflect the impact of sanctions while they are still in force, the relevant diagnostic is the difference between the two adjacent pre-exit effects rather than the magnitude of each individual estimate.

Figure 3: Pre-treatment effects



(a) Pre-entry effects



(b) Pre-exit effects

Notes: Panel (a) reports pre-trend estimates from an estimation of equation (1), to which additional placebo treatment indicators were added in the years before the sanction was imposed following Wooldridge (2025). Panel (b) reports pre-exit estimates from an estimation of equation (1), restricting the sample to country pairs observed in both periods (exit-2 and exit-1) to ensure a consistent comparison and avoid composition effects. For visualization, the resulting cohort-event-time-specific treatment effects were aggregated by event time. 95% confidence intervals are shown using standard errors clustered by country pair.

4.2 Main estimation results

This section presents our main findings in three steps. First, we discuss estimates of the effects of sanctions from our preferred ETWFE specification, which explicitly accounts for sanction lifting. Second, we compare these estimates with corresponding TWFE results, which also take into account the lifting of sanctions ('long' TWFE specification) as well as with the more prevalent TWFE estimates that do not account for exit, i.e., those with a single treatment ('short' TWFE specification). Finally, we present horizon-specific results for the baseline ETWFE specification and, for comparison, a dynamic TWFE model. Throughout, we interchangeably refer to the effect of sanctions on trade while they are in force as the 'sanction effect', 'ON effect', or 'entry effect', and to the effect of lifting sanctions as the 'lifting effect', 'OFF effect', or 'exit effect'.

Baseline ETWFE results. Table 2 presents our main findings and, based on the ETWFE estimates in column (1), we draw two conclusions about the effects of sanctions on trade. First, the impact of complete trade sanctions on bilateral trade flows between the sender and target states is negative, large, and statistically significant. Specifically, our estimate of -0.859 (SE 0.332) implies that the complete trade sanctions in our sample have eliminated about 58% (calculated as $(\exp(-0.859) - 1) \times 100 = -57.64$) of the bilateral trade flows between senders and targets.[26] The policy implications of this result are that sanctions are effective in decreasing bilateral trade, however, even complete trade sanctions do not fully eliminate trade between the sender and the target.[27] We do note, however, that our estimate is actually larger than corresponding results from the related literature. We discuss such comparisons in more detail below. The second main finding

[26]To rule out confounding by direct bilateral war, we use the Directed Dyadic Interstate War Dataset (Maoz et al., 2019) (Correlates of War) with data available until 2010. We drop both directions of any dyad with a bilateral war in 1950-2010 and truncate the sample at 2010. Re-estimating the baseline yields an imposition (ON) effect of -0.853 (SE 0.363) and a lifting (OFF) effect of 0.013 (SE 0.224). For comparability, the baseline on the full country set restricted to 1950-2010 gives very similar estimates: ON -0.856 (SE 0.374) and OFF 0.069 (SE 0.227). Hence the sanction effects are not driven by direct bilateral warfare.

[27]Possible explanations for this result could be that in the case of multilateral sanctions, for example, some countries obtain waivers while others do not enforce the sanctions completely.

from column (1) is that our estimate of post-sanction effects on trade is economically small and not statistically significant. Under the caveat that this result may be masking some heterogeneity over time or across sanctions (we investigate both possibilities later), the implication is that sanctions do not have negative effects on trade between senders and targets after the sanctions are lifted.

Comparison with (static) TWFE estimates. Next, we turn to some comparisons between our preferred estimates from column (1) and results based on TWFE specifications from the existing literature. The results in column (2) are obtained from a 'long' TWFE specification that also allows for ON and OFF sanction effects. The two main conclusions of this study are noteworthy. First, the ETWFE estimate of the impact of sanctions during the period when they are in place is substantially larger than the corresponding TWFE estimate, which implies a reduction of trade about 38% of the bilateral trade flows between senders and targets. Specifically, our estimates suggest that the ETWFE estimate is about 50% larger than the TWFE estimate (58% vs 38%). Second, we see from column (2) that the TWFE estimates of the sanction effects during the years after the sanctions are lifted are positive, sizable, and statistically significant. This is another important difference from the ETWFE estimates from column (1), which highlights the importance of the new estimation methods that we employ.[28]

Finally, column (3) presents results from a 'short' TWFE specification, which is the most prevalent in the literature and does not account for the lifting of sanctions. The difference between the ETWFE estimate and this TWFE estimate is notable. The TWFE specification in column (3) shows a reduction in trade by about 45%, which is still less than the ETWFE estimate (58%) but larger than the reduction observed with the 'long'

[28] As a robustness check, we also estimate (i) a conventional TWFE PPML model and (ii) a heterogeneity-robust PPML specification (Wooldridge, 2023; Moreau-Kastler, 2025; Nagengast and Yotov, 2025). For the baseline sample, the conventional TWFE PPML model yields an imposition effect of -0.658 (SE 0.133) and a post-lifting effect of -0.211 (SE 0.130). The corresponding proportional treatment effects from the heterogeneity-robust PPML are -0.878 (SE 0.104) during imposition and -0.475 (SE 0.361) post-lifting. The qualitative pattern mirrors our OLS results: sanctions depress trade during imposition, with the heterogeneity-robust PPML delivering slightly more negative effects. Post-lifting effects are negative in both models but imprecisely estimated and not statistically different from zero.

Table 2: Main results

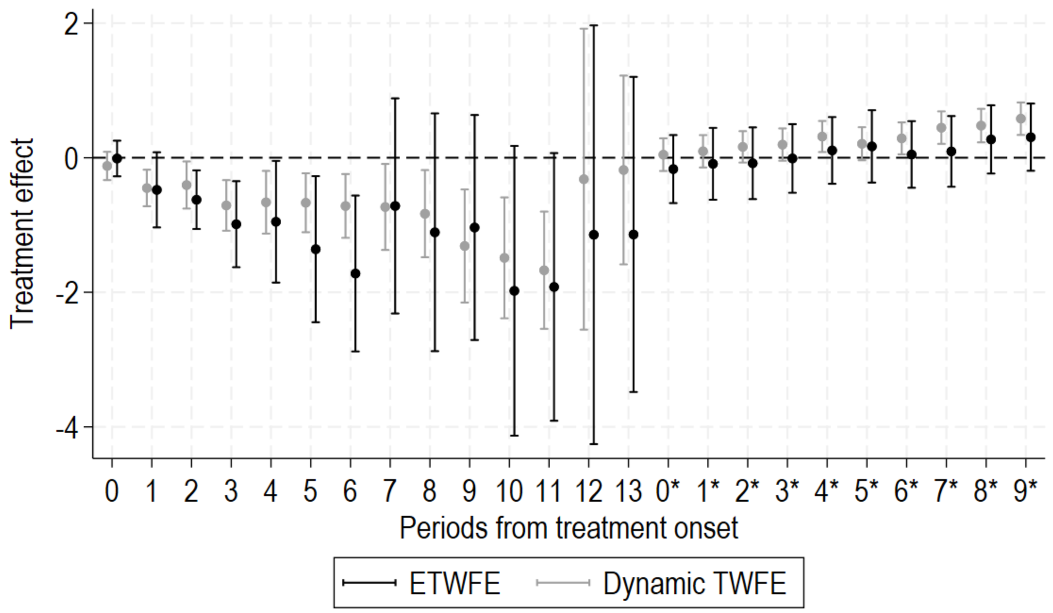
	(1) ETWFE	(2) TWFE	(3) TWFE
$ON_{ij,t}$	-0.859***	-0.482***	-0.604***
	(0.332)	(0.117)	(0.102)
$OFF_{ij,t}$	0.071	0.292***	
	(0.220)	(0.093)	
$OFF_{ij,t} - ON_{ij,t}$	0.930***	0.774***	
	(0.270)	(0.103)	
Observations	1,141,068	1,141,068	1,141,068
thereof: ON	3,550	3,550	3,550
thereof: OFF	7,305	7,305	7,305
Exporters	260	260	260
Importers	260	260	260
Years	70	70	70
Coefficients	526	2	1
Exporter × importer FE	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes
ON (own)		-0.581***	-0.583***
		(0.102)	(0.102)
ON (contam.)		0.099***	-0.021
		(0.032)	(0.049)
OFF (own)		0.287***	
		(0.088)	
OFF (contam.)		0.004	
		(0.007)	
Worst-case negative bias: ON (contam.)		-0.403***	-0.320***
		(0.036)	(0.050)
Worst-case positive bias: ON (contam.)		0.403***	0.539***
		(0.036)	(0.060)
Worst-case negative bias: OFF (contam.)		-0.191***	
		(0.009)	
Worst-case positive bias: OFF (contam.)		0.187***	
		(0.012)	

Notes: The table presents regression results using an ETWFE specification (equation (1)) in column (1), a 'long' TWFE specification (equation (2)) in column (2), and a 'short' TWFE specification (equation (3)) in column (3) in the terminology of de Chaisemartin and D'Haultfœuille (2023). For the ETWFE specification, the cohort-event-time-specific treatment effects were aggregated using equation (4) and (5) to obtain aggregate ON and OFF treatment effect estimates, respectively. The dependent variable is the natural logarithm of exports, which varies over the exporter-importer-year dimension. '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 bias decomposition into own and contamination bias terms follows the decomposition proposed by de Chaisemartin and D'Haultfœuille (2023). The worst-case negative and positive bias was computed following the methodology by Goldsmith-Pinkham et al. (2024).

TWFE specification. In sum, the comparisons between the ETWFE and TWFE estimates suggest that existing TWFE estimates of the effects of sanctions on trade may be subject to significant biases.

Horizon-specific results. The objectives of this section are twofold. First, motivated by recent TWFE event-type studies of the effects of sanctions (e.g., Dai et al. (2021) and Gutmann et al. (2023)), we obtain estimates of the evolution of the effects of sanctions over time from our preferred ETWFE specification and compare them to the corresponding TWFE estimates.

Figure 4: Horizon-specific effects



Notes: The figure reports event-time-specific treatment effects from equation (1) in dark color ('ETWFE') along with standard event-study results (column (1) of Table A7) in light color ('dynamic TWFE'). Years with a star (*) refer to post-sanction years. 95% confidence intervals are shown using standard errors clustered by country pair.

Figure 4 reports our ETWFE estimates (in dark color) vs. the corresponding TWFE estimates (in light color), and we draw four main conclusions based on it. First, consistent with our main point estimates, the event-type results show that the effects of sanctions are, overall, negative, sizable, and statistically significant during the period in which sanctions are in place. Second, the effects of sanctions seem stronger five years after their implementation. This result is consistent with the findings of Dai et al. (2021), and we explore it further in Section 4.4. Third, our estimates suggest that after sanctions are lifted, trade between senders and targets shows a rapid rebound and approaches its pre-sanction level within roughly a year. This result is encouraging from a policy perspective because it implies that complete trade sanctions do not cause long-lasting damage to trade after their lifting. Finally, a comparison between the ETWFE estimates and the TWFE results reveals that the TWFE estimates are smaller during the period in which sanctions are imposed and that they are larger and, in fact, positive after the lifting of sanctions. Both of these differences have significant policy implications and, consistent with our main conclusions from the previous section, they underscore the importance of the new methods that we apply to estimate the effects of sanctions.

4.3 Bias decompositions

This section presents three bias decompositions that clarify why the TWFE models yield different results. First, following de Chaisemartin and D'Haultfœuille (2023), we decompose the TWFE coefficients into contributions from (i) cohort-event-time cells under the same treatment and (ii) cells under other treatments. Second, we decompose the TWFE bias by comparing its implicit weights with those used in the ETWFE estimand, assessing whether the bias is driven by cohorts, event time, or both. Third, following Goldsmith-Pinkham et al. (2024), we implement a decomposition of contamination bias to isolate its sources. Additional evidence of the TWFE's implicit weights is reported in Section A.3 in the Appendix.

Own bias and contamination bias decomposition following de Chaisemartin and D'Haultfœuille (2023). We capitalize on recent developments in the econometrics literature to explore the nature and sources of these biases. The bottom two panels of Table 2 report the bias of the TWFE models in columns (2) and (3). First, we consider the decomposition proposed by de Chaisemartin and D'Haultfœuille (2023), which breaks down the TWFE coefficients into the product of cohort-event-time-specific treatment effects from the ETWFE regression and implicit weights attached to these cohort-event-time cells by the TWFE regression.[29] This allows us to decompose the TWFE coefficients into two terms: one related to cohort-event-time effects of the same treatment (labeled

[29]See also Sun and Abraham (2021) and Goldsmith-Pinkham et al. (2024).

'own') and another related to cohort-event-time effects of other treatments (labeled 'contamination').[30] The first term equals the bias of the TWFE regression with a single treatment (e.g., de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2024; Wooldridge, 2025), while the second term is new and appears only in the presence of multiple treatments with treatment effect heterogeneity (Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2023; Goldsmith-Pinkham et al., 2024).

The decomposition results show that the estimate of sanctions, $ON_{ij,t}$, in the 'long' TWFE specification (equation 2) in column (2) is downward-biased for two reasons: (i) the TWFE estimate overweights small sanction effects, which accounts for approximately two-thirds of the downward bias in the coefficient; (ii) the TWFE estimate also includes treatment effects from years after the sanction was lifted, accounting for around one-third of the bias. Both terms are also statistically significant. This illustrates the arbitrary weighting of within-category treatment effects by the TWFE specification in the presence of heterogeneous treatment effects (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2024; Wooldridge, 2025). Moreover, this supports the findings in de Chaisemartin and D'Haultfœuille (2023) and Goldsmith-Pinkham et al. (2024) that contamination bias in observational settings with multiple treatments can be economically and statistically significant. However, in the case of the OFF effect, we find that the bias derives essentially only from the own terms, and the contamination bias term is close to zero.

Next, we consider the 'short' TWFE specification (equation 3) in column (3), which includes a single indicator equal to one while a sanction is in force and zero otherwise. This is akin to the most widely used specification in the literature, which does not explicitly estimate the effect of exit. Encouragingly, the overall bias of the specification is smaller. While the bias terms resulting from the 'own' treatment effect remain the same, the contamination bias term is much smaller and close to zero in this particular case. This

[30] A detailed analysis of the implicit weights attached by the TWFE regression is provided in Section A.3 in the Appendix.

finding is consistent with the theoretical result in de Chaisemartin and D'Haultfœuille (2023), who show that omitting the second treatment can reduce the bias of the first treatment in the presence of treatment effect heterogeneity. In contrast, under constant treatment effects, omitting the second treatment generates standard omitted-variable bias, so including the second treatment is preferable.

Bias contributions from cohort- and event-year-dimensions. Building on the fact that the TWFE coefficients can be written as weighted averages of cohort-event-year treatment effects under treatment effect heterogeneity (de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2024), let $\Delta w_{c,\ell} \equiv w_{c,\ell}^{\mathrm{TWFE}} - w_{c,\ell}^{\mathrm{ETWFE}}$ denote the weight discrepancy for cohort c at event year ℓ, where, for notational convenience and with a slight abuse of notation, c indexes the (g,h) cohorts in equation (1).[31] We summarize this discrepancy as an additive decomposition into a cohort component, an event-year component, and a residual interaction:

$$\Delta w_{c,\ell} = \underbrace{\Delta w_{c\cdot}}_{\text{cohort}} + \underbrace{\Delta w_{\cdot\ell}}_{\text{event year}} + \underbrace{r_{c,\ell}}_{\text{interaction}},$$

where $\Delta w_{c\cdot}$ is the mean of $\Delta w_{c,\ell}$ over event years observed for cohort c, and $\Delta w_{\cdot\ell}$ is the mean over cohorts observed at event year ℓ.[32] The residual $r_{c,\ell} \equiv \Delta w_{c,\ell} - \Delta w_{c\cdot} - \Delta w_{\cdot\ell}$ captures what remains after removing the cohort and event-year means, i.e., a cohort×event-year-specific component.

We map these components into bias contributions by weighting with the ETWFE cohort-event-year treatment effects $\delta_{c,\ell}$:

$$B_{\mathrm{cohort}} = \sum_{c,\ell} \Delta w_{c\cdot}\, \delta_{c,\ell}, \qquad B_{\mathrm{event}} = \sum_{c,\ell} \Delta w_{\cdot\ell}\, \delta_{c,\ell}, \qquad B_{\mathrm{res}} = \sum_{c,\ell} r_{c,\ell}\, \delta_{c,\ell},$$

so that the total bias $B \equiv \beta_{\mathrm{TWFE}} - \beta_{\mathrm{ETWFE}} = B_{\mathrm{cohort}} + B_{\mathrm{event}} + B_{\mathrm{res}}$.

[31] For exposition, all objects in this subsection are presented as population quantities. In the empirical analysis we use their estimated counterparts (denoted with hats), and the expressions carry over unchanged.

[32] Because the number of event years differs across cohorts due to entry and exit, all sums and averages in this subsection are taken over the set of observed cohort-event-year cells only.

Table 3: Bias contributions from cohort and event-year dimensions

	ON		OFF	
	Bias	Share	Bias	Share
Cohort	0.549	1.453	0.195	0.883
Event-year	0.067	0.178	0.000	0.000
Residual	-0.238	-0.632	0.026	0.117
Total	0.378	1.000	0.220	1.000

Notes: The table splits the TWFE-ETWFE difference into three components: (i) a cohort margin, (ii) an event-year margin, and (iii) a residual cohort×event-year interaction. "Bias" reports each component in levels. "Share" reports the fraction of the total (shares can be negative or exceed one when other components offset). Deviations from totals are due to rounding. See the main text for definitions and implementation details.

The results in Table 3 indicate that the cohort component is the predominant source of the difference between the TWFE and ETWFE estimands for both the sanction effect (ON) and the lifting effect (OFF). For ON, the total bias equals 0.378, with contributions $B_{\text{cohort}} = 0.549$ (share: 1.453), $B_{\text{event}} = 0.067$ (share: 0.178), and a negative interaction contribution $B_{\text{res}} = -0.238$ (share: -0.632), which offsets a substantial part of the cohort and event-year components. For OFF, the total bias equals 0.220, with $B_{\text{cohort}} = 0.195$ (share: 0.883), a negligible event-year component $B_{\text{event}} = 0.000$ (share: 0.000), and a positive interaction contribution $B_{\text{res}} = 0.026$ (share: 0.117). Overall, these calculations point to reweighting across cohorts as the main source of bias in this setting, with reweighting across event years of second-order importance. At the same time, the interaction component is also relevant (mitigating the bias for ON and modestly amplifying it for OFF) so allowing for full cohort×event-year heterogeneity is necessary for unbiased estimates. This conclusion is consistent with our robustness analysis on the degree of heterogeneity in Section 6.

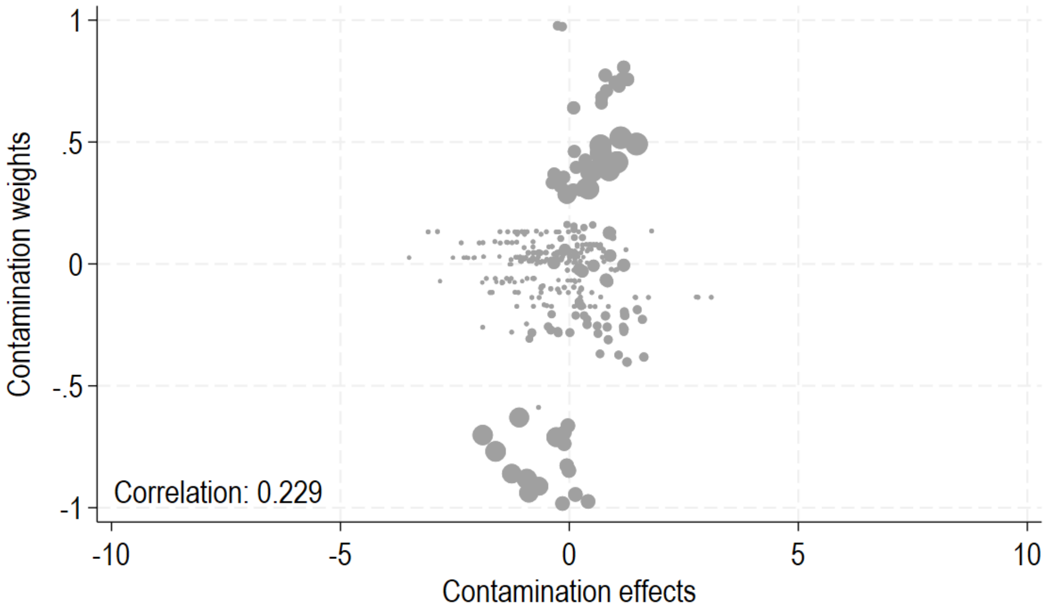
Contamination bias decomposition following Goldsmith-Pinkham et al. (2024). We complete the bias analysis by following Goldsmith-Pinkham et al. (2024) to study an additional decomposition of the contamination bias. Given that the contamination weights sum to zero in the setting with mutually exclusive treatments considered here, the contamination bias can be written as the correlation between contamination weights and the other treatment effects multiplied by their respective standard deviations (Goldsmith-Pinkham et al., 2024). Therefore, note that under treatment effect homogeneity, contamination bias is zero in our setting.

Figure 5 provides a graphical illustration of this decomposition for the sanction (ON) and lifting (OFF) effects. In both cases, there is substantial variation in the underlying treatment effect (i.e., treatment effect heterogeneity) as well as variation in the contamination weights.[33] For the sanction effect, the contamination weights are positively correlated with the treatment effects of the lifting of sanctions, with a (Pearson) correlation coefficient of 0.229. By contrast, for the lifting effect, the correlation between contamination weights and sanction effects is close to zero, which is behind the insignificant contamination bias in this case.

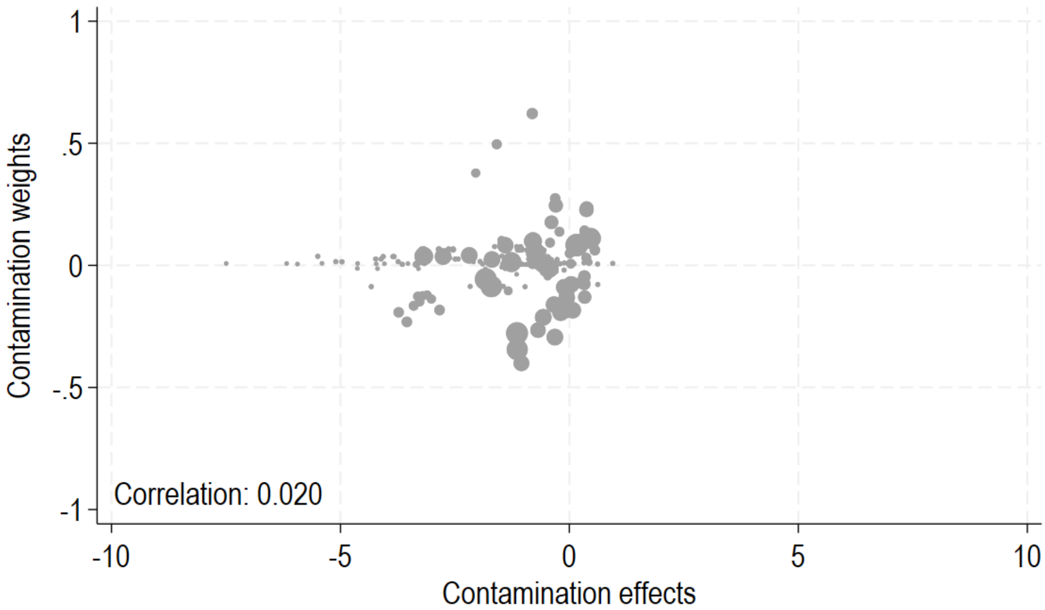
We also use this decomposition to compute a worst-case bias, which assumes maximal positive or negative correlation between contamination weights and treatment effects (Goldsmith-Pinkham et al., 2024). In practice, we reorder the contamination weights and treatment effects to maximize their correlation in each direction and recompute the corresponding weighted averages. The bottom panel of Table 2 shows that the potential contamination bias can be substantially larger than the actual contamination bias. For the ON effect, the interval is $[-0.403, 0.403]$ and, for the OFF effect, the interval is $[-0.191, 0.187]$. The worst-case estimates illustrate that the substantial treatment effect heterogeneity in our setting makes a sizable contamination bias possible in either direction, whereas the actual contamination is relatively small because the correlation between the contamination weights and treatment effects is low.

[33]For the ON effect, the (unadjusted) standard deviation of the contamination weights is 0.456 and the standard deviation of the OFF treatment effects is 0.956. For the OFF effect, the (unadjusted) standard deviation of the contamination weights is 0.117 and the standard deviation of the ON treatment effects is 1.800. Since the product of the standard deviations is the same order of magnitude (0.436 vs. 0.211), the main reason for the difference in contamination bias across the two cases is the correlation between contamination weights and treatment effects.

Figure 5: Correlation between contamination weights and treatment effects



(a) ON sanction effect



(b) OFF sanction effect

Notes: The figure reports the contamination weights attached by the OLS TWFE regression from column (2) in Table 2 computed following de Chaisemartin and D'Haultfœuille (2023) along with the treatment effects of the other treatment at the cohort-year level from column (1) in Table 2. Note that in this figure, the contamination weights (used in Figures A3 and A4) are scaled by the number of treated observations for enhanced visibility, in line with the definition by Goldsmith-Pinkham et al. (2024). However, due to scale invariance, this adjustment has no impact on the results. The size of the points is proportional to the number of observations in each cohort-year cell. Panel (a) depicts the correlation between the contamination effects (i.e., the OFF treatment effects) against the estimated contamination weight for the ON treatment effect. Panel (b) depicts the correlation between the contamination effects (i.e., the ON treatment effects) against the estimated contamination weight for the OFF treatment effect.

4.4 Ex-post heterogeneity of sanction effects by observable characteristics

In the next step of the analysis, we examine how average sanction effects vary with a rich set of observables: (i) *bilateral geography* (distance, contiguity); (ii) *colonial ties and language* (past colonial relationship, common language); (iii) *pre-sanction trade integration* (WTO co-membership and regional trade agreements (RTAs) measured at $g-1$); (iv) *sanction case attributes* (duration, war-related objective, ex-post outcome); (v) *concurrent financial sanctions and sanction history* (financial sanctions during the episode and any prior non-trade sanctions in the dyad); and (vi) *income level* (high-income exporter and importer indicators). Specifically, we take the cohort-event-time treatment effects $\hat{\delta}_{gh,s-g}$ from the baseline ETWFE gravity specification and regress them on these covariates. Time-varying variables are measured in the year before entry $(g-1)$. To absorb common dynamics, we include a full set of event-time fixed effects. Inference is based on a parametric bootstrap that draws 999 coefficient vectors $\tilde{\boldsymbol{\delta}}^{(b)} \sim \mathcal{N}\big(\hat{\boldsymbol{\delta}}, \widehat{\mathrm{V}}(\hat{\boldsymbol{\delta}})\big)$ from the first-step baseline ETWFE estimates, and for each draw we reestimate the ex-post regression. We estimate separate models for *ON* effects (during sanctions) and *OFF* effects (after lifting), and report the results in Table 4. Methodologically, the exercise follows the ex-post heterogeneity analyses in Shahn (2023) and de Chaisemartin et al. (2025). From a policy perspective, it is related to Felbermayr et al. (2025b), who study heterogeneity in sanctions' trade effects using interactions in a TWFE framework.

Geography. Distance robustly moderates the sanction effect on trade. During sanction imposition (columns (1)-(7)), the coefficient on $\ln \mathrm{Distance}_{ij}$ is negative and precisely estimated. Similarly, after the lifting of sanctions (columns (8)-(14)), the coefficient has the same sign and slightly larger magnitudes in parsimonious models. Contiguity is likewise associated with more negative treatment effects, both ON and OFF (although the coefficient loses significance in more stringent specifications for the ON sample), with large and precisely estimated coefficients. The distance pattern is consistent with evidence that long-haul trade is more time- and finance-intensive, which makes it more vulnerable to disruptions and slower to normalize (Ahn et al., 2011; Hummels and Schaur, 2013; Schmidt-Eisenlohr, 2013). A more negative effect for contiguous partners is consistent with higher sanction-related compliance and monitoring frictions at land crossings than on maritime or air routes, shifting trade away from the neighbor. This aligns with evidence that border procedures and clearance delays depress trade (Djankov et al., 2010; Volpe Martincus et al., 2015) and with structural-gravity estimates that border costs are a salient trade friction (Anderson and van Wincoop, 2003; Yotov et al., 2016).

Colonial ties and language. Conditioning on other observable characteristics, the coefficient on a colonial relationship is statistically indistinguishable from zero in most specifications. A common language is likewise not systematically related to the estimated trade effect of sanctions before and after lifting.

Pre-sanction trade integration. We next ask whether rules-based ties, measured prior to entry, insulate trade during sanctions or facilitate a post-lifting rebound once sanctions are lifted. WTO co-membership in the pre-sanction period is positively associated with the treatment effects in the OFF sample, implying stronger recovery after exit. Pre-existing RTAs are positively associated with the treatment effect during sanction imposition in the more comprehensive models (columns (5)-(7)). One interpretation is that rules-based arrangements reduce policy uncertainty, standardize procedures, and sustain trading relationships, thereby cushioning contemporaneous disruptions and lowering the fixed costs of re-entry, broadly in line with evidence on RTAs and WTO-related institutional effects in gravity settings (e.g., Baier and Bergstrand, 2007; Handley and Limão, 2017; Baier et al., 2019; Larch et al., 2025).

Sanction case attributes: duration, objective, and success. Sanction-case characteristics are also systematically associated with variation in the estimated sanction effects.[34] The indicator for long duration (i.e., sanctions lasting more than six years) is not significantly different from zero for the ON effect across specifications, yet strongly associated with more negative OFF effects. We simply note this pattern here and analyze it in more detail below. Cases coded as unsuccessful in achieving their stated objective are associated with larger ON losses in one specification (column (5)) and more negative OFF effects, a pattern in line with accounts that difficult cases often involve escalation and leave deeper trade scars (Hufbauer et al., 2007). Finally, sanctions motivated by war-related objectives (war prevention, war termination, or territorial disputes) are linked to larger trade contractions during imposition, consistent with evidence that security-oriented sanctions tend to be broader in scope and more stringently enforced (Felbermayr et al., 2020; Crozet and Hinz, 2020).[35]

Concurrent financial sanctions and sanction history. We also consider concurrent financial sanctions and a history of other sanctions in the dyad.[36] First, contemporaneous financial sanctions are neither significantly associated with the treatment effect during sanction imposition nor after lifting. Second, a prior history of other (non-trade) sanctions is linked to larger ON losses. This pattern is consistent with enforcement complementarity and reputational risk, whereby earlier episodes establish compliance routines and tighter monitoring, which increases the marginal bite of trade sanctions (Besedeš and Prusa, 2006a; Crozet et al., 2021).

Income level. Conditioning on other observables, indicators for a high-income exporter and a high-income importer are positively associated with the sanction effect in both the ON and OFF samples, albeit only the high-income importer indicator is marginally sta-

[34]For completeness and transparency, Table A2 in the Appendix reports additional subgroup-specific regressions that disaggregate objectives (end war, prevent war, territorial conflict, democracy, regime destabilization, human rights, policy change) and outcomes (partial success, negotiated settlement).

[35]The 'war-related objective' captures the sanction's stated motive and does not imply an ongoing bilateral war. See footnote 26 for a sample restriction that excludes dyads ever at war.

[36]The history variable pertains to non-trade sanctions only (e.g., arms, military, finance, travel, and other measures). In our main analysis, we treat the first complete trade sanction in the dataset as the treatment, so by construction there is no prior complete trade sanction.

tistically significant during sanction imposition. One interpretation is that richer partners may better sustain permitted flows in some cases and adjust along compliant margins supported by deeper trade finance and firm capabilities thereby attenuating contemporaneous losses and facilitating post-lifting normalization (Manova, 2013; Crozet et al., 2021).

Table 4: Subgroup-specific results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	ON	*ON*	*ON*	*ON*	*ON*	*ON*	*ON*	*OFF*	*OFF*	*OFF*	*OFF*	*OFF*	*OFF*	*OFF*
ln Distance$_{ij}$	-0.232***	-0.214***	-0.207***	-0.203***	-0.206***	-0.181**	-0.178**	-0.172***	-0.232***	-0.177**	-0.183***	-0.162***	-0.166***	-0.179***
	(0.055)	(0.060)	(0.056)	(0.057)	(0.057)	(0.071)	(0.072)	(0.059)	(0.085)	(0.071)	(0.068)	(0.054)	(0.047)	(0.050)
Contiguity$_{ij}$	-0.391**	-0.398***	-0.393***	-0.409***	-0.389***	-0.249	-0.219	-0.698***	-0.601***	-0.551***	-0.560***	-0.544***	-0.520***	-0.510***
	(0.162)	(0.106)	(0.102)	(0.099)	(0.096)	(0.164)	(0.163)	(0.211)	(0.160)	(0.142)	(0.142)	(0.143)	(0.110)	(0.106)
Colonial relationship$_{ij}$	-0.188	-0.137	-0.109	-0.160	-0.008	-0.032	-0.018	-0.339	-0.355	0.035	0.001	0.022	0.026	0.060
	(0.122)	(0.150)	(0.131)	(0.119)	(0.099)	(0.100)	(0.094)	(0.240)	(0.249)	(0.125)	(0.129)	(0.142)	(0.156)	(0.137)
Common language$_{ij}$	-0.012	-0.022	-0.059	0.014	0.008	-0.051	-0.039	0.368*	0.340*	0.034	0.066	0.076	0.067	0.067
	(0.124)	(0.122)	(0.098)	(0.086)	(0.087)	(0.094)	(0.097)	(0.210)	(0.197)	(0.093)	(0.100)	(0.096)	(0.094)	(0.093)
WTO$_{ij,g-1}$		0.415	0.391	0.426	0.246	0.327*	0.290		0.673**	0.387*	0.402*	0.398*	0.394***	0.379***
		(0.327)	(0.309)	(0.301)	(0.237)	(0.192)	(0.188)		(0.330)	(0.218)	(0.215)	(0.213)	(0.149)	(0.141)
RTA$_{ij,g-1}$		0.210	0.224	0.189	0.326***	0.198***	0.269***		-0.156	0.058	0.027	0.037	0.034	0.109
		(0.182)	(0.175)	(0.164)	(0.122)	(0.063)	(0.062)		(0.226)	(0.158)	(0.152)	(0.149)	(0.125)	(0.085)
Long duration$_{gh}$			-0.137	-0.117	0.042	0.268	0.275			-0.896**	-0.876**	-0.840**	-0.827**	-0.847**
			(0.242)	(0.239)	(0.230)	(0.299)	(0.311)			(0.411)	(0.412)	(0.405)	(0.329)	(0.350)
Outcome: Failed$_{gh}$				-0.494	-0.844***	-0.459	-0.401				-0.302	-0.424*	-0.382**	-0.327*
				(0.340)	(0.307)	(0.350)	(0.323)				(0.291)	(0.219)	(0.181)	(0.178)
Objective: War related$_{gh}$					-0.871**	-1.091***	-1.105***					-0.251	-0.247	-0.275
					(0.384)	(0.261)	(0.259)					(0.355)	(0.337)	(0.331)
Financial sanction$_{ijt}$						0.544	0.597						-0.050	0.014
						(0.458)	(0.437)						(0.585)	(0.546)
Other sanctions in the past$_{ij}$						-0.263***	-0.236***						-0.087	-0.044
						(0.089)	(0.087)						(0.128)	(0.127)
High income exporter$_{i}$							0.094							0.176
							(0.137)							(0.204)
High income importer$_{j}$							0.182*							0.097
							(0.096)							(0.094)
Observations	3,517	3,505	3,505	3,505	3,505	3,505	3,505	7,272	7,186	7,186	7,186	7,186	7,186	7,186

Notes: The table reports OLS regression results of the estimated cohort-event-time effects $\hat{\delta}_{g,h,s-g}$ from the baseline ETWFE results of column (1) in Table 2 on dyad, sanction-episode, and country covariates. All specifications include event-time fixed effects and are estimated separately for the *ON* (treatment effects during imposition) and *OFF* (treatment effects after lifting) samples. *Gravity covariates*: $\ln$ Distance$_{ij}$, Contiguity$_{ij}$, Colonial relationship$_{ij}$, Common language$_{ij}$, WTO$_{ij,g-1}$, and RTA$_{ij,g-1}$ are taken from the CEPII's Gravity database (Mayer and Zignago, 2011). Distance, contiguity, colonial relationship, and common language vary at the country-pair level (i,j); WTO co-membership and RTA status are measured in the year before entry $(g-1)$. *Sanction characteristics*: Long duration$_{gh}$ (episode lasts more than six years), Outcome : Failed$_{gh}$, Objective : War related$_{gh}$ (objective coded as *end war*, *prevent war*, or *territorial conflict*), Financial sanction$_{ijt}$ (financial sanctions present in year t of the episode), and Other sanctions in the past$_{ij}$ (any prior non-trade sanctions in the dyad before the first complete trade sanction) come from the Global Sanctions Database (Syropoulos et al., 2024). *Income*: High income exporter$_{i}$ and High income importer$_{j}$ are based on World Bank income groups. Standard errors in parentheses are obtained using a parametric bootstrap procedure with 999 repetitions. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Horizon-specific effects by sanction duration. Following Dai et al. (2021), we use our ETWFE baseline results and separately aggregate the cohort-year effects for short sanctions (lasting 6 years or less) versus long sanctions (lasting more than 6 years).[37] The resulting paths in Figure 6 show clear differences.[38] For short sanctions (Figure 6a), the treatment effects are initially negative and statistically significant, especially in the later years. However, after sanctions are lifted, the negative effect diminishes and the post-lifting coefficients turn positive for short-duration episodes, indicating a clear rebound in trade flows relative to pre-sanction levels.[39] Long sanctions (Figure 6b) exhibit a similarly pronounced negative impact, which generally increases with duration, with significant reductions observed over time. In contrast to short sanctions, the post-sanction periods show some modest improvement, but the coefficients remain negative and do not return to their pre-sanction baseline values (although they are not statistically different from zero). One possible interpretation is that short interruptions mainly generate pent-up demand and inventory dynamics that permit quick normalization (Alessandria et al., 2010; Behrens et al., 2013), whereas long episodes destroy relationships and induce re-sorting to third markets making re-entry costly and incomplete (Melitz, 2003; Besedeš and Prusa, 2006b; Monarch, 2016; Martin et al., 2023).[40]

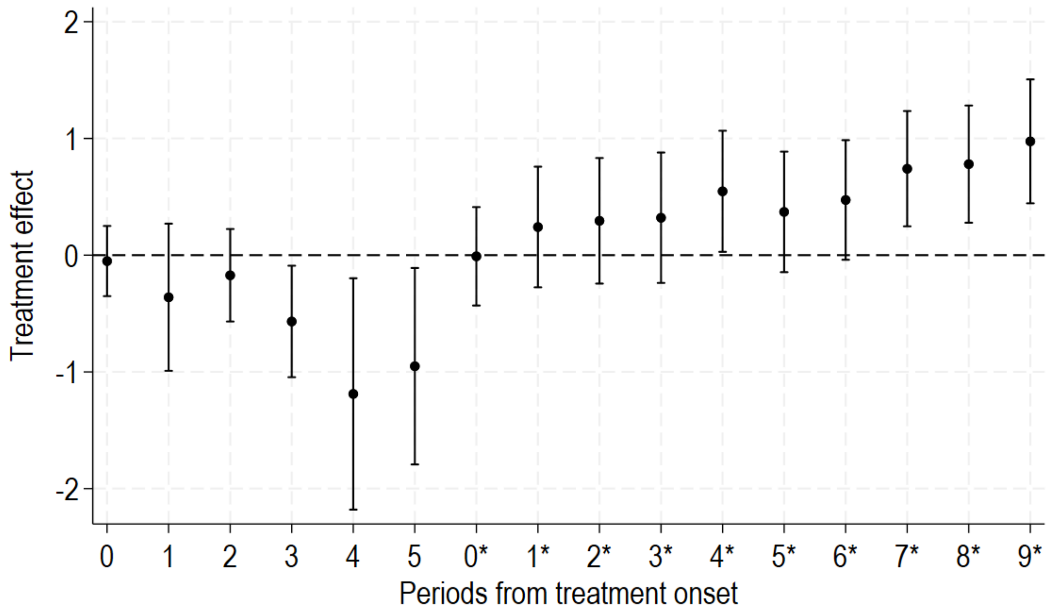
[37]Note that there is no need to allow for duration-specific effects separately, as the entry and exit years of cohorts implicitly define duration.

[38]These contrasts are not driven by pre-trends, and a duration-interacted dynamic TWFE specification (Appendix Figure A7) delivers very similar profiles.

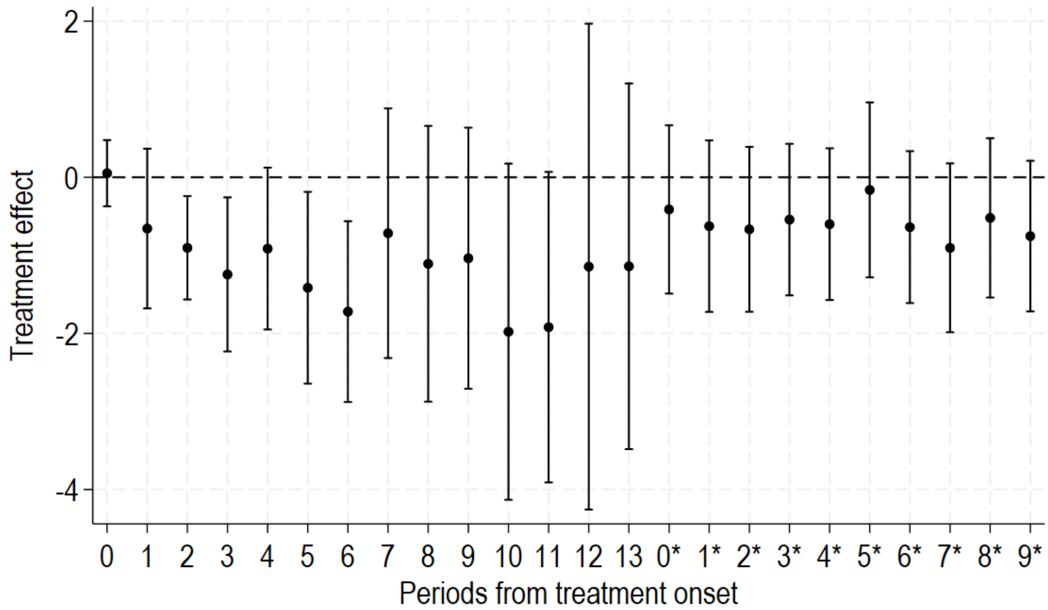
[39]This rebound reflects period-by-period deviations from the pre-sanction baseline and does not identify whether cumulative post-lifting gains offset the losses incurred during imposition. Assessing full cumulative recovery would require additional analysis beyond what the event-study design can provide.

[40]Complementary results from regressions that interact a long-duration indicator with covariates (geography, colonial ties/language, pre-sanction integration, sanction attributes, concurrent financial sanctions and sanction history, and income) are reported in Section A.4.1 and Figure A6 in the Appendix.

Figure 6: Horizon-specific effects by duration



(a) Short sanctions



(b) Long sanctions

Notes: The figure reports event-time-specific treatment effects from equation (1). Years with a star (*) refer to post-sanction years. Short sanctions are sanctions with a duration of 6 years or less. Long sanctions are sanctions with a duration between 7 and 14 years. 95% confidence intervals are shown using standard errors clustered by country pair.

5 Additional results

To deepen our understanding of the mechanisms behind the baseline results and to broaden their empirical scope, we present a set of additional analyses. For clarity, we organize the material into five themes: (i) responses at the extensive margin, (ii) outcomes under other types of sanctions, (iii) sectoral heterogeneity, (iv) raw data evidence from major sanction cohorts, and (v) an application to currency unions. For brevity, we summarize the headline findings in the main text and relegate implementation details and full results tables to the Appendix.

Extensive margin. We examine year-to-year switches between zero and non-zero trade flows at the country-pair level to assess extensive-margin adjustments. Sanctions raise the probability that an active trade link falls to zero by 4.3 percentage points and reduce the probability that a dormant link turns positive by 5.3 percentage points. When sanctions are lifted, entry probabilities rise significantly (by 4.9 percentage points), consistent with the resumption of relationships disrupted during the sanction episode, while exit probabilities return to pre-sanctions levels.

Other types of sanctions. Next, we examine whether other sanction types also reduce trade. We estimate specifications for partial trade sanctions, financial sanctions, and other sanctions (arms, military, travel). None of these alternative sanction types significantly affect bilateral trade, indicating that the trade-reducing effects we document are specific to complete trade sanctions rather than a general feature of sanctions.

Sectoral heterogeneity. Using UN Comtrade data from 1963 onward, we classify bilateral flows into five sectors and find substantial heterogeneity. Dual-use manufacturing goods, which are strategically most sensitive, exhibit the strongest negative effect during sanctions (-0.727). Mining shows no significant impact, consistent with common exemptions for energy products. For agriculture, we find only limited evidence of post-sanction improvement, consistent with the possibility that disrupted trade relationships were replaced by alternative suppliers during sanctions.

Raw data evidence from major sanction cohorts. To provide raw data evidence that complements our econometric estimates, we plot trade flows for all sanction episodes in aggregate as well as for the six largest sanction cohorts individually, which together account for over 80% of all sanctioned pairs. The figures reveal clear declines in trade following sanction imposition, with particularly pronounced effects for comprehensively enforced episodes such as UN sanctions on Iraq (1990-2003), while the magnitude of effects varies across cases in ways consistent with anecdotal reports of enforcement stringency.

Currency union application. To demonstrate the broader applicability of our methodology, we apply it to currency unions, which exhibit a similar staggered treatment structure with entry and exit. Re-estimating our baseline specification, we find positive and significant effects of currency union formation on trade across all specifications, with ETWFE yielding the largest estimate. Dissolutions also show positive effects, though less precisely estimated due to the small number of exits, where TWFE results exhibit contamination effects. Overall, the exercise confirms that the ETWFE methodology with exit extends well beyond sanctions.

6 Robustness analysis

To demonstrate the robustness of our baseline findings, we conduct a series of robustness checks across five key dimensions: (i) different samples, (ii) the role of small trade flows, (iii) degree of model heterogeneity, (iv) inclusion of domestic trade flows, and (v) alternative clustering approaches. For brevity, we summarize the main results in the text and relegate full implementation details and results tables to the Appendix.

Different samples. We assess the robustness to alternative sample definitions, including (i) all sanction durations rather than restricting to episodes lasting 14 years or less, (ii) all post-exit observations rather than limiting to 10 years after lifting, and (iii) requiring a minimum of five observations per treated cohort-event-time cell to improve precision. Across all specifications, the ETWFE estimates remain consistent with our baseline results, confirming that our findings are not driven by specific sample restrictions.

The role of small trade flows. To assess whether our results are driven by small trade flows, we conduct two tests: restricting the sample to the 200 largest exporters and implementing weighted least squares using time-invariant weights based on initial bilateral trade flows. In both cases, the ETWFE estimates remain consistent with our baseline findings, while the TWFE estimates prove more sensitive to the inclusion or weighting of small flows.

Degree of heterogeneity of the model. To understand the drivers of the differences between the ETWFE and TWFE estimates, we impose restrictions on treatment effect heterogeneity. Allowing variation only along event time produces estimates close to the TWFE baseline, while allowing variation only across cohorts yields intermediate results. Models with full cohort heterogeneity but restricted time heterogeneity (grouped every 2, 3, or 5 years) converge toward the baseline ETWFE estimates as time flexibility increases. These results underscore that cohort heterogeneity is crucial for capturing treatment effects, while time heterogeneity plays a secondary but non-negligible role.

Including domestic trade flows. We re-estimate our models using a dataset that incorporates domestic trade flows from the CEPII TradeProd database (Mayer et al., 2023) for the years 1966-2020. Comparing results with and without internal flows (using the same 1966-2019 window), we find that ETWFE and TWFE estimates closely mirror our baseline patterns. When domestic trade is included together with cross-border$\times$year fixed effects to absorb globalization effects the estimated ON effects become more negative, reflecting the diversion of sanctioned countries' trade toward domestic markets. OFF effects remain similar, i.e., insignificant for ETWFE and positive for TWFE. Overall, including domestic trade flows leaves our main conclusions unchanged.

Alternative clustering. We examine the robustness of our inference to alternative clustering methods, including double-clustering by exporter and importer, by exporter-year and importer-year, and triple-clustering by exporter, importer, and year. The ON estimate remains statistically significant across all specifications, confirming the robustness of our main findings.

7 Conclusion

The objective of our study was to estimate the impact of sanctions on trade. Unlike many other international policies that remain indefinitely once enacted, economic sanctions are frequently lifted, often as frequently as they are imposed. Utilizing recent advancements in DiD methods for treatments with both entry and exit, we show that traditional TWFE approaches significantly underestimate the impact of sanctions during their enforcement and overestimate their lingering effects after being lifted.

Our analysis provides compelling empirical evidence for the existence and significance of contamination bias in an economically relevant context. We find that the bias in TWFE estimates arises not only from the negative weighting of treatment effects but also from substantial contamination across different treatments. The ETWFE approach that we employ avoids these issues, revealing that sanctions reduce bilateral trade by approximately 58% while they are in effect, i.e., about 50% more than what TWFE estimates suggest. Furthermore, we find no significant persistent effects on trade after sanctions are lifted, in contrast to the positive and significant post-sanction estimates derived from TWFE models. Our worst-case bias calculations further illustrate the potential for even larger contamination bias, given the heterogeneity in treatment effects across different sanction episodes.

These results complement existing approaches to avoid contamination bias, particularly those focusing on non-mutually exclusive treatments. Although the ETWFE approach is limited to cases involving exit, it has the advantage of being directly applicable in a gravity model setting. The ETWFE framework can be readily applied to other bilateral policies involving treatments with exit, such as trade agreements and migration policies, where contamination bias may similarly distort conventional estimates. Future research could extend existing bias decomposition methods to non-linear settings and explore their implications for estimating the effects of other time-varying bilateral policies on trade and other economic outcomes.

In conclusion, our study emphasizes the need for adopting heterogeneity-robust methods for treatments with both entry and exit, such as the ETWFE approach, to avoid contamination bias and achieve more accurate estimates of policy impacts. By highlighting the importance of accounting for both the imposition and removal of economic sanctions, our study contributes to a more precise understanding of their true economic impacts and provides a robust methodological framework for future policy evaluations in international economics. Our ex-post heterogeneity analysis further shows that the magnitude and, in several cases, the persistence of trade losses vary systematically with observable dyad and episode characteristics. In particular, distant or contiguous partners, war-related and unsuccessful cases, and dyads with a history of other sanctions suffer larger trade contractions, whereas pre-sanction rule-based trade integration and higher income levels, especially on the importer side, are associated with smaller contemporaneous losses and, in several cases, faster post-lifting improvements.

References

Afesorgbor, Sylvanus Kwaku. 2019. "The impact of economic sanctions on international trade: How do threatened sanctions compare with imposed sanctions?" *European Journal of Political Economy* 56 11–26.

Ahn, JaeBin, Mary Amiti, and David E. Weinstein. 2011. "Trade Finance and the Great Trade Collapse." *American Economic Review* 101 (3): 298–302. 10.1257/aer.101.3.298.

Alessandria, George, Joseph P. Kaboski, and Virgiliu Midrigan. 2010. "Inventories, Lumpy Trade, and Large Devaluations." *American Economic Review* 100 (5): 2304–2339.

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

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

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

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

Behrens, Kristian, Gregory Corcos, and Giordano Mion. 2013. "Trade Crisis? What Trade Crisis?" *Review of Economics and Statistics* 95 (2): 702–709.

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

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2009. "The Margins of US Trade." *American Economic Review Papers and Proceedings* 99 (2): 487–493.

Besedeš, T., and T.J. Prusa. 2006a. "Ins, Outs, and the Duration of Trade." *Canadian Journal of Economics* 39 (1): 266–295.

Besedeš, T., and T.J. Prusa. 2006b. "Product Differentiation and Duration of US Import Trade." *Journal of International Economics* 70 (2): 339–358.

Besedeš, Tibor, Stefan Goldbach, and Volker Nitsch. 2017. "You're banned! The effect of sanctions on German cross-border financial flows." *Economic Policy* 32 (90): 263–318. 10.1093/epolic/eix001.

Bista, Rishav, and Brandon Sheridan. 2025. "Economic Sanctions and Export Margins." *Review of International Economics* 33 (2): 423–435. https://doi.org/10.1111/roie.12784.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies* 91 (6): 3253–3285. 10.1093/restud/rdae007.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics* 225 (2): 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001, Themed Issue: Treatment Effect 1.

Caruso, Raul. 2003. "The impact of international economic sanctions on trade: An empirical analysis." *Peace economics, peace science and public policy* 9 (2): 41–66.

Crozet, Matthieu, and Julian Hinz. 2020. "Friendly fire: The trade impact of the Russia sanctions and counter-sanctions." *Economic Policy* 35 (101): 97–146.

Crozet, Matthieu, Julian Hinz, Amrei Stammann, and Joschka Wanner. 2021. "Worth the pain? Firms' exporting behaviour to countries under sanctions." *European Economic Review* 134 103683. https://doi.org/10.1016/j.euroecorev.2021.103683.

Dai, Mian, Gabriel Felbermayr, Aleksandra Kirilakha, Constantinos Syropoulos, Erdal Yalcin, and Yoto Yotov. 2021. "Timing the Impact of Sanctions on Trade." Edward Elgar Publishing Chapter 22, pages 411-437, Research Handbook on Economic Sanctions.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964–2996. 10.1257/aer.20181169.

de Chaisemartin, Clément, Xavier D'Haultfœuille, Félix Pasquier, Doulo Sow, and Gonzalo Vazquez-Bare. 2025. "Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period." `https://arxiv.org/abs/2201.06898`, arXiv:2201.06898.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2022. "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey." *The Econometrics Journal* 26 (3): C1–C30. 10.1093/ectj/utac017.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2023. "Two-way fixed effects and differences-in-differences estimators with several treatments." *Journal of Econometrics* 236 (2): 105480. https://doi.org/10.1016/j.jeconom.2023.105480.

de Sousa, José. 2012. "The Currency Union Effect on Trade Is Decreasing over Time." *Economics Letters* 117 (3): 917–920. 10.1016/j.econlet.2012.07.009.

Djankov, Simeon, Caroline Freund, and Cong S Pham. 2010. "Trading on Time." *The Review of Economics and Statistics* 92 (1): 166–173. 10.1162/rest.2009.11498.

Dube, Arindrajit, Daniele Girardi, Òscar Jordà, and Alan M. Taylor. 2025. "A Local Projections Approach to Difference-in-Differences." *Journal of Applied Econometrics* n/a (n/a): 1–18. https://doi.org/10.1002/jae.70000.

Dutt, Pushan, Ana Maria Santacreu, and Daniel A. Traça. 2022. "The gravity of experience." *Canadian Journal of Economics/Revue canadienne d'économique* 55 (1): 213–248. https://doi.org/10.1111/caje.12583.

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

Felbermayr, Gabriel, Aleksandra Kirilakha, Constantinos Syropoulos, Erdal Yalcin, and Yoto V. Yotov. 2020. "The global sanctions data base." *European Economic Review* 129 103561.

Felbermayr, Gabriel, Clifton Morgan, Constantinos Syropoulos, and Yoto Yotov. 2025a. "Economic Sanctions: Stylized Facts and Quantitative Evidence." *Annual Review of Economics* 17 175–195.

Felbermayr, Gabriel, T. Clifton Morgan, Constantinos Syropoulos, and Yoto V. Yotov. 2021. "Understanding economic sanctions: Interdisciplinary perspectives on theory and evidence." *European Economic Review* 135 103720.

Felbermayr, Gabriel, Constantinos Syropoulos, Erdal Yalcin, and Yoto V. Yotov. 2025b. "On the heterogeneous effects of sanctions on trade." *Canadian Journal of Economics/Revue canadienne d'économique* 58 (1): 247–280. https://doi.org/10.1111/caje.12754.

Frankel, Jeffrey, and Andrew Rose. 2002. "An Estimate of the Effect of Common Currencies on Trade and Income." *The Quarterly Journal of Economics* 117 (2): 437–466.

Glick, Reuven, and Andrew K. Rose. 2002. "Does a Currency Union Affect Trade? The Time-Series Evidence." *European Economic Review* 46 (6): 1125–1151. 10.1016/S0014-2921(01)00202-1.

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár. 2024. "Contamination Bias in Linear Regressions." *American Economic Review* 114 (12): 4015–51. 10.1257/aer.20221116.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* 225 (2): 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014, Themed Issue: Treatment Effect 1.

Gutmann, Jerg, Matthias Neuenkirch, and Florian Neumeier. 2023. "The economic effects of international sanctions: An event study." *Journal of Comparative Economics* 51 (4): 1214–1231. https://doi.org/10.1016/j.jce.2023.05.005.

Handley, Kyle, and Nuno Limão. 2017. "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States." *American Economic Review* 107 (9): 2731–83. 10.1257/aer.20141419.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *The Quarterly Journal of Economics* 123 (2): 441–487. 10.1162/qjec.2008.123.2.441.

Hinz, Julian, Amrei Stammann, and Joschka Wanner. 2021. "State Dependence and Unobserved Heterogeneity in the Extensive Margin of Trade." CEPA Discussion Papers 36, Center for Economic Policy Analysis. 10.25932/publishup-51191.

Hufbauer, G. C., J. J. Schott, K. A. Elliott, and B. Oegg. 2007. "Economic Sanctions Reconsidered." *(3rd edition). Washington, DC: Peterson Institute for International Economics.*

Hufbauer, Gary C., and Barbara Oegg. 2003. "The Impact of Economic Sanctions on US Trade: Andrew Rose's Gravity Model." *Peterson Institute for International Economics.*

Hull, Peter. 2018. "Estimating treatment effects in mover designs." `https://arxiv.org/abs/1804.06721`, arXiv:1804.06721.

Hummels, David L., and Georg Schaur. 2013. "Time as a Trade Barrier." *American Economic Review* 103 (7): 2935–59. 10.1257/aer.103.7.2935.

Koeze, Ella. 2022. "Boycotts, Not Bombs: Sanctions Are a Go-To Tactic, With Uneven Results." *New York Times* March 11.

Larch, Mario, José-Antonio Monteiro, Roberta Piermartini, and Yoto V. Yotov. 2025. "On the trade effects of GATT/WTO membership: They are positive and large after all." *Canadian Journal of Economics/Revue canadienne d'économique* 58 (1): 281–328. https://doi.org/10.1111/caje.12755.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2024. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* 68 (1): 160–176. https://doi.org/10.1111/ajps.12723.

Manova, Kalina. 2013. "Credit Constraints, Heterogeneous Firms, and International Trade." *The Review of Economic Studies* 80 (2): 711–744. 10.1093/restud/rds036.

Maoz, Zeev, Paul L. Johnson, Jasper Kaplan, Fiona Ogunkoya, and Aaron P. Shreve. 2019. "The Dyadic Militarized Interstate Disputes (MIDs) Dataset Version 3.0: Logic, Characteristics, and Comparisons to Alternative Datasets." *Journal of Conflict Resolution* 63 (3): 811–835. 10.1177/0022002718784158.

Martin, Julien, Isabelle Mejean, and Mathieu Parenti. 2023. "Relationship Stickiness, International Trade, and Economic Uncertainty." *The Review of Economics and Statistics* 1–45. 10.1162/rest_a_01396.

Mayer, Thierry, Gianluca Santoni, and Vincent Vicard. 2023. "The CEPII Trade and Production Database." Working Papers 2023-01, CEPII research center, `https://ideas.repec.org/p/cii/cepidt/2023-01.html`.

Mayer, Thierry, and Soledad Zignago. 2011. "Notes on CEPII's Distances Measures: The GeoDist Database." *CEPII Working Paper 2011 - 25.*

Melitz, Marc J.. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Monarch, Ryan. 2016. "'It's Not You, It's Me': Breakups in U.S.-China Trade Relationships." *Journal of International Economics* 102 242–261.

Moreau-Kastler, Ninon. 2025. "Proportional Treatment Effects in Staggered Settings: An Approach for Poisson Pseudo-Maximum Likelihood." Working Papers 031, EU Tax Observatory, https://ideas.repec.org/p/dbp/wpaper/031.html.

Morgan, T. Clifton, Navin Bapat, and Yoshiharu Kobayashi. 2014. "Threat and imposition of economic sanctions 1945-2005: Updating the TIES dataset." *Conflict Management and Peace Science* 31 (5): 541–558.

Morgan, T. Clifton, Constantinos Syropoulos, and Yoto V. Yotov. 2023. "Economic Sanctions: Evolution, Consequences, and Challenges." *Journal of Economic Perspectives* 37 (1): 3–30.

Mou, Hongyu, and Yiqing Xu. 2022. "PANELVIEW: Stata module to visualize panel data." Statistical Software Components, Boston College Department of Economics, January, https://ideas.repec.org/c/boc/bocode/s459034.html.

Nagengast, Arne J., and Yoto V. Yotov. 2025. "Staggered Difference-in-Differences in Gravity Settings: Revisiting the Effects of Trade Agreements." *American Economic Journal: Applied Economics* 17 (1): 271–96. 10.1257/app.20230089.

Rose, Andrew K. 2000. "One Money, One Market: The Effect of Common Currencies on Trade." *Economic Policy* 15 (30): 08–45. 10.1111/1468-0327.00056.

Roth, Jonathan. 2022. "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends." *American Economic Review: Insights* 4 (3): 305–22. 10.1257/aeri.20210236.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics* 235 (2): 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008.

Schmidt-Eisenlohr, Tim. 2013. "Towards a theory of trade finance." *Journal of International Economics* 91 (1): 96–112. https://doi.org/10.1016/j.jinteco.2013.04.005.

Shahn, Zach. 2023. "Subgroup difference in differences to identify effect modification without a control group." https://arxiv.org/abs/2306.11030, arXiv:2306.11030.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2): 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006, Themed Issue: Treatment Effect 1.

Syropoulos, Constantinos, Gabriel Felbermayr, Aleksandra Kirilakha, Erdal Yalcin, and Yoto V. Yotov. 2024. "The global sanctions data base–Release 3: COVID-19, Russia, and multilateral sanctions." *Review of International Economics* 32 (1): 12–48.

Tyazhelnikov, Vladimir, and John Romalis. 2024. "Russian counter-sanctions and smuggling: Forensics with structural gravity estimation." *Journal of International Economics* 152 104014. https://doi.org/10.1016/j.jinteco.2024.104014.

Volpe Martincus, Christian, Jerónimo Carballo, and Alejandro Graziano. 2015. "Customs." *Journal of International Economics* 96 (1): 119–137. https://doi.org/10.1016/j.jinteco.2015.01.011.

Wooldridge, Jeffrey M. 2023. "Simple Approaches to Nonlinear Difference-in-Differences with Panel Data." *The Econometrics Journal* 26 C31–C66.

Wooldridge, Jeffrey M. 2025. "Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators." *Empirical Economics* 1–43.

Yotov, Yoto V., Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. 2016. *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model.* Geneva, Switzerland: United Nations and World Trade Organization, available for download at `http://vi.unctad.org/tpa/index.html`.

Supplementary Appendix

A.1 Additional descriptive statistics

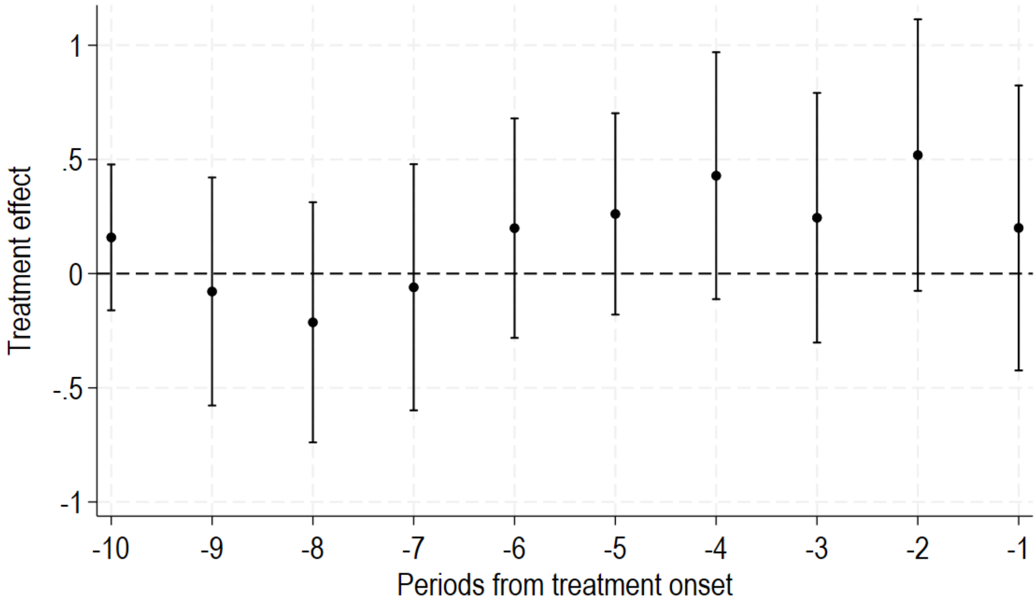
Table A1: Descriptive statistics on the number of observations by cohort

Cohort	Obs	Obs (ON)	Obs (OFF)	Years	Years (ON)	Years (OFF)	Pairs	Exporters	Importers
g = 1955; h = 1962	28	14	14	17	7	10	2	2	2
g = 1957; h = 1959	36	6	30	12	2	10	3	3	2
g = 1957; h = 1964	17	7	10	17	7	10	1	1	1
g = 1963; h = 1968	30	10	20	15	5	10	2	2	2
g = 1963; h = 1975	452	218	234	22	12	10	26	1	26
g = 1964; h = 1976	247	87	160	22	12	10	30	14	18
g = 1966; h = 1975	1,329	626	703	19	9	10	73	1	73
g = 1971; h = 1977	30	10	20	15	5	10	2	2	2
g = 1973; h = 1975	663	121	542	12	2	10	77	37	5
g = 1974; h = 1975	19	1	18	11	1	10	2	1	2
g = 1974; h = 1979	30	10	20	15	5	10	2	2	2
g = 1975; h = 1979	28	8	20	14	4	10	2	2	2
g = 1978; h = 1980	24	4	20	12	2	10	2	2	2
g = 1978; h = 1984	280	107	173	16	6	10	18	1	18
g = 1979; h = 1980	394	35	359	11	1	10	38	20	20
g = 1982; h = 1983	231	21	210	11	1	10	21	10	13
g = 1982; h = 1986	14	4	10	14	4	10	1	1	1
g = 1982; h = 1990	18	8	10	18	8	10	1	1	1
g = 1983; h = 1984	99	8	91	11	1	10	10	6	6
g = 1985; h = 1991	32	12	20	16	6	10	2	2	2
g = 1987; h = 1988	22	2	20	11	1	10	2	2	1
g = 1987; h = 1994	68	28	40	17	7	10	4	3	3
g = 1987; h = 1999	43	23	20	22	12	10	2	2	2
g = 1989; h = 1991	24	4	20	12	2	10	2	2	2
g = 1990; h = 1991	9	7	2	2	1	1	7	5	4
g = 1990; h = 1992	2,004	259	1,745	12	2	10	209	103	108
g = 1990; h = 2004	2,824	1,397	1,427	24	14	10	206	98	110
g = 1991; h = 1992	23	2	21	11	1	10	3	3	2
g = 1991; h = 1995	736	185	551	14	4	10	65	34	33
g = 1993; h = 1997	27	7	20	14	4	10	2	2	2
g = 1996; h = 2000	624	146	478	14	4	10	68	36	34
g = 1997; h = 2004	340	135	205	17	7	10	24	13	13
g = 2002; h = 2004	22	2	20	12	2	10	2	2	2
g = 2006; h = 2012	28	12	16	14	6	8	2	2	2
g = 2007; h = 2013	26	12	14	13	6	7	2	2	2
g = 2010; h = 2012	20	4	16	10	2	8	2	2	2
g = 2013; h = 2017	14	8	6	7	4	3	2	2	2
Treated	10,855	3,550	7,305	65	60	61	795	142	149
Never treated	1,115,728			70			46,495	260	260

Notes: This table provides descriptive statistics on the number of observations by cohort for the baseline sample that is used to obtain our main results in Table 2. The columns represent the cohort (g = entry year, h = exit year), total observations (Obs), observations while sanctions were imposed (Obs (ON)), observations after sanctions were lifted (Obs (OFF)), total years, years while sanctions were imposed (Years (ON)), years after sanctions were lifted (Years (OFF)), pairs, exporters, and importers. In the bottom section of the table, the data is divided into treated and never treated observations.

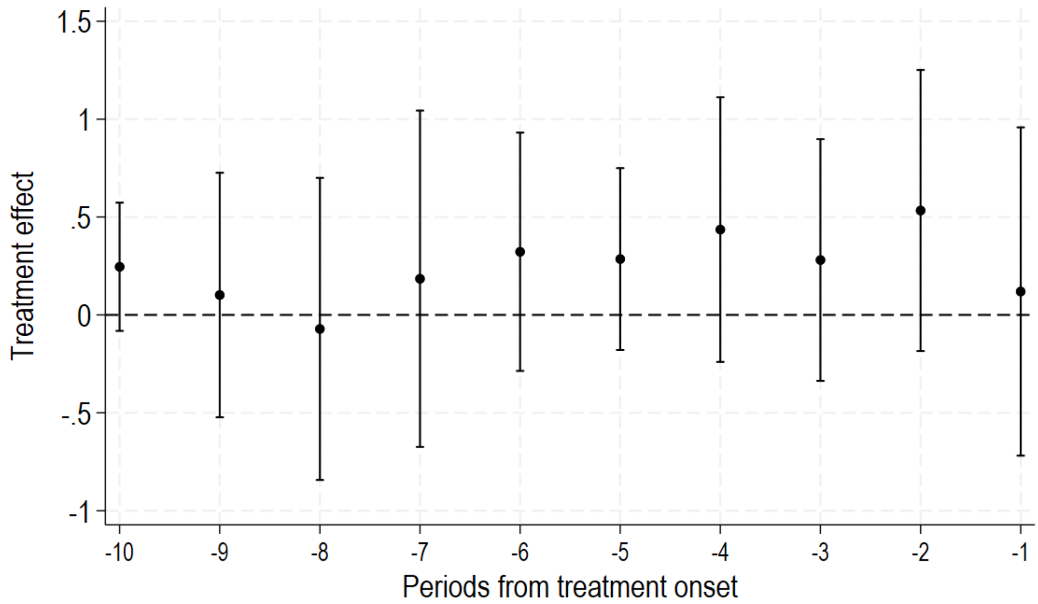
A.2 Alternative pre-treatment tests

Figure A1: Alternative pre-treatment test I



Notes: The figure reports pre-trend estimates from an estimation of equation (1), to which additional placebo treatment indicators were added in the years before the sanction was imposed following Wooldridge (2025). The regression is estimated with untreated observations only in the spirit of Borusyak et al. (2024). For visualization, the resulting cohort-event-time-specific treatment effects were aggregated by event time. The coefficients before sanction imposition are jointly insignificant (p-value of 0.195). 95% confidence intervals are shown using standard errors clustered by country pair.

Figure A2: Alternative pre-treatment test II

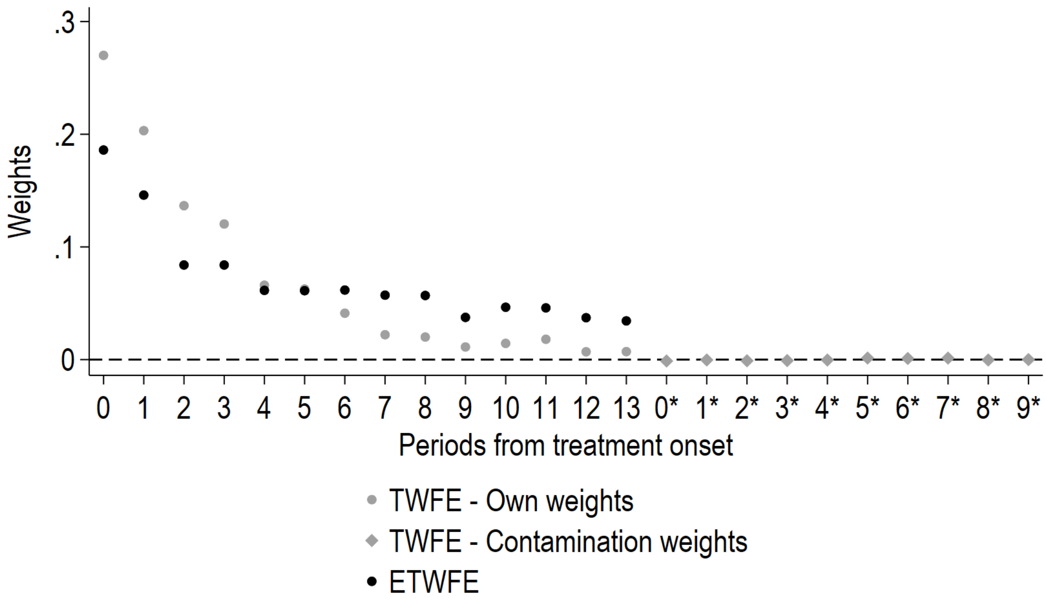


Notes: The figure reports pre-trend estimates in the spirit of Liu et al. (2024), obtained from an imputation-type estimator (Borusyak et al., 2024) using only never-treated and not-yet-treated observations. For treated country pairs, the 10 pre-treatment years are excluded when estimating exporter-year, importer-year, and pair fixed effects. Predicted trade flows are then imputed beginning 10 years prior to the treatment onset (sanction imposition). Treatment effects are the residual gaps between observed outcomes and the fixed-effects fit across event-time horizons. For visualization, the resulting cohort-year effects are aggregated by event time (with $t = 0$ at sanction imposition). 95% confidence intervals are shown using block-bootstrap standard errors with 999 repetitions.

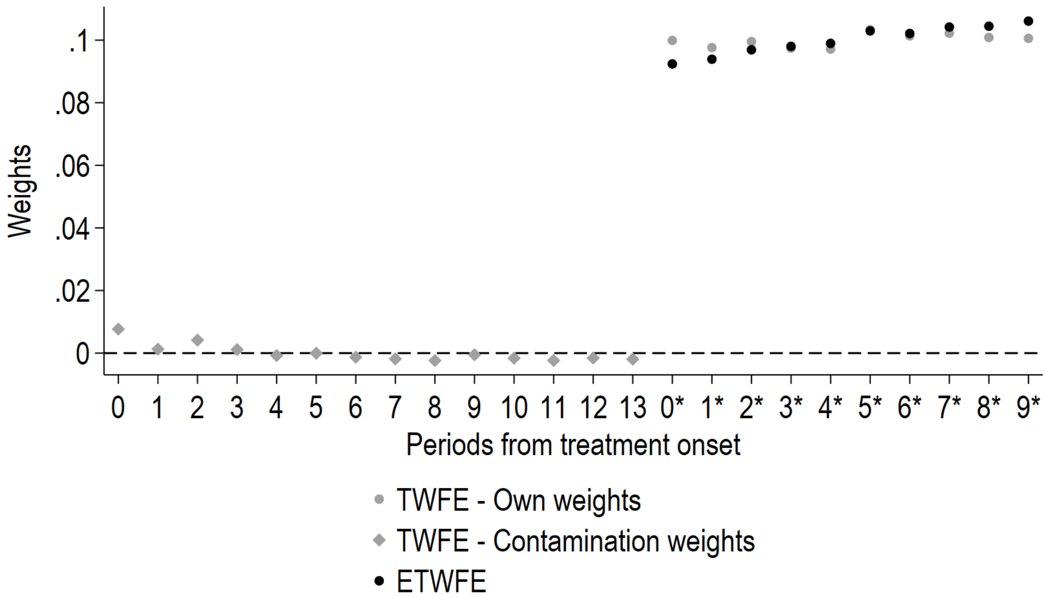
A.3 Implicit weights attached by the TWFE model

In this section, we consider a detailed analysis of the implicit weights attached by the TWFE model from the decomposition analysis described in Section 4.3 and we also provide a bias decomposition analysis of the dynamic TWFE model following Sun and Abraham (2021).

Figure A3: Weights of ETWFE and TWFE model - By event time



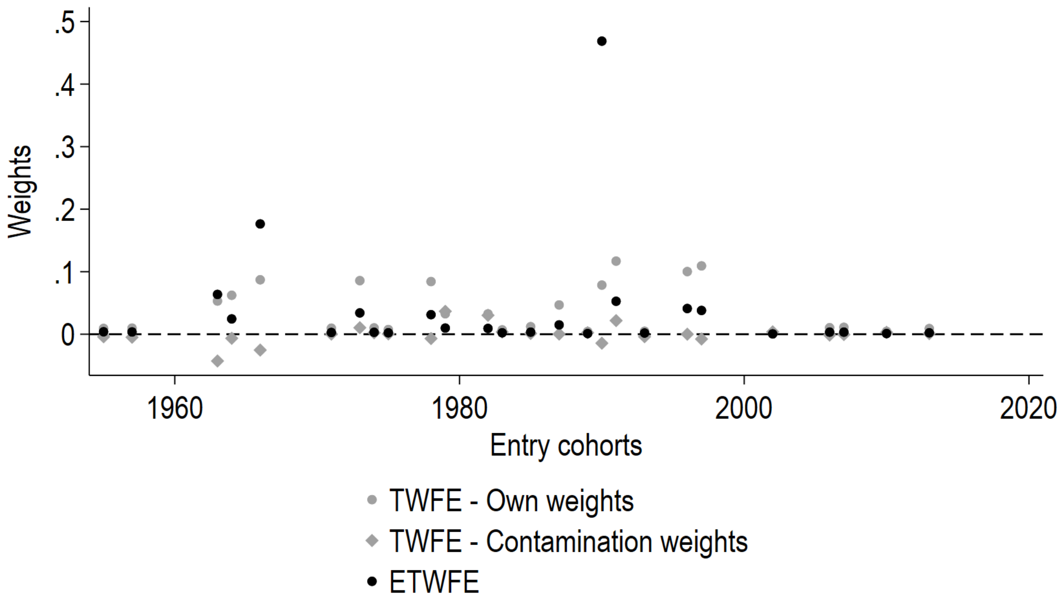
(a) ON effect



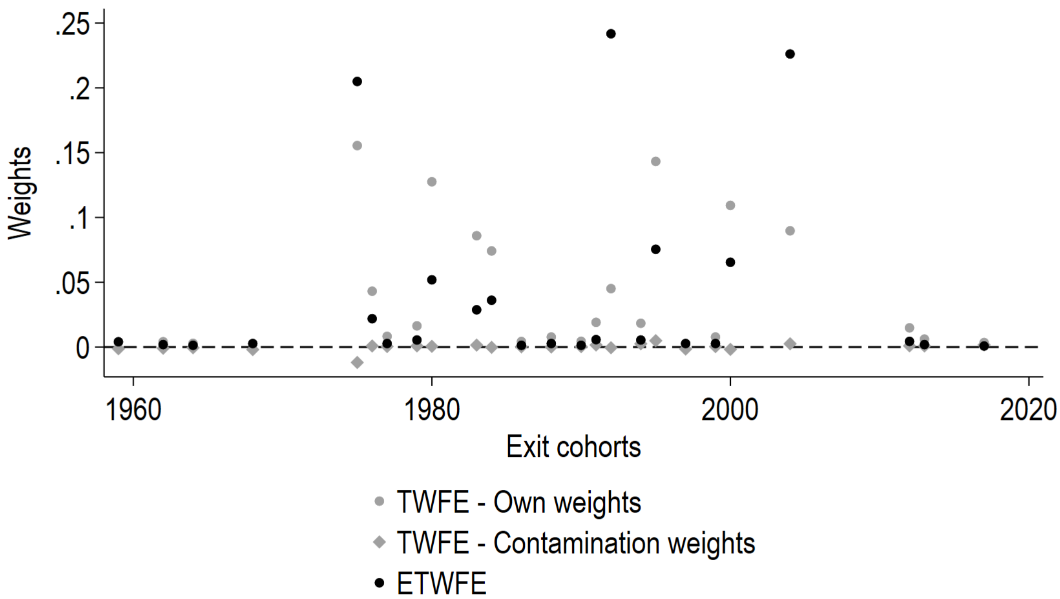
(b) OFF effect

Notes: The figure reports the weights used in the computation of the aggregate treatment effects of the ETWFE from column (1) of Table 2 in dark color ('ETWFE') along with the implicit weights attached by the (long) TWFE model from column (2) of Table 2 to cohort-event-time cells computed following de Chaisemartin and D'Haultfœuille (2023) in light color ('TWFE') split into weights on own treatment effects ('own weights') and weights on the effects of other treatments ('contamination weights'). Panel (a) reports weights aggregated by event time for the ON effect. Panel (b) reports weights aggregated by event time for the OFF effect.

Figure A4: Weights of ETWFE and TWFE model - By cohort



(a) By entry cohort (ON effect)



(b) By exit cohort (OFF effect)

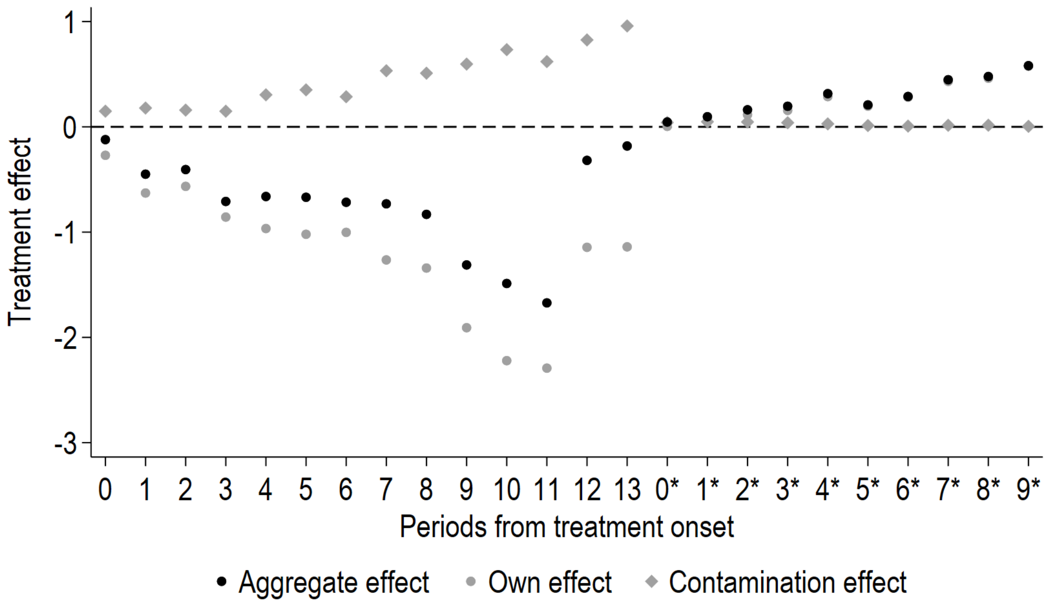
Notes: The figure reports the weights used in the computation of the aggregate treatment effects of the ETWFE from column (1) of Table 2 in dark color ('ETWFE') along with the implicit weights attached by the (long) TWFE model from column (2) of Table 2 to cohort-event-time cells computed following de Chaisemartin and D'Haultfœuille (2023) in light color ('TWFE') split into weights on own treatment effects ('own weights') and weights on the effects of other treatments ('contamination weights'). Panel (a) reports weights aggregated by entry cohorts the ON effect. Panel (b) reports weights aggregated by exit cohorts for the OFF effect.

The bias decomposition results of the TWFE model from column (2) of Table 2 are presented in Figure A3 by event time and in Figure A4 by cohort. We report the weights used in the computation of the aggregate treatment effects of the ETWFE from column (1) of Table 2 in dark color ('ETWFE') along with the implicit weights attached by the (long) TWFE model from column (2) of Table 2 to cohort-event-time cells computed following de Chaisemartin and D'Haultfœuille (2023) in light color ('TWFE') split into weights on own treatment effects ('own weights') and weights on the effects of other treatments ('contamination weights').

Figure A3 reports the resulting weights aggregated by event time. Panel (a) shows that the ON effect of the TWFE suffers from a strong short-term bias. In combination with the increasing effects of sanctions over time (in absolute terms), this leads to a downward bias of the ON effect. The contamination weights roughly sum to zero across event time highlighting that the contamination bias stems from the cohort dimension in this case. Panel (b) shows that the OFF effect of the TWFE is afflicted by a short-term bias albeit less so. Given the absence of increasing effects over time (and the opposite direction of the bias), the source of the own bias term stems from the cohort dimension in this case. As for the ON effect, the contamination weights sum to zero in event time.

For completeness, Figure A4 reports the corresponding weights aggregated by entry and exit cohorts. Given that the ON and OFF treatment effects do not show a clear pattern over time, here, we highlight only two results. First, for some cohorts, the TWFE own weights differ substantially from the corresponding ETWFE weights. Second, for the ON effect, some cohorts have large positive and negative contamination weights, while, for the OFF effect, the contamination weights aggregated by cohort are mostly close to zero.

Figure A5: Decomposition of dynamic TWFE results



Notes: The figure reports event-time-specific treatment effects from a dynamic OLS regression. The aggregate effects are decomposed into own effects, and contamination effects following Sun and Abraham (2021). Own effects are defined as the effects for the same event year, while contamination effects are defined as the effects from other event years. The black dots represent the aggregate effects, the gray dots represent the own effects, and the gray diamonds represent the contamination effects. Periods marked with a star (*) indicate post-sanction years.

We also perform a similar decomposition for the dynamic TWFE results into 'own' and 'contamination' effects in the spirit of Sun and Abraham (2021). In this case, the definition of the effects differs slightly, as they are defined relative to the same event year. More specifically, 'own' effects are defined as the effects for the same event year, while 'contamination' effects are defined as the effects from other event years (i.e., not only entry vs. exit as in the previous analysis). The decomposition results are presented in Figure A5. The following three results stand out: First, the contamination effects are positive in all periods while sanctions are imposed reducing the aggregate effect as for the main analysis. Second, the contamination effects increase over time. A possible interpretation is that the negative impact of sanctions accumulates over the years and, therefore, has a higher correlation with the other dynamic treatment effect terms as sanctions last longer. Third, after sanctions are lifted, contamination effects do not play a role and the aggregate effect is only driven by 'own' effects.

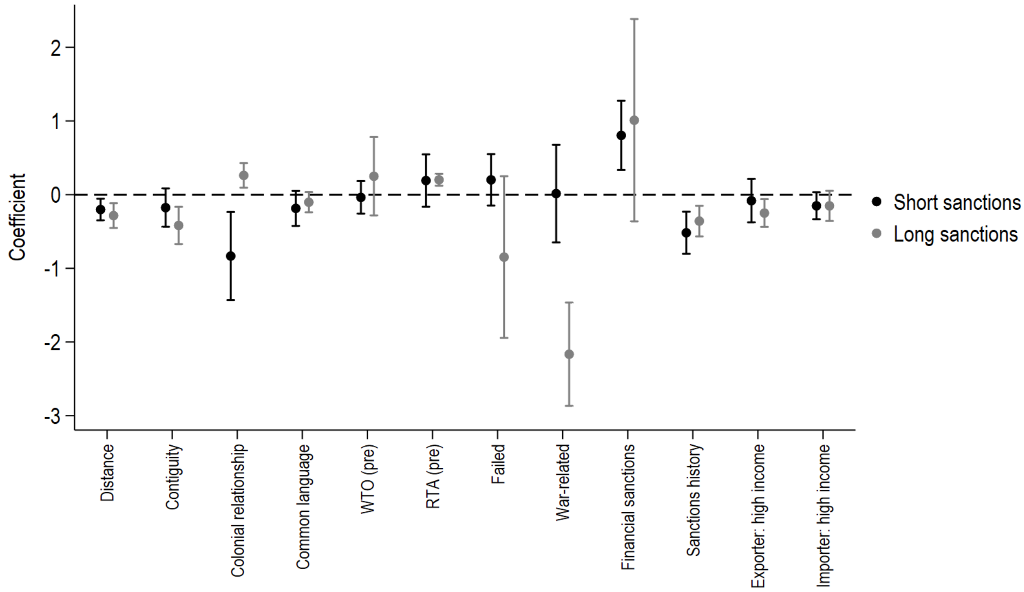
A.4 Additional subgroup-specific results

Table A2: Additional subgroup-specific results

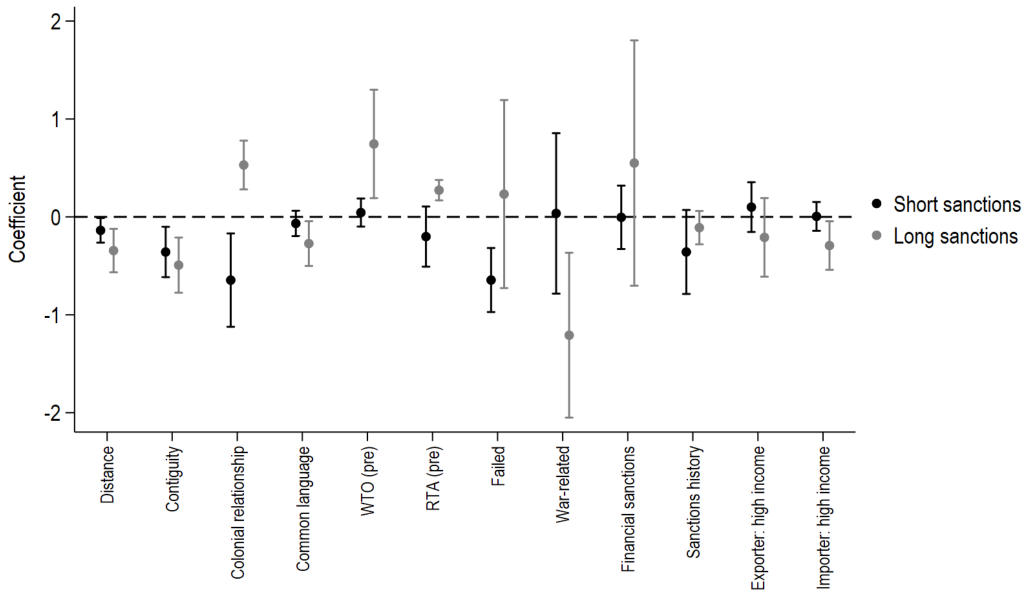
	(1)	(2)	(3)	(4)	(5)	(6)
ln Distance$_{ij}$	-0.198***	-0.214***	-0.204***	-0.187***	-0.174***	-0.159**
	(0.055)	(0.056)	(0.065)	(0.065)	(0.051)	(0.079)
Contiguity$_{ij}$	-0.419***	-0.268**	-0.427***	-0.552***	-0.484***	-0.551***
	(0.090)	(0.119)	(0.113)	(0.134)	(0.145)	(0.180)
Colonial relationship$_{ij}$	-0.166	-0.059	0.000	0.000	0.078	0.060
	(0.119)	(0.097)	(0.098)	(0.127)	(0.132)	(0.144)
Common language$_{ij}$	-0.008	-0.122	0.027	0.070	-0.011	0.019
	(0.091)	(0.089)	(0.077)	(0.105)	(0.082)	(0.080)
WTO$_{ij,g-1}$	0.424	0.326	0.250	0.405*	0.408*	0.422*
	(0.318)	(0.221)	(0.239)	(0.220)	(0.210)	(0.225)
RTA$_{ij,g-1}$	0.137	0.089	0.331***	0.058	-0.035	0.117
	(0.176)	(0.077)	(0.107)	(0.146)	(0.143)	(0.114)
Long duration	0.004	-0.023	0.043	-0.903**	-0.874**	-0.953*
	(0.253)	(0.233)	(0.285)	(0.424)	(0.403)	(0.489)
Outcome: Failed	-0.446	-0.734**	-0.800***	-0.314	-0.329	-0.758**
	(0.361)	(0.303)	(0.311)	(0.306)	(0.210)	(0.300)
Outcome: Partial success	0.192			-0.085		
	(0.563)			(0.385)		
Outcome: Negotiated settlement	0.550			-0.176		
	(0.370)			(0.400)		
Objective: End war		-0.516			-0.142	
		(0.491)			(0.418)	
Objective: Prevent war		-0.853			-0.973*	
		(0.695)			(0.555)	
Objective: Territorial Conflict		-2.079***			-0.789***	
		(0.281)			(0.168)	
Objective: Democracy			0.865**			0.207
			(0.338)			(0.381)
Objective: Destabilize regime			0.662			1.140
			(0.722)			(0.709)
Objective: Human rights			0.060			-0.210
			(0.349)			(0.295)
Objective: Policy change			0.955**			0.308
			(0.442)			(0.408)
Observations	3,505	3,505	3,505	7,186	7,186	7,186

Notes: The table presents OLS regression results of the estimated treatment effects $\hat{\delta}_{g,h,s-g}$ from the baseline ETWFE results of column (1) in Table 2 on a set of $s-g$ fixed effects to control for dynamics, using gravity variables from the CEPII's Gravity database (Mayer and Zignago, 2011), sanction characteristics from the Global Sanctions Database Syropoulos et al. (2024), and country characteristics from the World Bank. Standard errors in parentheses are obtained using a parametric bootstrap procedure with 999 repetitions. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Figure A6: Covariate heterogeneity in sanction effects: short vs. long-duration episodes



(a) During sanction imposition (ON sample)



(b) After sanction lifting (OFF sample)

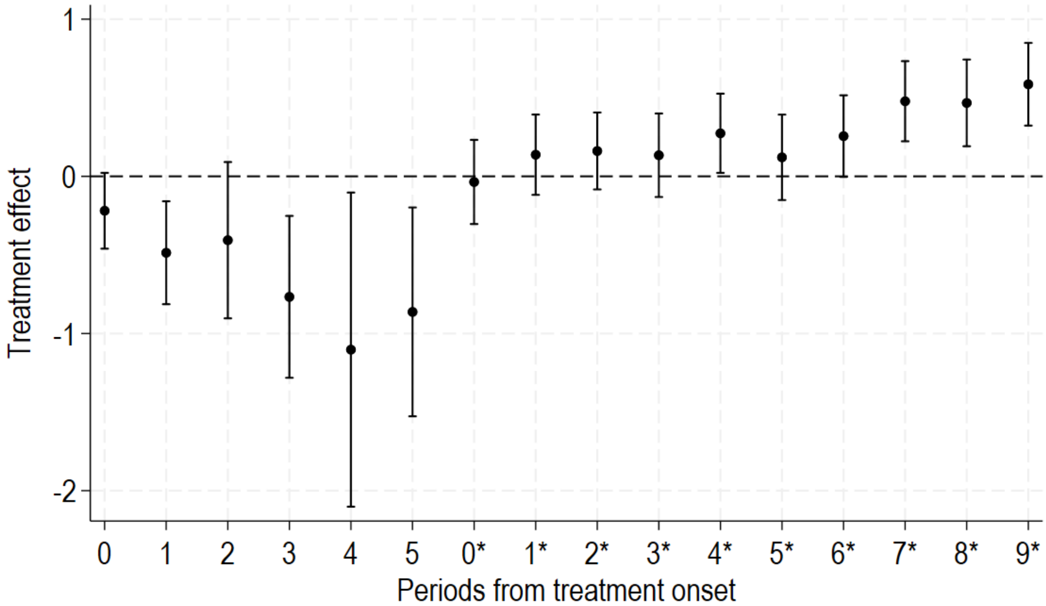
Notes: Points plot coefficients from OLS regressions of the cohort-event-time effects $\hat{\delta}_{gh,s-g}$ on dyad, episode, and country covariates, including event-time fixed effects. Black markers show *short sanctions* (baseline effects), gray markers show *long sanctions*, computed as the baseline + interaction with the long-duration indicator (episodes lasting between 7 and 14 years). The 95% confidence intervals are shown with standard errors obtained using a parametric bootstrap procedure with 999 repetitions. Estimation is performed separately for the ON (during imposition) and OFF (after lifting) samples. *Gravity covariates* from the CEPII's Gravity database (Mayer and Zignago, 2011): $\ln \text{Distance}_{ij}$, Contiguity_{ij}, Colonial relationship_{ij}, Common language_{ij}, $\text{WTO}_{ij,g-1}$, and $\text{RTA}_{ij,g-1}$. *Sanction characteristics* from the Global Sanctions Database (Felbermayr et al., 2020): Failed_{gh}, War-related_{gh}, Financial sanctions_{ijt} (in year t), and Sanctions history_{ij} (any prior non-trade sanctions in the dyad). *Income*: High income exporter_i and High income importer_j follow World Bank income groups.

A.4.1 Covariate heterogeneity in sanction effects: short vs. long-duration episodes

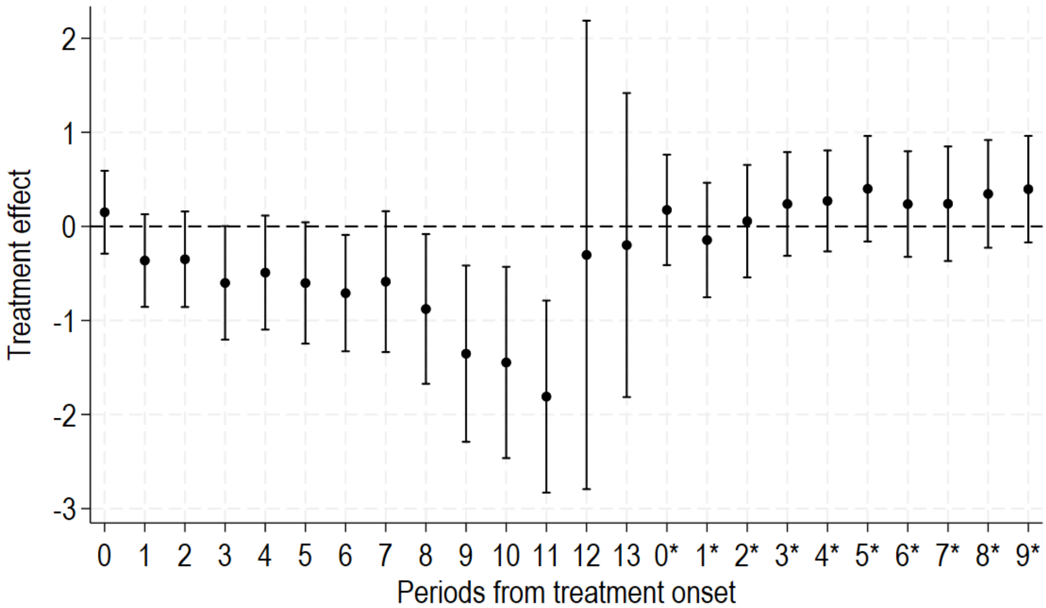
Figure A6 complements Table 4 by allowing the associations in columns (7) and (14) to vary with episode length. Starting from those specifications, we interact each covariate with an indicator for long duration (episodes lasting more than 6 years) and plot the implied coefficients for short and long episodes in the ON and OFF samples. This decomposition clarifies whether the average patterns in Table 4 are driven by long cases or also present for short ones. During imposition (ON sample), the negative associations for failed outcomes are much larger for long episodes. For short episodes, the corresponding coefficients are small and close to zero. Together with the negative uninteracted coefficients, this implies that the average ON penalties for failed sanctions are primarily concentrated in long sanctions, consistent with broader scopes and tighter enforcement when episodes persist. We also find a significantly more negative coefficient for colonial relationships for short sanctions. After lifting (OFF sample), rules-based ties matter mainly for long episodes. Pre-entry WTO co-membership and RTA status are associated with more positive post-lifting effects, whereas the corresponding short-episode coefficients are muted. We also find a significantly more positive coefficient on colonial relationships for long sanctions.

A.4.2 Horizon-specific effects by duration estimated using TWFE model

Figure A7: Horizon-specific effects by duration estimated using TWFE model



(a) Short sanctions



(b) Long sanctions

Notes: The figure reports event-time-specific treatment effects estimated using a (dynamic) TWFE model that allows for different effects for short and long sanctions. Years with a star (*) refer to post-sanction years. Short sanctions are sanctions with a duration of 6 years or less. Long sanctions are sanctions with a duration between 7 and 14 years. 95% confidence intervals are shown using standard errors clustered by country pair.

A.5 Details on additional results

This section provides implementation details and full results for the additional analyses summarized in the main text. We organize the material into five subsections covering (i) responses at the extensive margin, (ii) outcomes under other types of sanctions, (iii) sectoral heterogeneity, (iv) raw data evidence from major sanction cohorts, and (v) an application to currency unions.

A.5.1 Extensive margin

Extensive-margin adjustments, i.e., entry and exit of firms, products, and bilateral links, are central to trade dynamics (e.g., Helpman et al., 2008; Bernard et al., 2009). Although most visible at the firm and product level, in principle, these forces can also shape aggregate trade by altering the number of active country pairs (e.g., Hinz et al., 2021; Dutt et al., 2022). To gauge the relevance of the extensive margin in aggregate bilateral trade data, we examine year-to-year switches between zero and non-zero flows at the exporter-importer pair level. Specifically, we estimate a linear probability model using the same ETWFE specification as in the baseline and aggregate the cohort-year effects into 'ON' (sanctions in force) and 'OFF' (sanctions lifted) periods. The results are shown in columns (1) and (2) of Table A3, respectively. During sanctions, the probability that a previously positive bilateral flow falls to zero rises by 4.3 percentage points (SE 0.024), while the corresponding effect after sanctions are lifted is close to zero and statistically insignificant. Conversely, sanctions reduce the probability that a zero flow turns positive by 5.3 percentage points (SE 0.024), whereas lifting sanctions increases that entry probability by 4.9 percentage points (SE 0.018). Taken together, sanctions operate partly through entry-exit dynamics even in aggregate data. They raise the hazard that ongoing flows fall to zero and suppress the formation or resumption of links. When sanctions are lifted, entry rises above its pre-sanctions level, consistent with the resumption of trade relationships disrupted during sanctions, while the exit hazard is statistically indistinguishable from its pre-sanctions level. These results indicate that entry-exit dynamics materially contribute to the aggregate response to sanctions, consistent with evidence using more disaggregated data (e.g. Bista and Sheridan, 2025).

A.5.2 Other types of sanctions.

To determine whether our results are specific to complete trade sanctions, we experiment with other types of sanctions as alternative treatments. Specifically, we consider partial trade sanctions, financial sanctions, and a category encompassing arms sanctions, military sanctions, travel sanctions, and other types of sanctions. The results are shown in columns (3) to (8) of Table A3.

The estimated effects of partial trade sanctions are not statistically significant (column (3)), suggesting that these measures may not substantially impact bilateral trade flows. A possible explanation for this result is that partial trade sanctions usually target specific companies, and even if they are effective in punishing those companies, the impact on overall trade is not significant. Similarly, financial sanctions (column (4)), which restrict access to capital and financial services, do not significantly affect trade volumes, likely because they do not directly impede the physical exchange of goods and services. On a related note, such indirect effects are probably captured by country-year fixed effects in our econometric model. Other types of sanctions (column (5)) also fail to produce significant trade effects, possibly because they are aimed at achieving specific geopolitical or security objectives rather than disrupting overall economic activities.

Table A3: Extensive margin and other types of sanctions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$ON_{ij,t}$	0.043*	-0.053**	0.005	-0.048	-0.061	0.020	-0.050	-0.023
	(0.024)	(0.024)	(0.118)	(0.083)	(0.070)	(0.123)	(0.085)	(0.076)
$OFF_{ij,t}$	0.009	0.049***	0.126	0.045	-0.044	0.230	-0.018	-0.068
	(0.026)	(0.018)	(0.148)	(0.070)	(0.071)	(0.165)	(0.073)	(0.076)
$OFF_{ij,t} - ON_{ij,t}$	-0.034	0.102***	0.122	0.093	0.018	0.210	0.032	-0.045
	(0.024)	(0.021)	(0.138)	(0.077)	(0.064)	(0.154)	(0.076)	(0.068)
Extensive margin	Market entry	Market entry						
Treatment								
Complete trade sanctions	Yes	Yes						
Partial trade sanctions			Yes			Yes		
Financial sanctions				Yes			Yes	
Any other sanction					Yes			Yes
Sample								
Excl. complete trade sanctions						Yes	Yes	Yes
Observations	1,106,727	856,554	1,121,586	1,020,945	932,735	1,063,489	1,003,688	924,750
thereof: ON	3,544	3,409	12,528	21,962	26,760	11,529	20,696	24,890
thereof: OFF	7,050	3,676	14,553	31,748	41,816	12,902	27,552	37,073
Exporters	259	259	260	260	260	260	260	260
Importers	259	259	260	260	260	260	260	260
Years	69	69	70	70	70	70	70	70
Coefficients	570	329	1,226	2,036	2,647	1,155	1,936	2,471
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 regression results using ETWFE specifications (equation (1)), for which the cohort-event-time-specific treatment effects were aggregated using equations (4) and (5) to obtain aggregate ON and OFF treatment effect estimates, respectively. In column (1), the dependent variable is market exit, a dummy equal to one if a previously positive bilateral flow drops to zero, and zero otherwise. In column (2), the dependent variable is market entry, a dummy equal to one if a previously zero bilateral flow becomes positive, and zero otherwise. In columns (3)-(8), the dependent variable is the natural logarithm of exports. The treatment variable is an indicator of complete trade sanctions (columns (1) and (2)), partial trade sanctions (columns (3) and (6)), financial sanctions (columns (4) and (7)), or any other sanction (columns (5) and (8)). Columns (6)-(8) exclude all post-treatment observations involving complete trade sanctions (ON and OFF). '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% levels, respectively.

So far, we have estimated regressions identifying the effects of partial trade sanctions, financial sanctions, and other sanctions, without separately controlling for complete trade sanctions, which remain in the sample. This follows de Chaisemartin and D'Haultfœuille (2023), who show that omitting a second treatment can reduce bias when treatment effects are heterogeneous. To ensure that estimates for other sanctions are not confounded by complete trade sanctions, whether through their presence in the control group or as a concurrent treatment, we replicate the analysis from columns (3)-(5) in columns (6)-(8) of Table A3, excluding all post-treatment observations involving complete trade sanctions (ON and OFF). The results are unchanged: estimates remain near zero and statistically insignificant.

We therefore conclude that trade effects are driven specifically by complete trade sanctions, while other sanctions alone have no measurable impact.

A.5.3 Sectoral heterogeneity

We expand our heterogeneity analysis to examine how the effects of trade sanctions vary across sectors. Using UN Comtrade product-level data (available from 1963 onward), we classify bilateral trade flows into three broad sectors: Agriculture, Mining, and Manufacturing. In addition, we further distinguish two sub-sectors within Manufacturing, High-Tech Manufacturing and Dual-Use Manufacturing, as these are of particular relevance in the context of trade sanctions due to their strategic and security importance. The classification follows the HS 2-digit code groupings below:

- Agricultural: HS codes 01-24
- Mining: HS codes 25-27
- Manufacturing: HS codes 28-96
- High Tech Manufacturing: HS codes 28-31, 32-33, 72-76, 78-81, 84-92
- Dual-Use Manufacturing: HS codes 26-27, 28-31, 38, 72-76, 78-81, 84-85, 88-90

Table A4: Results by sector

	(1) Agri.	(2) Mining	(3) Manu.	(4) High-tech	(5) Dual-use
$ON_{ij,t}$	-0.473**	-0.161	-0.356**	-0.415***	-0.727**
	(0.189)	(0.201)	(0.178)	(0.160)	(0.319)
$OFF_{ij,t}$	-0.720***	0.462*	-0.157	-0.242	-0.288
	(0.271)	(0.279)	(0.230)	(0.210)	(0.322)
$OFF_{ij,t} - ON_{ij,t}$	-0.247	0.622**	0.199	0.172	0.439*
	(0.251)	(0.276)	(0.217)	(0.214)	(0.246)
Observations	764,457	542,664	945,299	878,803	877,598
thereof: ON	2,174	1,814	2,693	2,463	2,566
thereof: OFF	4,924	4,190	6,186	5,763	5,810
Exporters	245	245	245	245	245
Importers	245	245	245	245	245
Years	57	57	57	57	57
Coefficients	385	377	398	398	398
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes

Notes: The table presents regression results using ETWFE specifications, for which the cohort-event-time-specific treatment effects were aggregated using equation (4) and (5) to obtain aggregate ON and OFF treatment effect estimates, respectively. The dependent variable is the natural logarithm of exports, which varies over the exporter-importer-year dimension. Data are from UN Comtrade product-level data (available from 1963 onward) aggregated to sectoral level as described in the text. 'Agri.' denotes 'Agriculture', 'Manu.' denotes 'Manufacturing', 'High-tech' denotes 'High-tech manufacturing', and 'Dual-use' denotes 'Dual-use manufacturing'. '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 detailed regression results are reported in Table A4. A few key findings emerge. First, the ETWFE estimates reveal substantial heterogeneity across sectors when compared to our baseline aggregate ON effect of -0.859 (SE 0.332). In the agricultural sector, sanctions significantly reduce trade during sanction periods (-0.473, SE 0.189), and this negative effect persists even after sanctions are lifted (-0.720, SE 0.271), although the two estimates are not statistically different. For the mining sector, we find no significant effect (-0.161, SE 0.201), likely due to the smaller sample size and special exemptions often applied to energy-related products (e.g., Iraq's "oil-for-food" program in the 1990s). In the manufacturing sector, sanctions significantly reduce trade during the sanction period (-0.356, SE 0.178). The OFF effect is statistically insignificant (-0.157, SE 0.230) consistent with our aggregate baseline. As anticipated, the high-tech and dual-use manufacturing sectors exhibit strong negative effects during sanction periods (-0.415, SE 0.160 and -0.727, SE 0.319, respectively), reflecting their strategic sensitivity and the central role of technology controls in modern sanction regimes.

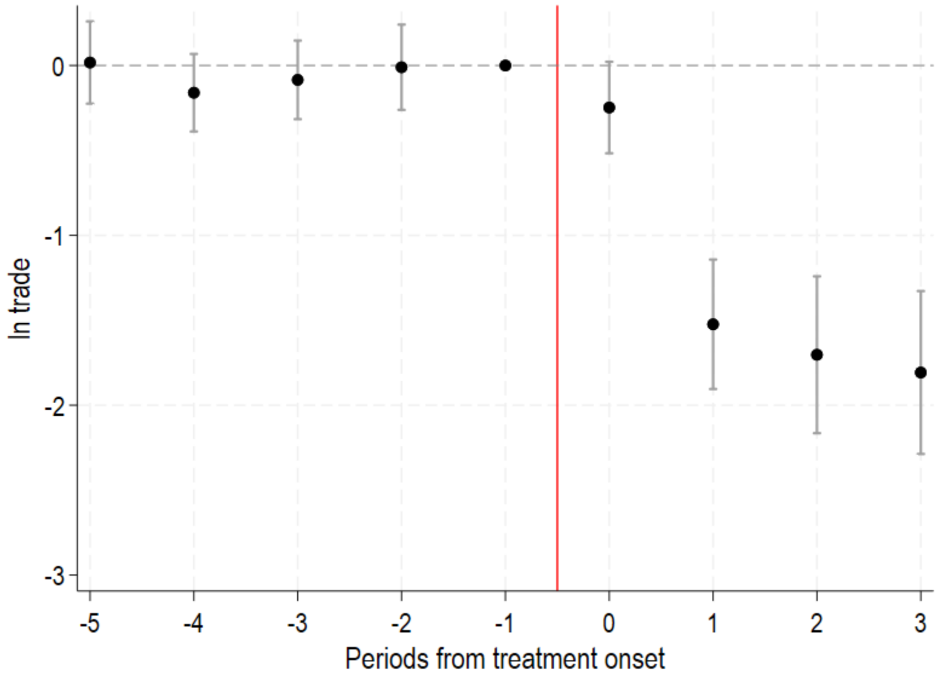
A.5.4 Raw data evidence from major sanction cohorts

To provide visual evidence of raw trade dynamics, we plot trade volume changes for all sanctions, as well as the largest sanction cohorts. Figure A8 presents the aggregate dynamics across all treated country pairs, while Figure A9 focuses on six major sanction episodes that correspond to the largest cohorts in our sample. As shown in Table A1, these six cohorts together account for more than 80% of all sanctioned pairs, making them highly representative of the overall pattern in the data. Each panel in Figure A9 plots the natural logarithm of bilateral trade flows between the sender and target countries over time, with the vertical line marking the imposition of sanctions. To enable comparability across episodes, trade flows are normalized relative to their pre-sanction averages in year $t-1$.

The figures reveal clear visual evidence of sanctions' effects in the raw data, with distinct declines in trade flows following sanction imposition. As shown in Figure A8, trade flows tend to decline slightly in the first year of sanctions and drop more sharply thereafter. The magnitude of the first-year decline may appear moderate because several sanctions were imposed mid-year, leading to annual averages that blend pre- and post-sanction trade. For example, the largest treated case (UN sanctions on Iraq) was imposed in August 1990, while UN sanctions on Portugal began in November 1966. Consequently, the annualized first-year impact is partially diluted.

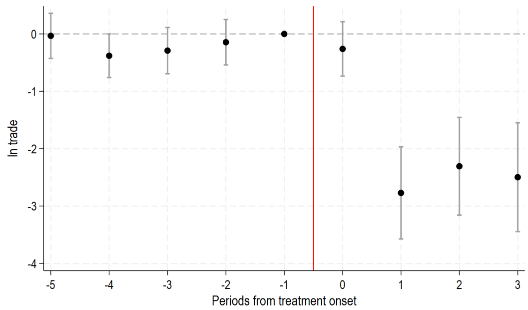
The figures also highlight considerable heterogeneity across sanction episodes. For instance, UN sanctions on Portugal (Figure A9b) did not lead to a pronounced decline in trade flows. These sanctions, adopted in 1966 through Resolution 2184 (XXI) in response to Portugal's colonial rule in Africa, urged member states to sever economic, diplomatic, and military ties until Portugal granted its colonies its independence. However, given the absence of specific enforcement mechanisms, compliance appears to have been partial,

Figure A8: Raw data evidence on entry effects across all sanction episodes

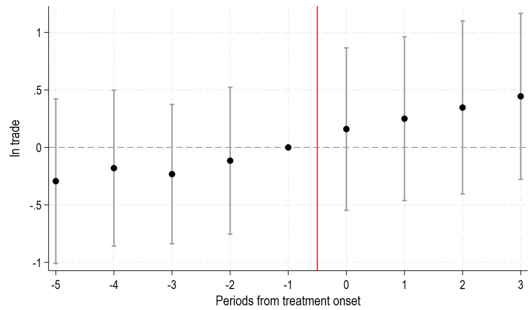


Notes: This figure plots the natural logarithm of bilateral trade flows for all sanction episodes in the baseline sample that is used to obtain our main results in Table 2. Point estimates represent the average at each event time relative to the pre-sanction period, with trade flows normalized by their pre-sanction average in year $t-1$. The vertical bars denote the 95% confidence intervals. The red vertical line marks the timing of the sanction imposition.

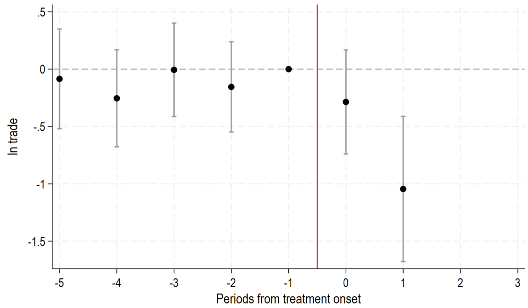
Figure A9: Raw data evidence on entry effects for major sanction cohorts



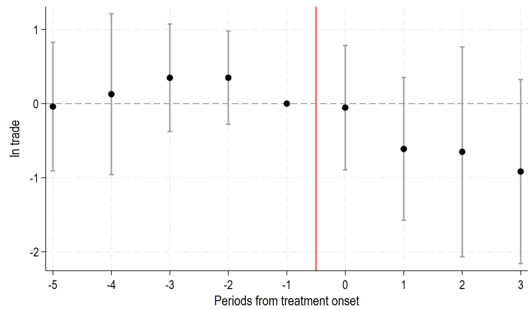
(a) UN-Iraq Sanctions (1990-2003)



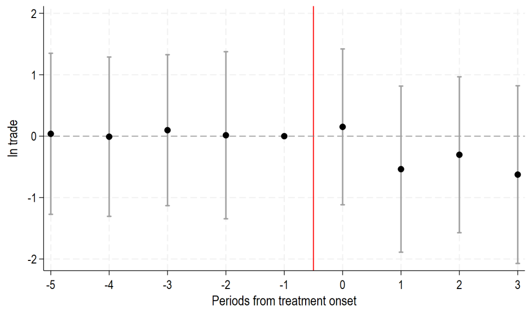
(b) UN-Portugal Sanctions (1966-1974)



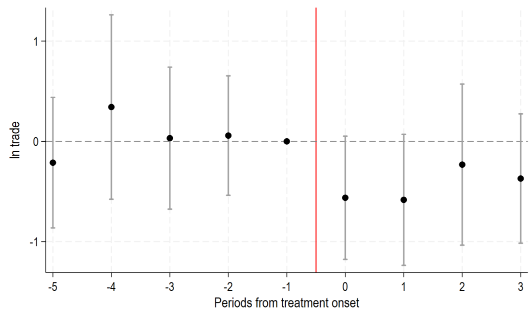
(c) UN-Kuwait, Venezuela-Suriname Sanctions (1990-1991)



(d) OAU-Portugal Sanctions (1963-1974)



(e) Canada, US, OAS-Haiti Sanctions (1991-1994)



(f) OAU-Burundi Sanctions (1996-1999)

Notes: Each panel shows the natural logarithm of bilateral trade flows for the corresponding sanctions episode in the baseline sample that is used to obtain our main results in Table 2. Point estimates represent the average at each event time relative to the pre-sanction period, with trade flows normalized by their pre-sanction average in year $t-1$. Vertical bars denote 95% confidence intervals. The red vertical line marks the timing of sanction imposition. UN denotes the United Nations sanctions, OAU denotes the Organization of African Unity, and OAS denotes the Organization of American States. Figures are labeled in the *sender-target* format.

leading to limited observable trade effects.

A brief summary of the historical context for each major sanction cohort is as follows:

- **UN-Iraq Sanctions (1990-2003):** Imposed following Iraq's invasion of Kuwait, these comprehensive sanctions were broadly enforced, leading to a dramatic reduction in Iraq's international trade.

- **UN-Portugal Sanctions (1966-1974):** Adopted in response to Portugal's colonial policies in Africa. Due to weak enforcement and limited participation, trade effects were relatively muted.

- **UN-Kuwait, Venezuela-Suriname Sanctions (1990-1991):** The UN sanctioned Iraq (and consequently trade with occupied Kuwait) after the invasion of Kuwait, as stated in Resolution 661: "all States shall prevent the import of commodities and products originating in Iraq or Kuwait." Separately, Venezuela imposed sanctions on Suriname following a military coup.

- **OAU-Portugal Sanctions (1963-1974):** Similar in motivation to the UN sanctions, these were declared by the Organization of African Unity to pressure Portugal to decolonize its African territories.

- **Canada, US, OAS-Haiti Sanctions (1991-1994):** These sanctions followed the military coup that overthrew President Jean-Bertrand Aristide and sought to restore democratic governance.

- **OAU-Burundi Sanctions (1996-1999):** Imposed by the Organization of African Unity after a coup in Burundi, aiming to promote a return to democratic rule.

A.5.5 Currency union application

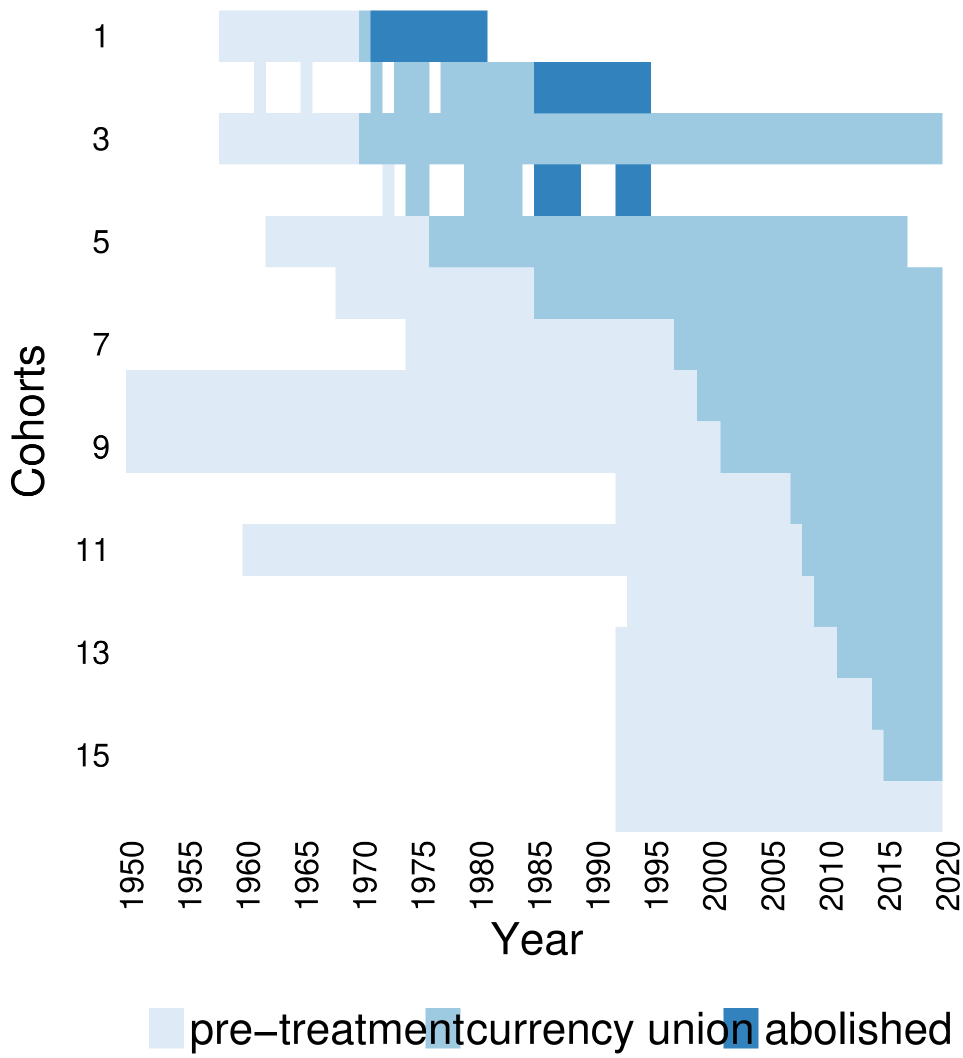
To demonstrate the broader applicability of our methodology, we apply it beyond sanctions to another trade policy-related setting. In particular, we examine the case of *currency unions*, which have been extensively studied in the trade literature (Rose, 2000;

Glick and Rose, 2002; Frankel and Rose, 2002; de Sousa, 2012). According to the definition in Frankel and Rose (2002), a currency union is considered dissolved once a currency ceases to be pegged at a 1:1 rate to the partner currency. Using this definition, we replicate Figure 2 from the main text for currency unions in Figure A10, which likewise displays a staggered treatment structure. We then re-estimate our baseline specification including currency union episodes, as shown in Table A5.

The results are qualitatively consistent with our main findings on sanctions. Specifically, the effect of currency union formation is positive and statistically significant (0.619, SE 0.067) under ETWFE, and larger than the corresponding estimates from both the 'long' TWFE (0.545, SE 0.059) and 'short' TWFE (0.540, SE 0.059) specifications. The dissolution of a currency union ('OFF' effect) is also positive and statistically significant in both the ETWFE model (0.708, SE 0.326) and the 'long' TWFE model (1.087, SE 0.348), providing suggestive evidence that the effects of currency unions may persist even after their dissolution, at least for the set of country pairs in our sample. Notably, the 'long' TWFE estimate of the OFF effect is subject to a sizable contamination component, underscoring the importance of using estimators that are robust to treatment-effect heterogeneity and, in particular, that correctly account for treatment exit.

Compared to the sanctions analysis, the currency union application features far fewer 'OFF' observations (62), which naturally limits the precision with which post-dissolution dynamics can be estimated. This caveat suggests that caution is warranted when interpreting the magnitude of OFF effects. A more detailed investigation of the post-dissolution adjustment process represents a promising avenue for future research.

Figure A10: Cohorts and treatment status of currency union episodes over time



Notes: This figure illustrates the treatment status of currency union formation and dissolution episodes over time. Pre-treatment years are shown in light blue, post-formation ("ON") years in medium blue, and post-dissolution ("OFF") years in dark blue. Years excluded from the estimation sample such as those with missing or zero trade flows are shown in white. To limit the influence of unrelated events, the estimation sample is restricted to 10 years after dissolution for treated pairs. For untreated pairs, all available observations are included. The figure is generated using the Stata module `panelview`.

Table A5: Regression results for currency union episodes

	(1) ETWFE	(2) TWFE	(3) TWFE
$ON_{ij,t}$	0.619***	0.545***	0.540***
	(0.067)	(0.059)	(0.059)
$OFF_{ij,t}$	0.708**	1.087***	
	(0.326)	(0.348)	
$OFF_{ij,t} - ON_{ij,t}$	0.089	0.542	
	(0.332)	(0.348)	
Observations	1,089,256	1,089,256	1,089,256
thereof: ON	6,199	6,199	6,199
thereof: OFF	62	62	62
Exporters	260	260	260
Importers	260	260	260
Years	70	70	70
Coefficients	292	2	1
Exporter × importer FE	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes
ON (own)		0.542***	0.540***
		(0.058)	(0.058)
ON (contam.)		0.003**	0.000
		(0.001)	(0.001)
OFF (own)		0.545*	
		(0.303)	
OFF (contam.)		0.542***	
		(0.136)	

Notes: The table presents regression results using an ETWFE specification (equation (1)) in column (1), a 'long' TWFE specification (equation (2)) in column (2), and a 'short' TWFE specification (equation (3)) in column (3) in the terminology of de Chaisemartin and D'Haultfœuille (2023). For the ETWFE specification, the cohort-event-time-specific treatment effects were aggregated using equation (4) and (5) to obtain aggregate ON and OFF treatment effect estimates, respectively. The dependent variable is the natural logarithm of exports, which varies over the exporter-importer-year dimension. '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.

A.6 Details on robustness analysis

This section provides implementation details and full results for the robustness checks summarized in the main text. We organize the material into five subsections covering (i) different samples, (ii) the role of small trade flows, (iii) degree of model heterogeneity, (iv) inclusion of domestic trade flows, and (v) alternative clustering approaches.

A.6.1 Different samples

With regard to different samples, we consider both variants that have larger coverage as well as variants that are more restricted. Our findings are reported in columns (1)-(6) of Table A6.

All sanction durations. In the baseline, we focus on sanctions lasting 14 years or less to maintain a more manageable scope for estimation. Therefore, we first use a sample that includes all sanction durations in columns (1) and (2). This increases the number of ON observations from 3,550 in the baseline to 3,778 and the number of OFF observations from 7,305 to 7,378. At the same time, the number of parameters to be estimated for the ETWFE regression increases noticeably from 526 in the baseline to 694.[A1] Most importantly for our purposes, the ETWFE estimates in column (1) are similar to our baseline estimate. Additionally, in column (2), the estimated TWFE ON coefficient is smaller in absolute value, and the OFF coefficient is larger than the corresponding ETWFE estimates, with the conclusions regarding the source of the bias remaining unchanged.

All post-exit observations. In our baseline analysis, we focused on 10 years of post-exit data to ensure a more targeted estimation and minimize the potential impact of other events on our estimates. Next, we expand the sample to include all post-exit observations, as presented in columns (3) and (4). This adjustment substantially increases the number

[A1] While estimating a large number of parameters in a linear model is not problematic in itself, we report the number of cohort-event-time coefficients to highlight the information-per-parameter trade-off in the ETWFE design. Unlike nuisance fixed effects, these coefficients are parameters of interest, and their precision depends on the number of observations underlying each estimate.

of OFF observations from 7,305 to 23,767. At the same time, the number of parameters to be estimated for the ETWFE regression nearly triples, rising from 526 in the baseline to 1,341. Crucially, the ETWFE estimates in column (3) remain consistent with our baseline results. Furthermore, in column (4), the estimated TWFE ON coefficient is reduced in absolute value, whereas the OFF coefficient is larger than the corresponding ETWFE estimates, with the conclusions regarding the source of bias remaining unchanged.

Requiring a minimum number of observations per treated cohort-year-cell. Next, to increase the precision of the cohort-event-time-specific estimates of ETWFE, we omit treated cohort-event-time cells with less than five observations(columns (5) and (6)). This only slightly reduces the number of treated ON (OFF) observations from 3,550 to 3,359 (from 7,305 to 6,898). At the same time, the number of coefficients to be estimated declines substantially to less than half (220) in comparison to the ETWFE baseline (526). However, the resulting aggregate ETWFE estimates in column (5) remain essentially unchanged. The corresponding TWFE ON coefficient appear even smaller in absolute terms and the OFF coefficient slightly larger than in the baseline. Importantly, the contamination bias term of the ON coefficient remains more or less unchanged and statistically significant.

Table A6: Robustness with regard to different samples and small trade flows

	(1) ETWFE	(2) TWFE	(3) ETWFE	(4) TWFE	(5) ETWFE	(6) TWFE	(7) ETWFE	(8) TWFE	(9) ETWFE	(10) TWFE
$ON_{ij,t}$	-0.949***	-0.711***	-0.763***	-0.482***	-0.897*	-0.190	-0.894***	-0.499***	-1.115***	-2.284***
	(0.313)	(0.141)	(0.294)	(0.121)	(0.484)	(0.117)	(0.339)	(0.121)	(0.295)	(0.424)
$OFF_{ij,t}$	0.065	0.241***	0.166	0.363***	0.094	0.438***	0.095	0.298***	0.029	-0.277**
	(0.218)	(0.093)	(0.158)	(0.096)	(0.253)	(0.103)	(0.229)	(0.096)	(0.161)	(0.126)
$OFF_{ij,t} - ON_{ij,t}$	1.014***	0.953***	0.929***	0.845***	0.990***	0.627***	0.989***	0.798***	1.144***	2.008***
	(0.256)	(0.120)	(0.286)	(0.112)	(0.361)	(0.106)	(0.279)	(0.108)	(0.233)	(0.396)
Sample	all durations		all post-exit periods		N(ghs)>=5		200 largest exporters		baseline	
Weighted Least Squares									Yes	Yes
Observations	1,141,641	1,141,641	1,157,530	1,157,530	1,139,438	1,139,438	1,012,166	1,012,166	1,141,068	1,141,068
thereof: ON	3,778	3,778	3,550	3,550	3,359	3,359	3,517	3,517	3,550	3,550
thereof: OFF	7,378	7,378	23,767	23,767	6,898	6,898	7,133	7,133	7,305	7,305
Exporters	260	260	260	260	260	260	200	200	260	260
Importers	260	260	260	260	260	260	200	200	260	260
Years	70	70	70	70	70	70	70	70	70	70
Coefficients	694	2	1,341	2	220	2	501	2	526	2
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ON (own)		-0.810***		-0.552***		-0.308**		-0.600***		
		(0.109)		(0.101)		(0.122)		(0.106)		
ON (contam.)		0.099***		0.070*		0.118***		0.100***		
		(0.029)		(0.039)		(0.041)		(0.031)		
OFF (own)		0.263***		0.359***		0.412***		0.291***		
		(0.086)		(0.087)		(0.101)		(0.090)		
OFF (contam.)		-0.022***		0.005		0.025**		0.007		
		(0.008)		(0.005)		(0.010)		(0.007)		

Notes: The table presents regression results using an ETWFE specification (equation (1)) and a 'long' TWFE specification (equation (2)) in the terminology of de Chaisemartin and D'Haultfœuille (2023). For the ETWFE specification, the cohort-event-time-specific treatment effects were aggregated using equation (4) and (5) to obtain aggregate ON and OFF treatment effect estimates, respectively. The dependent variable is the natural logarithm of exports, which varies over the exporter-importer-year dimension. The 'all durations' sample (columns (1) and (2)) includes sanctions of all durations and not just those of up to 14 years as in the baseline in Table 2. The 'all post-exit periods' sample (columns (3) and (4)) includes all post-exit observations and not just those of up to 10 years after sanctions were lifted as in the baseline in Table 2. The 'N(ghs)>=5' sample (columns (5) to (6)) contains only treated cohort-event-time cells with at least five observations (as well as all control observations). The '200 largest exporters' sample (columns (7) and (8)) contains only observations from the 200 largest exporters. The 'baseline' sample (columns (9) to (10)) corresponds to the sample in Table 2. Columns (9)-(10) were estimated using weighted least squares with time-invariant weights based on initial bilateral trade flows. 'Coefficients' reports the number of estimated coefficients apart from the fixed effects. The bias decomposition into own and contamination bias terms follows the decomposition proposed by de Chaisemartin and D'Haultfœuille (2023). Standard errors in parentheses are clustered by country pair. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

A.6.2 The role of small trade flows

Next, to assess whether our results are driven by small trade flows, we conduct two tests: restricting the sample to the 200 largest exporters and implementing weighted least squares using time-invariant weights based on initial bilateral trade flows. Our findings are reported in columns (7)-(10) of Table A6.

Sample limited to 200 largest exporters. In columns (7) and (8), we restrict the sample to include only the 200 largest exporters to limit the impact of small trade flows on our estimates. This reduces the overall number of observations by approximately 130,000. At the same time, the composition of the treatment group, and hence the composition of sanction episodes, remains similar to the baseline. Specifically, the number of ON (OFF) observations only decreases slightly by 33 (172), and the number of parameters to be estimated decreases by 25. However, the ETWFE estimates in column (7), the TWFE coefficients in column (8), and the conclusions regarding the source of bias remain essentially unchanged.

Weighted Least Squares. Next, as an additional robustness check, we implement a weighted least squares version of the baseline specification using time-invariant weights based on initial bilateral trade flows. Because our panel is unbalanced and some country pairs first appear after 1950, we construct weights as follows. For pairs observed in 1950, the weight is their 1950 trade flow, $w_{ij} = X_{ij,1950}$. For pairs first observed in year $s > 1950$, we back-cast an implied 1950 flow using global trade growth rates computed on the balanced set of pairs present in both adjacent years:

$$\hat{X}_{ij,1950} = X_{ij,s} \Big/ \prod_{\tau=1951}^{s} \big(1 + g_\tau\big),$$

where g_τ is the world trade growth rate between $\tau - 1$ and τ calculated over country pairs available in both years. Re-estimating our baseline ETWFE model with these weights (column (9) of Table A6) produces an ON estimate that is only slightly more negative than the baseline specification (-1.115, SE 0.295) and an OFF estimate that remains near zero and is statistically insignificant (0.029, SE 0.161). In contrast, the weighted TWFE model (column (10) of Table A6) delivers a substantially more negative ON estimate (-2.284, SE 0.424) and an OFF estimate that turns significantly negative (-0.277, SE 0.126). We interpret these results as confirming that our ETWFE estimates are robust to economically meaningful re-weighting that down-weights very small flows, whereas the TWFE estimates are sensitive to weighting and differ from the baseline.

A.6.3 Degree of heterogeneity of the model

The ETWFE specification in equation (1) potentially requires estimating a large number of cohort-event-time-specific coefficients. However, full treatment effect heterogeneity along the cohort and year dimension might not be necessary. Therefore, in this section, we impose restrictions on the treatment effect heterogeneity in the estimation, i.e., the coefficients $\delta_{gh,s-g}$ in equation (1), to explore the drivers behind the different estimates of the ETWFE and TWFE regressions. Our findings are reported in Table A7.

Only time heterogeneity. First, we allow the treatment effect to vary only across event time (column (1)). This is a very strong restriction, for which the number of coefficients to be estimated drops substantially from 526 in the baseline to just 24. Note that this specification is akin to a standard event-study or dynamic TWFE regression without including leads of the intervention (cf. Figure 4). As expected from the corresponding results in Figure 4, the resulting ON and OFF estimates are very close to the baseline TWFE coefficients. The implication is that event-time heterogeneity alone is not sufficient to account for the underlying treatment effect heterogeneity, which highlights the importance of cohort heterogeneity in this setting.

Only cohort heterogeneity. Therefore, next, we impose another strong restriction on the model by allowing the treatment effect to vary only across cohorts (column (2)), i.e.,

Table A7: Robustness with regard to degree of heterogeneity

	(1)	(2)	(3)	(4)	(5)
$ON_{ij,t}$	-0.610***	-0.663**	-0.700**	-0.840**	-0.804**
	(0.156)	(0.322)	(0.333)	(0.331)	(0.341)
$OFF_{ij,t}$	0.288***	0.165	0.087	0.071	0.076
	(0.094)	(0.201)	(0.223)	(0.222)	(0.222)
$OFF_{ij,t} - ON_{ij,t}$	0.898***	0.828***	0.787***	0.911***	0.880***
	(0.139)	(0.277)	(0.261)	(0.269)	(0.281)
Unit heterogeneity		Cohort	Cohort	Cohort	Cohort
Time heterogeneity	Year		5yr	3yr	2yr
Observations	1,141,068	1,141,068	1,141,068	1,141,068	1,141,068
thereof: ON	3,550	3,550	3,550	3,550	3,550
thereof: OFF	7,305	7,305	7,305	7,305	7,305
Exporters	260	260	260	260	260
Importers	260	260	260	260	260
Years	70	70	70	70	70
Coefficients	24	74	160	216	272
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes

Notes: The table presents regression results using variants of an ETWFE specification (equation (1)), in which the cohort-event-time-specific treatment effects were aggregated using equations (4) and (5) to obtain aggregate ON and OFF treatment effect estimates, respectively. The dependent variable is the natural logarithm of exports, which varies across the exporter-importer-year dimension. Column (1) imposes complete homogeneity along the cohort dimension, allowing only event-time-specific treatment effects. Column (2) imposes complete homogeneity along the time dimension, allowing only cohort-specific treatment effects. Columns (3), (4), and (5) restrict the cohort-time-specific treatment effects to change every five years, every three years, or every two years, respectively. Standard errors, in parentheses, are clustered by country pair. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

we allow each sanction cohort to have one separate coefficient for the period during which the sanction was imposed, and another coefficient for the period after the sanction was lifted. In this case, the number of estimated coefficients only decreases to 74. The resulting ON estimate (ETWFE) is slightly larger in absolute value than the corresponding TWFE baseline coefficient, but still smaller in absolute value than the ETWFE baseline estimate. Similarly, the resulting OFF estimate (ETWFE) is slightly smaller than the corresponding TWFE baseline coefficient and slightly larger than the ETWFE baseline estimate. Note that the OFF estimate loses its statistical significance in this case. We conclude that cohort heterogeneity plays a very important role in this setting, but the interaction with time heterogeneity is also important.

Cohort heterogeneity with restricted time heterogeneity. Therefore, we next consider a model with full cohort heterogeneity, but with limited time heterogeneity.

Specifically, for each cohort, we allow treatment effect heterogeneity to vary every five years (column (3)), every three years (column (4)), or every two years (column (5)). The number of coefficients to be estimated ranges between 160 (column (3)) and 272 (column (5)). The resulting ETWFE ON estimates generally become larger in absolute value as the time heterogeneity increases, though even in the most flexible model in column (5) the ON estimate remains slightly below the baseline estimate. In contrast, the OFF estimate converges more rapidly to the baseline estimate as time heterogeneity increases.

Overall, the estimates presented in Table A7 underscore the pronounced heterogeneity in treatment effects across different cohorts. At the same time, the heterogeneity of treatment effects across event time appears to have a smaller impact from this perspective. In summary, our findings suggest that the heterogeneity of the ETWFE model may be somewhat limited along the time dimension, with minimal repercussions for the aggregate treatment effects. However, caution is advised when making such adjustments.

A.6.4 Including domestic trade

The inclusion of domestic trade flows is consistent with standard gravity theory and can be beneficial for identifying certain policy effects. However, our baseline model excludes these flows. This choice is driven by data limitations: our primary objective is to cover the longest possible time span, thereby maximizing the number of sanction entries and exits, which is best achieved using an international-only trade dataset that offers greater temporal coverage.

Nevertheless, we also estimate our models using a dataset that includes domestic trade flows. Specifically, we incorporate domestic trade data from the CEPII Trade-Prod database (Mayer et al., 2023) available from 1966 onward. We proceed in two steps. First, we re-estimate our baseline ETWFE and TWFE models on the 1966-2019 sample without domestic trade flows, providing a clean reference point. Second, we augment the dataset with domestic trade flows and re-estimate the same models, now including cross-border×year fixed effects to absorb common globalization-related shocks (Bergstrand et al., 2015), which otherwise risk biasing sanction estimates when internal and external trade move in different directions over time. Columns (1)-(4) of Table A8 report these results.

First, for the 1966-2019 sample without domestic trade flows (columns (1)-(2)), we obtain patterns very similar to our main baseline, if anything, slightly more pronounced. For the ETWFE model (column (1)), the ON effect is -1.028 (SE 0.447), while the corresponding TWFE estimate (column (2)) is smaller in magnitude at -0.451 (SE 0.132). This mirrors the attenuation pattern observed at the baseline. For the OFF effect, ETWFE yields an imprecise estimate (0.062, SE 0.256), whereas the TWFE estimate is positive and highly significant (0.349, SE 0.101), again consistent with our baseline results.

Second, once domestic trade flows are included and we control for globalization effects with cross-border×year fixed effects (columns (3)-(4)), the magnitude of the ON estimates becomes larger in both models. The ETWFE estimate increases in absolute value to -1.891 (SE 0.495), while the TWFE estimate increases in absolute value to -0.701 (SE 0.228). Including domestic trade flows introduces internal trade as an additional, always-untreated comparison group. When sanctions are imposed, sanctioned countries divert part of their trade toward the domestic market, causing internal flows to rise while external flows fall. In the international-only dataset, this inward diversion is unobserved. Once domestic flows are included, however, the estimator compares declining sanctioned international flows to rising domestic flows, which widens the treated-control difference and makes the estimated ON effect more negative. For the OFF effect, the ETWFE estimate remains statistically insignificant (-0.291, SE 0.204), while the TWFE estimate remains positive and statistically significant (0.219, SE 0.119).

Overall, the inclusion of domestic trade flows does not alter our main conclusions. The qualitative patterns of ETWFE and TWFE estimates remain unchanged.

Table A8: Robustness with regard domestic trade flows and alternative clustering

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$ON_{ij,t}$	-1.028**	-0.451***	-1.891***	-0.701***	-0.859*	-0.859***	-0.859*
	(0.447)	(0.132)	(0.495)	(0.228)	(0.457)	(0.155)	(0.450)
$OFF_{ij,t}$	0.062	0.349***	-0.291	0.219*	0.071	0.071	0.071
	(0.256)	(0.101)	(0.204)	(0.119)	(0.171)	(0.119)	(0.167)
$OFF_{ij,t} - ON_{ij,t}$	1.090***	0.800***	1.601***	0.919***	0.930*	0.930***	0.930*
	(0.364)	(0.116)	(0.449)	(0.172)	(0.511)	(0.142)	(0.505)
Sample	1966-2019	1966-2019	1966-2019	1966-2019	baseline	baseline	baseline
Incl. domestic trade			Yes	Yes			
Model	ETWFE	TWFE	ETWFE	TWFE	ETWFE	ETWFE	ETWFE
Standard error clustering							
Exp × Imp	Yes	Yes	Yes	Yes			
Exp, Imp					Yes		
Exp × Year, Imp × Year						Yes	
Exp, Imp, Year							Yes
Observations	1,079,433	1,079,433	1,085,304	1,085,304	1,141,068	1,141,068	1,141,068
thereof: ON	2,571	2,571	2,571	2,571	3,550	3,550	3,550
thereof: OFF	6,117	6,117	6,117	6,117	7,305	7,305	7,305
Exporters	257	257	257	257	260	260	260
Importers	257	257	257	257	260	260	260
Years	54	54	54	54	70	70	70
Coefficients	402	2	402	2	526	526	526
Exporter × importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross-border × year FE			Yes	Yes			

Notes: The table presents regression results using variants of an ETWFE specification (equation (1)), in which the cohort-event-time-specific treatment effects were aggregated using equations (4) and (5) to obtain aggregate ON and OFF treatment effect estimates, respectively. The dependent variable is the natural logarithm of exports, which varies across the exporter-importer-year dimension. Columns (1)-(2) re-estimate the baseline ETWFE and TWFE models on the 1966-2019 sample without domestic trade flows. Columns (3)-(4) augment the dataset with domestic trade flows from the CEPII TradeProd database and include cross-border×year fixed effects to absorb globalization-related shocks. Columns (5)-(7) return to the baseline sample and ETWFE specification but vary the clustering of standard errors Standard errors, in parentheses, are clustered by country pair in columns (1)-(4), by exporter and importer in column (5), by exporter-year and importer-year in column (6), and by exporter, importer, and year in column (7). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

A.6.5 Alternative clustering.

In our last experiment, we examine the robustness of our results to alternative clustering methods for the standard errors in columns (5)-(7) of Table A8. In column (5), we double-cluster the standard errors by exporter and importer, resulting in larger standard errors, and the ON estimate becomes significant at the 10% level. In column (6), we double-cluster the standard errors by exporter-year and importer-year, leading to smaller standard errors, with the ON estimate achieving significance at the 1% level. Finally, in column (7), we triple-cluster standard errors by exporter, importer, and year. This approach increases the standard errors again, resulting in the ON estimate being significant only at the 10% level.