



CITP
CENTRE FOR
INCLUSIVE
TRADE POLICY

Heterogeneous difference-in-differences for monadic treatments in dyadic flow data

Rodolphe Desbordes, Markus Eberhardt, and Mario Larch

September 2026

Centre for Inclusive Trade Policy
Working Paper No.043



Economic
and Social
Research Council

Centre for Inclusive Trade Policy

<https://citp.ac.uk/> info@citp.ac.uk

Established in 2022, the Centre for Inclusive Trade Policy (CITP) is the first research centre dedicated to trade policy to be funded by the Economic and Social Research Council. As a centre of excellence for innovative research on trade policy and its inclusiveness, we aim to equip the UK with the capability to formulate and implement a trade policy tailored to the needs of the whole of the UK, while recognising the importance of the multilateral trading system and the UK's role within it. The CITP is funded by the Economic and Social Research Council [grant number ES/W002434/1]

This Working Paper is issued under the auspices of the Centre's research programme. Any opinions expressed here are those of the author(s) and not those of the Centre for Inclusive Trade Policy. Research disseminated by CITP may include views on policy, but the Centre itself takes no institutional policy positions.

These Working Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character. The author(s) asserts their moral right to be identified as the author of this publication. For online use, we ask readers to link to the webpage for the resource on the Centre's website.

Abstract

How do country-level (policy) changes affect bilateral flows? Estimating such monadic effects in dyadic data is challenging because treatment is collinear with directional-time fixed effects, pair-specific exposure to global shocks violates parallel trends, and network-wide equilibrium adjustments contaminate standard workarounds. We develop a common correlated effects difference-in-differences estimator that augments pair-level PPML regressions with cross-sectional averages from never-treated countries. Under a testable factor-spanning assumption, the estimator recovers the total general equilibrium effect. Simulations confirm identification. Applied to democratic regime change, we find large treatment effects and, using results from a border exercise alongside exact hat algebra, corroborate our estimator.

©: Rodolphe Desbordes, Markus Eberhardt, and Mario Larch

Suggested citation

R, Desbordes; M, Eberhardt; M, Larch (2026) Heterogeneous difference-in-differences for monadic treatments in dyadic flow data, Centre for Inclusive Trade Policy, Working Paper 043

Non-Technical Summary

How can we measure the impact of country-level policy changes on international flows?

Many important policy questions concern changes that occur at the national level and then affect its relationships with all other countries. For example: What happens to trade flows when a country becomes more democratic? How do capital controls affect international investment flows? How do migration reforms change migration flows?

Answering such questions is surprisingly difficult. Economists often study trade, investment, and migration using data on flows between pairs of countries. Existing statistical methods work well when a policy directly targets a pair of countries, such as a free trade agreement between two countries. When the policy affects individual countries, but changes interactions with every partner (and even non-partners) simultaneously, the approach cannot measure the magnitude of the economic effect of such a 'treatment'.

This paper develops a new method to overcome these challenges. The method is designed specifically for situations in which a policy or institutional change occurs at the country level while outcomes are observed at the level of pairs of countries. Rather than comparing affected countries directly with unaffected ones, the approach models and removes the influence of global forces that affect all countries differently, such as global business cycles or commodity-price shocks. This allows the researcher to isolate the effect of the country-specific policy change more credibly.

Application: Does democratisation increase trade?

The paper applies the new method to a longstanding question in political economy: does transition to democracy increase a country's international trade? Using bilateral trade data covering more than 200 countries between 1960 and 2018, and four alternative measures of democratic transition, the analysis finds a consistently positive answer. Democratisation increases a country's exports by approximately 18% to 34%, depending on the democracy measure used.

Theory predicts that democracy affects trade through three channels. First, democratic transitions reduce barriers to international commerce, making it easier for firms to trade across borders (the direct 'trade cost channel'). Second, democracy promotes economic growth and higher incomes in the democratising economies, which in turn support greater trade activity (the indirect 'own-income channel'). And finally, the effects of democratic regime change (like other shocks) propagate to a country's trading partners and eventually across the entire network, changing trading patterns for all countries (the indirect 'network channel'). The estimated effects combine lower trade costs, income gains resulting from democracy, and a global adjustment mechanism.

Our estimated effect captures all three channels together. In other words, the results measure the overall impact of democratisation on trade as it actually unfolded in the economy, including global adjustments (affected by who else democratised), rather than measuring only the direct effect of reduced trading costs.

A particularly interesting finding relates to the durability of the benefits of democratic regime change: since a substantial part of the democracy effect is driven by a reduction in trade costs, it is perhaps not that surprising that we find that total effects are not substantially larger if we limit our analysis to those countries who experienced regime change and then remained democratic for the remainder of our sample period. This implies that democratic collapse may not result in a substantive export penalty, even if incomes may drop as a result.

Why does this matter?

The paper develops a new econometric method for estimating the effects of country-level policies in bilateral flow data. The contribution extends beyond the specific case of democracy and trade. Many policy questions involve country-level interventions that affect bilateral relationships, yet researchers have lacked reliable tools for studying them. The new methodology provides a practical way to evaluate policies that influence cross-border flows (of goods, services, capital, or people) while accounting for global shocks, network effects, and changing treatment patterns over time.

Heterogeneous Difference-in-Differences for Monadic Treatments in Dyadic Flow Data*

Rodolphe Desbordes¹, Markus Eberhardt², and Mario Larch³

¹SKEMA Business School and Université Côte d’Azur

²School of Economics, University of Nottingham and CEPR

³University of Bayreuth, CEPR, ifo/CESifo, WIFO, and GEP

This version: September 16, 2026

Abstract: How do country-level (policy) changes affect bilateral flows? Estimating such monadic effects in dyadic data is challenging because treatment is collinear with directional-time fixed effects, pair-specific exposure to global shocks violates parallel trends, and network-wide equilibrium adjustments contaminate standard workarounds. We develop a common correlated effects difference-in-differences estimator that augments pair-level PPML regressions with cross-sectional averages from never-treated countries. Under a testable factor-spanning assumption, the estimator recovers the total general-equilibrium effect. Simulations confirm identification. Applied to democratic regime change, we find large treatment effects and, using results from a border exercise alongside exact hat algebra, corroborate our estimator.

Keywords: difference-in-differences, dyadic flow data, monadic treatments, heterogeneous treatment effects, common correlated effects, democratic regime change

JEL codes: C23, F13, F14, P16

*We thank Arturas Juodis, Clément de Chaisemartin, and seminar participants at Nottingham GEP, Kent University, FIW Vienna, IfW Kiel, SKEMA Business School Paris, the Midwest International Trade Conference, Freiburg University, the AMSE Conference in Applied Econometrics, and the TRISTAN workshop at Bayreuth University for useful comments and suggestions. We are indebted to Lena Boneva and Oliver Linton for sharing their rough derivations for a CCE-augmentation in the generalised linear model. This work was supported by the Centre for Inclusive Trade Policy (CITP) Innovation Fund, ESRC (Eberhardt) [grant ES/W002434/1]. We acknowledge the use of generative AI (Claude, Anthropic) to assist with code development, formatting of the paper, and editorial suggestions. The inception of the estimator as well as all analytical decisions, empirical specifications, and interpretations are the authors’ own. The usual disclaimers apply. Correspondence: Mario Larch, Chair of Empirical Economics, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany. Email: mario.larch@uni-bayreuth.de.

1 Introduction

The heterogeneous difference-in-differences (het-DID) revolution has transformed causal inference in panel data. Recent contributions have documented that the standard two-way fixed effects (TWFE) estimator can produce severely biased estimates when treatment effects vary across units or over time, particularly under staggered adoption (Goodman-Bacon 2021, de Chaisemartin & D’Haultfœuille 2020, Callaway & Sant’Anna 2021, Sun & Abraham 2021, Borusyak et al. 2024). Robust alternatives now exist that avoid the negative weighting and forbidden comparison problems inherent in pooled TWFE estimation (Roth et al. 2023), including for nonlinear models (Wooldridge 2023, 2025).

Many policy-relevant questions involve monadic treatments in dyadic flow data: a country-level intervention that reshapes the country’s bilateral flows with all its partners. Does democratic regime change alter export patterns? Do capital controls redirect cross-border equity flows? Does the lifting of emigration restrictions reshape a country’s migration flows with all destinations? Does regulatory harmonisation expand or divert trade in services? In each case, the treatment D_{it} varies at the country-time level while the outcome $Y_{ij,t}$ is observed at the country-pair-time level. The standard empirical specification for bilateral flows includes directional-time fixed effects (μ_{it}, ν_{jt}) , which absorb everything that varies at the country-time level: output, expenditure, and the multilateral resistance terms (MRTs) through which a country-level shock propagates to its partners and across the entire network. With a bilateral treatment such as a free trade agreement or a bilateral investment treaty, this works well, because the policy varies across pairs within the same origin-time cell. With a monadic treatment, D_{it} varies at the exactly same level as μ_{it} and is perfectly absorbed. This situation represents an empirical ‘straitjacket’, whereby researchers are severely constrained in the questions they can ask, namely about the effects of dyadic but not of monadic treatments or shocks, if identification is to remain in line with the rigorous canonical empirical framework.

Dropping the fixed effects and turning to the het-DID toolkit instead, as we propose, runs into additional challenges. First, the parallel trends assumption is implausible, because the same globalisation wave moves every country pair’s flows by a pair-specific amount. And second, SUTVA, the stable unit treatment value assumption, that one unit’s treatment leaves other units’ outcomes unaffected, always fails: a country’s treatment shifts the MRTs and hereby also flows even between untreated countries, so that untreated flows are contaminated controls. We introduce the heterogeneous Common Correlated Effects Difference-in-Differences (CCE-DID)

estimator to escape the ‘straitjacket’.

Our empirical strategy implicitly replaces the directional-time fixed effects with a narrower control that absorbs what has an immediate effect on every country’s bilateral flows: global shocks, such as the world business cycle or waves of globalisation. Following the common factor, or ‘interactive fixed effects’, literature ([Pesaran 2006](#), [Bai 2009](#), [Boneva & Linton 2017](#), [Chen et al. 2021](#)), which deals with such unobserved, time-varying heterogeneity, we estimate a separate flow regression for each treated pair but augment it with the cross-sectional averages (CAs) of income computed from never-treated countries. This approach provides us with solutions to all three problems raised above. The treatment is no longer absorbed, because the averages are common to all country pairs but constructed from countries that are never treated. Heterogeneous exposure to global shocks is accommodated, because each country pair has its own coefficients (‘factor loadings’) on the averages, and hence its own sensitivity to global shocks. And contaminated controls are avoided, because each country pair’s treatment effect is identified *against its own path predicted from the global shocks*, not against the flows of untreated pairs. Never-treated countries enter only through their averaged income, which, unlike their bilateral flows, the treatment reaches only weakly and gradually through the network. What the averages do not absorb is the response of output, expenditure, and MRTs to the treatment itself, so our estimate is the *total general-equilibrium effect* of the treatment, an object we argue below is of interest in its own right. Because the estimator delivers a separate effect for every pair-wise flow of a treated country, it is by construction unaffected by the negative weighting that afflicts staggered two-way fixed effects ([de Chaisemartin & D’Haultfœuille 2020](#), [Goodman-Bacon 2021](#)), and *inter alia* can be aggregated into an average treatment effect on the treated, or used more flexibly to study how effects vary across partners. In our empirical application on the trade effects of democratic regime change, an independent border exercise on domestic trade flows corroborates our total general-equilibrium effect estimate, while exact hat algebra decomposes it into a trade-cost and an income component.

The price to pay for breaking the straitjacket is first and foremost an assumption about our proxies for the common shocks. Identification rests on ‘factor spanning’: the cross-section averages, in combination with a pair-specific linear trend, must adequately capture the global shocks. The trend takes up deterministic drift, such as the secular rise in trade. Adequate factor spanning *rules out* a shock that the averages miss but that coincides with the treatment,

most likely one that exclusively hits individual treated countries¹ — similar to the conditional unconfoundedness assumption in standard DID estimators.

The spanning assumption itself is testable: we introduce a diagnostic, the ‘ Δ CSD’ (cross-section dependence difference) test, which flags unabsorbed structure in the treatment regression residuals. Estimating bilateral flow regressions *pair by pair* requires long panel data, though many existing bilateral trade, capital, and migration flow datasets fit this requirement. That income is itself an equilibrium outcome poses no problem, since the cross-section averages need only be uncorrelated with the treatment, not with the shocks they absorb.

We make four contributions. First, we establish identification of monadic effects in dyadic flow data. Pair-specific regressions augmented with never-treated averages identify a monadic treatment effect under factor spanning, and we bound the contamination of those averages by the treatment as a second-order term that the pair-specific trend absorbs.

Second, we characterise the estimand and argue that it is an object of interest in its own right. The coefficient identified by a monadic treatment in our CCE-DID is its *total general-equilibrium* effect: the *direct partial effect* together with indirect effects operating through the treated country’s *own-income effect*, and through the induced movement of the multilateral resistances (the *network effect*), quantified together rather than netted out. It is a reduced-form estimate, in line with those analysed in standard het-DID applications in other fields and data contexts, intended as a measurement exercise, instead of a precursor to the counterfactual exercise which gives structural gravity estimation its main purpose.² To our knowledge, this is the first reduced-form route to that object for a monadic regressor, and it is attractive for its policy-relevance, quantifying the change in a country’s own trade that its treatment brought about in the equilibrium that prevailed: what *actually* happened, rather than a counterfactual holding all else equal. A simulation study confirms that our estimator recovers this total effect, inclusive of the induced income response, to within Monte Carlo sampling noise. The simulation data-generating process is a structural gravity economy: bilateral flows solve the multilateral-resistance system exactly, and income enters as a factor-loaded fundamental that the estimator conditions on. The partial-equilibrium cost parameter, by contrast, is not attainable by our estimation, but only by an infeasible oracle. Because the trade data used in our empirical

¹Some candidate shocks, such as a natural resource discovery or a banking crisis, on closer inspection can be argued to be driven by global factors (see [Cotet & Tsui 2013](#), [Arellano et al. 2017](#), [Arezki et al. 2017](#), [Cesa-Bianchi et al. 2019](#)).

²This distinction matters for trade economists, who associate structural gravity with deep, transportable parameters, such as the trade elasticity, that feed counterfactual simulations, rather than with a reduced-form measurement of what a policy did in a given sample.

application also contain domestic flows, a border exercise enables us to decompose the total: holding income fixed while the price system adjusts isolates the trade-cost and multilateral-resistance component (the income-fixed *conditional* GE), and the remaining gap to our estimate is the induced income channel. In our trade flow application, an entirely separate empirical model thereby corroborates that our CCE-DID estimator identifies the reduced-form total effect. Since many applications lack domestic flow data, this corroboration is a rare opportunity which underscores the usefulness of our estimator.

Third, we develop a testing procedure that provides a diagnostic check on the factor-spanning identification assumption, exactly sized under a pure null and with power against economically meaningful violations in simulation.

Fourth, we apply our method to the study of democratic regime change and bilateral exports, on trade data that include domestic flows ([Head & Mayer 2021](#)) and hence permit a treatment decomposition unavailable in the vast majority of applications. Democratisation onsets raise exports by 0.17–0.29 log points (roughly 18 to 34%) across four alternative democracy measures ([Lührmann et al. 2018](#), [Treisman 2020](#), [Coppedge et al. 2021](#), [Maerz et al. 2024](#)). The border exercise identifies a larger partial-equilibrium trade-cost effect, and exact hat algebra on the same trade matrix maps that partial effect, through multilateral-resistance feedback, to an income-fixed conditional GE effect of about 0.11 to 0.19 log points; treated countries’ own induced income gains account for the remainder. Our estimate is therefore already the total general-equilibrium effect, not a lower bound to be grossed up; the border exercise and hat algebra attribute it to a trade-cost component and an income component (the latter +0.10 to +0.11 for three of the democracy codings, and +0.02 for the fourth, [Lührmann et al., 2018](#)), reconstructing what our estimator recovers directly.

Our work contributes to several literatures. Within the het-DID literature, we extend these methods to settings where the treatment unit (country) and the outcome unit (country pair) do not coincide, and where network-wide equilibrium propagation creates complex contamination channels. Our factor model-based approach relates to [Gobillon & Magnac \(2016\)](#), [Xu \(2017\)](#), [Chan & Kwok \(2022\)](#), and [Liu et al. \(2024\)](#), who use interactive fixed effects to capture or impute counterfactual measures, but these methods have not been adapted to the dyadic setting with its additional challenges to identification.

We build on the common factor panel literature ([Pesaran 2006](#), [Boneva & Linton 2017](#)) by adapting the CCE approach to a causal inference setting in dyadic data, where the causal

hierarchy of variables used for CA construction is central. [Chen et al. \(2021\)](#) show that nonlinear factor models can capture important features of dyadic network data and confirm in a network application that multiple latent factors are empirically relevant. Whereas these authors estimate the full nonlinear factor structure, we use CCE proxies constructed from never-treated controls, thereby avoiding high-dimensional factor estimation and the associated bias corrections. We further study a distinct estimand, monadic treatment effects under staggered adoption in dyadic panels, and develop a heterogeneous DID estimator with pair-specific effects.

In the empirical literature using bilateral flow data, several recent contributions address related but distinct identification problems. [Nagengast & Yotov \(2025\)](#) apply het-DID methods to bilateral treatments, where treatment varies at the pair-level and is directly identified within standard structural approaches using directional-time fixed effects. [Heid et al. \(2021\)](#), [Beverelli et al. \(2024\)](#) and [Freeman et al. \(2025\)](#) develop structural approaches for country-specific policies in bilateral settings, imposing parametric assumptions and estimating average (pooled) partial effects. Our contribution is complementary: we address *monadic* treatments, such as democratisation, that move several primitives at once (trade costs, productivity), so that no single structural parameter summarises the policy; the object of interest is then the total general-equilibrium effect defined above. We recover it without estimating the complete parametric structural model, yet the coefficient remains a structurally interpretable object. Where the data include domestic flows, the clean partial effect is identified directly by the border exercise, which serves as our empirical benchmark (Section 5.7).

The remainder of the paper is organised as follows. Section 2 describes the setup. Section 3 presents the CCE-DID estimator and discusses inference. Section 4 summarises simulation evidence on finite-sample properties and introduces the factor spanning diagnostic. Section 5 applies the estimator to study the effect of democratic regime change on bilateral exports. Section 6 concludes.

2 Setup

Consider bilateral flow data $Y_{ij,t}$, representing trade, capital, or migration flows from origin i to destination j at time t , where the treatment $D_{it} \in \{0, 1\}$ is a monadic (unit-level) binary indicator. Treatment is staggered (different units adopt at different times) and may reverse. For ease of exposition, and in light of our empirical application, we consider democratic regime change in the setting of bilateral trade flows between countries when developing the empirical

setup and in the simulation exercises in Section 4. Let $Y_{ij,t}(d)$ denote the potential outcome for pair (i, j) at time t when origin i 's treatment status is $d \in \{0, 1\}$, with the observed outcome $Y_{ij,t} = D_{it} \cdot Y_{ij,t}(1) + (1 - D_{it}) \cdot Y_{ij,t}(0)$. For a total of N origin countries, we have N_{tr} treated and N_c never-treated (control) countries.

A subtlety arises from the network structure of dyadic flows. Standard potential outcomes assume that one unit's treatment does not affect another unit's outcome (SUTVA). In dyadic flow data this is violated through a specific mechanism: treatment changes country i 's primitive characteristics, these enter a system-wide equilibrium as inputs, and the equilibrium solution shifts for all countries, altering flows even between untreated countries.

Our estimand, a *total reduced-form treatment effect*, is defined conditional on this equilibrium response. The potential outcome $Y_{ij,t}(1)$ denotes the flow observed when $D_{it} = 1$, given the full vector of treatment assignments realised in the data. The estimand is specific to the realised treatment history: it measures the effect of country i 's treatment in the equilibrium that actually prevails, which depends on which other countries are treated simultaneously. It is the change in i 's own flows, inclusive of whatever equilibrium adjustments occur, and it is a relevant object for ex-post policy evaluation when the treatment moves several structural primitives jointly. In the language of the structural gravity literature, the estimator lets the multilateral resistances and the treated country's own income adjust rather than holding them fixed, and the estimand includes their response: a *total general-equilibrium effect*. This is distinct from the income-fixed *conditional* GE, the trade-cost component that the border exercise of Section 5.7 recovers separately. The simulations of Section 4.2 make the distinction from the partial-equilibrium cost parameter exact, and Section 5.7 decomposes the total in the data. For applications where the treatment maps cleanly into a single structural parameter (for instance, a bilateral cost-shifter), a partial-effect decomposition is feasible and existing structural techniques are appropriate; Section 5.7 makes the relationship explicit in our application.

In line with common practice in the literature on dyadic flow data, we model the outcome as an exponential linear model, but introduce a multi-factor error structure:

$$Y_{ij,t} = \exp(\alpha_{ij} + \lambda'_{ij} \mathbf{F}_t + \theta_{ij} \cdot D_{it} + \delta_{ij} \cdot D_{jt} + \beta'_{ij} \mathbf{X}_{ij,t}) \cdot \eta_{ij,t} \quad (1)$$

where $\mathbf{F}_t$ is a $K \times 1$ vector of unobserved common factors (capturing global business cycles, technology, commodity prices, financial conditions), λ_{ij} is a $K \times 1$ vector of pair-specific factor loadings, α_{ij} is a pair-specific fixed effect (capturing time-invariant bilateral characteristics),

θ_{ij} is the pair-specific origin treatment effect (the object of interest), δ_{ij} captures the effect of destination treatment, $X_{ij,t}$ are additional dyadic controls with pairwise coefficients β_{ij} , and $\eta_{ij,t}$ is an idiosyncratic error with $\mathbb{E}[\eta_{ij,t} \mid F_t, D_{it}, D_{jt}, X_{ij,t}] = 0$.

This formulation treats the origin as the primary treated unit but does not restrict destinations. Including D_{jt} absorbs destination treatment variation that would otherwise confound the origin effect θ_{ij} ; because effects are estimated pair by pair, they can additionally be aggregated by destination characteristics if that heterogeneity is of interest.

We further consider country-level variables that capture observable characteristics of each country, drawn from any source that loads on the common factor structure. In our empirical application we use country-level income; the framework equally accommodates population, demographic shares, capital stock measures, electricity consumption, and other aggregates. Let $z_{m,it}$ denote the m -th such variable:

$$\ln z_{m,it} = \alpha_i^{z_m} + \gamma_i^{z_m'} F_t + \theta^{z_m} D_{it} + u_{m,it}, \quad m = 1, \dots, M. \quad (2)$$

Some of these variables may respond to treatment ($\theta^{z_m} \neq 0$), others may not. The framework requires only that some such variables load on the common factor structure and remain available in the never-treated subsample, where θ^{z_m} is zero by construction.

Bilateral frictions (due to transaction costs, regulatory distance, informational barriers) also follow a factor structure:

$$\ln \tau_{ij,t} = \alpha_{ij}^\tau + \gamma_i^{\tau'} F_t + \gamma_j^{\tau'} F_t + \theta^\tau D_{it} + \xi_{ij,t}. \quad (3)$$

As is standard in the factor-model panel literature (Pesaran 2006, Bai 2009, Chen et al. 2021, Chan & Kwok 2022) we assume that the country-level variables and the bilateral flow outcome share a common factor structure. Under the maintained assumption of time-constant effects, the object of interest is the pair-level treatment effect $\theta_{ij} = \ln Y_{ij,t}(1) - \ln Y_{ij,t}(0)$ for all t . An event-study analysis in Section 5.5 relaxes time-constancy to allow $\theta_{ij,\ell}$ indexed by event time ℓ .

3 Estimation and Inference

3.1 Construction of the CCE-DID Estimator

Cross-sectional averages. Our approach augments each pair-level regression with cross-sectional averages (CAs) that absorb confounding common factor variation without also

absorbing the treatment. Following [Pesaran \(2006\)](#), consider the cross-sectional average of an observable $z_{c,t}$ but over never-treated countries $c \in \mathcal{C}$:

$$\bar{z}_t^{\text{ctrl}} = \frac{1}{N_c} \sum_{c \in \mathcal{C}} \ln z_{c,t} = \bar{\alpha}_{\text{ctrl}} + \bar{\gamma}'_{\text{ctrl}} \mathbf{F}_t + \bar{u}_t^{\text{ctrl}} \quad (4)$$

where $\bar{\gamma}_{\text{ctrl}} = N_c^{-1} \sum_{c \in \mathcal{C}} \gamma_c$ is the average factor loading across control countries. As $N_c \rightarrow \infty$, $\bar{u}_t^{\text{ctrl}} \xrightarrow{p} 0$ by the law of large numbers, yielding:

$$\bar{z}_t^{\text{ctrl}} \approx \bar{\alpha}_{\text{ctrl}} + \bar{\gamma}'_{\text{ctrl}} \mathbf{F}_t \quad (5)$$

Since $\bar{\gamma}_{\text{ctrl}} \neq 0$, the control CA is a (rotated) proxy for the common factors. Including it as a regressor absorbs the common factor variation that would otherwise confound the DID comparison. More generally, with M distinct CA variables (each computed from a different observable), the CAs can span a factor space of dimension $K \leq M$. In our application the primary specification pairs a single control-country (ln) GDP average with a pair-specific linear trend; the estimating equation below sets out the distinct role each plays.

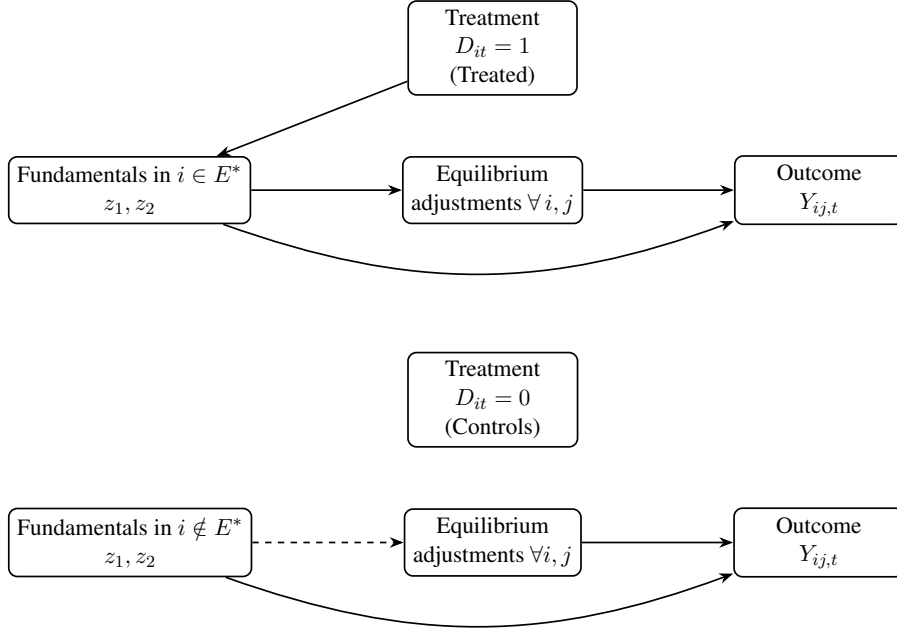
Two restrictions govern our CA construction. First, CAs must be computed from never-treated countries only. If all-sample CAs were used instead:

$$\bar{z}_t^{\text{all}} = \bar{\alpha} + \bar{\gamma}' \mathbf{F}_t + \theta_z \cdot \underbrace{\frac{N_{\text{tr}}}{N} \bar{D}_t}_{\text{treatment contamination}} + \bar{u}_t, \quad (6)$$

where $\bar{D}_t = N_{\text{tr}}^{-1} \sum_{i \in \text{treated}} D_{it}$ is the average treatment intensity among treated countries at time t and N_{tr}/N is the treated share of the sample. The all-sample CA absorbs part of the treatment effect, making it a ‘bad control’: a post-treatment variable that is itself a function of the treatment. Control-only CAs avoid this contamination. Since control countries are never treated, no treatment term enters equation (4), and the cross-sectional average remains a clean factor proxy. This is the CCE analogue of the ‘never-treated’ restriction in the het-DID literature ([Callaway & Sant’Anna 2021](#), [Chan & Kwok 2022](#)) and reflects the same principle as the ‘forbidden comparison’ concern raised by [de Chaisemartin & D’Haultfœuille \(2020\)](#): treated units’ outcomes must not enter the counterfactual.

The second restriction concerns which variables to average. One might want to use CAs of bilateral flows themselves or of proxies for the network-wide equilibrium adjustment terms, since these also load on $\mathbf{F}_t$. The causal hierarchy of the data-generating process (DGP), illustrated in

Figure 1: Causal Hierarchy in Treated versus Control Countries



Notes: The diagram proposes the causal hierarchy in treated (top) versus control (bottom) countries. The dashed line signals that control-country GDP responds to treatment elsewhere through trade multipliers; Section 3.2 bounds this feedback by $O(\omega(N_{tr}/N)\theta)$, the product of openness, treated country share, and shock size, small and smooth, hence absorbed by the pair trend.

Figure 1, shows why this fails.³ Treatment moves country-level primitives: productivity and bilateral costs in our trade application, upstream objects in general (Head & Mayer 2014). The equilibrium then determines incomes, price indices, and flows for every country jointly, so when country i is treated, equilibrium objects shift everywhere, including for never-treated countries.⁴ What matters for CA construction is *how* treatment reaches an observable, not whether it is structurally exogenous. For control countries ($D_{ct} = 0$), country-level observables carry no own-treatment term and reflect the common factors and noise,

$$z_{m,ct} = \alpha_c^{z_m} + \gamma_c^{z_m'} \mathbf{F}_t + u_{m,ct}, \quad m = 1, \dots, M. \quad (7)$$

Any treatment spillover from elsewhere in the network is small and smooth: it reaches a control unit only in proportion to its exposure to the treated bloc (openness times the treated share)

³The average of the dependent variable is inadmissible for a further reason, specific to the nonlinear model: under the exponential link the average of Y enters the index nonlinearly and no longer proxies the factors cleanly, so the CCE augmentation is built from covariate averages only (Boneva & Linton 2017, Juodis & Reese 2026).

⁴Measured GDP includes trade-related value added, so control-country GDP responds to treatment elsewhere through trade multipliers; Section 3.2 bounds this feedback by $O(\omega(N_{tr}/N)\theta)$, the product of openness, treated share, and shock size, small and smooth, hence absorbed by the pair trend, whereas contamination of constructed multilateral-resistance terms is first-order.

times the pass-through of each shock, and it co-moves with the roll-out of treatment, so the pair-specific trend we add to our empirical implementation absorbs it (Section 3.2).

Constructed pair-level equilibrium terms are different in kind: they are solved from the full system that includes treated countries' changed primitives, so each treated country contributes at first order, and CAs of such objects (or of bilateral flows themselves) inherit treatment contamination even when computed from control countries only; the simulations and the border benchmark return to the equilibrium-term case (Sections 4.2 and 5.7). We therefore construct CAs from country-level observables of never-treated countries, whose contamination is negligible, and never from pair-level equilibrium objects, whose contamination is structural. Control-country income is itself an equilibrium outcome, but that does not disqualify it. The cross-sectional average need only be approximately exogenous with respect to the *treatment*, which the second-order dilution above delivers, while its response to the common shocks is exactly what lets it proxy the factors the estimator must remove (Section 3.2); we return to which averages are admissible below.

Estimation. The pair-level estimation equation is then:

$$\mathbb{E}[Y_{ij,t} \mid D_{it}, D_{jt}, t, \bar{\mathbf{z}}_t^{\text{ctrl}}] = \exp(\alpha_{ij} + \theta_{ij} \cdot D_{it} + \delta_{ij} \cdot D_{jt} + \kappa_{ij} t + \phi'_{ij} \bar{\mathbf{z}}_t^{\text{ctrl}}). \quad (8)$$

Here α_{ij} is a pair fixed effect absorbing time-invariant bilateral heterogeneity, D_{it} and D_{jt} are the origin and destination treatment indicators, $\bar{\mathbf{z}}_t^{\text{ctrl}}$ is the vector of CAs from never-treated countries with pair-specific coefficients ϕ_{ij} , and $\kappa_{ij} t$ is a pair-specific linear trend. The trend and the averages divide the confounding structure between them: the cross-sectional averages span the *stochastic* common factor, while the pair-specific trend spans its *smooth* components, secular growth in bilateral trade and the gradual roll-out of treatment across the panel, which a mean-zero stochastic average cannot absorb. The two are complementary, not redundant, and the simulation of Section 4.3 shows that neither alone recovers the effect while the two together do. Including D_{jt} ensures that θ_{ij} captures the origin treatment effect net of any destination regime change occurring over the same period.

Dyadic controls (e.g., FTA membership, common currency) may be included in $\mathbf{X}_{ij,t}$:

$$\mathbb{E}[Y_{ij,t} \mid \cdot] = \exp(\alpha_{ij} + \theta_{ij} \cdot D_{it} + \delta_{ij} \cdot D_{jt} + \kappa_{ij} t + \mathbf{X}'_{ij,t} \boldsymbol{\beta}_{ij} + \phi'_{ij} \bar{\mathbf{z}}_t^{\text{ctrl}}). \quad (9)$$

The PPML estimator is consistent under correct specification of the conditional mean,

without requiring correct specification of the conditional variance (Santos Silva & Tenreyro 2006). The exponential form naturally accommodates zero flows and is robust to heteroskedasticity. The CCE consistency results of Pesaran (2006) are established for the linear case. Boneva & Linton (2017) extend the CCE to a binary dependent variable, explicitly excluding the cross-sectional average of the dependent variable. Juodis & Reese (2026) formalise the general principle for single-index nonlinear models, the Poisson (PPML) case among them: the factors are identified only from the covariate averages, and the dependent-variable average is inadmissible because it proxies a nonlinear function of the factors rather than the factors themselves. They also note that consistency and asymptotic-normality theory for the empirically relevant case of more averages than factors ($R < K + 1$) is not yet established in the nonlinear setting. Formal asymptotic theory for our CCE-DID in the nonlinear setting therefore remains to be developed; our simulation study, which uses the PPML estimator throughout, examines finite-sample performance in data calibrated to realistic trade panels (Section 4).

The pair-specific coefficients ϕ_{ij} on the CAs reflect the structure of dyadic flow data. In network settings, pair-level factor loadings inherit dependence from both endpoint countries (the two nodes joined by each directed pair, or edge) and from each country’s position in the network of flows. Chen et al. (2021) formalise this for nonlinear factor models in network and panel data: the latent factor structure underlying dyadic outcomes generates pair-specific loadings that vary across the $N(N - 1)$ directed pairs even when the underlying factor space has low dimension, capturing features such as degree heterogeneity, homophily in latent traits, and clustering. The relevant implication for our purposes is that the pair-level loading λ_{ij} on common factors F_t is a function of country-level loadings together with a network-position term that itself depends on the joint loading distribution. Whatever the precise functional form, pair-level sensitivity to common shocks is heterogeneous in a way that a common time effect μ_t cannot absorb.

This heterogeneity has three implications. First, the parallel trends assumption fails because treated pairs have systematically different λ_{ij} from control pairs. Second, pooled estimation fails because a common coefficient ϕ on the CAs cannot absorb pair-specific factor exposure. Third, Mean Group estimation is therefore necessary: estimating $(\alpha_{ij}, \theta_{ij}, \phi_{ij})$ separately for each pair and averaging allows each pair to absorb its own factor exposure. Pesaran (2006) showed in the two-way panel that the MG estimator is consistent for the average effect under loading heterogeneity, as long as a law of large numbers holds. Chan & Kwok (2022), specifying estimated factor proxies following Bai (2009) rather than CAs, extend this result to the DID

setting and establish consistency under $N_{\text{tr}}, T \rightarrow \infty$ with $\sqrt{N_{\text{tr}}}/T \rightarrow 0$.

Robust Mean Group Aggregation and Inference. To aggregate the pair-specific $\hat{\theta}_{ij}$ into an overall/sample ATET, we apply robust regression (iteratively reweighted least squares with Tukey’s biweight function; rreg) to all pair-wise estimates:

$$\hat{\theta}_{ij} = \mu + e_{ij}, \quad \hat{\theta}^{\text{rreg}} = \hat{\mu}, \quad (10)$$

where $\hat{\mu}$ is the robust intercept estimate, downweighting extreme outlier pairs.⁵ Since we estimate the granular pair-wise treatment effect, researchers are free to aggregate treatment effects by alternative sub-groups, enabling them to highlight structural differences in the result (e.g., advanced vs. developing country effects) or to investigate ‘mechanical’ distortions (e.g., quantifying the potential bias from treatment reversal).

Inference must account for the dependence structure inherent in dyadic data. Pairs sharing an origin share both D_{it} and likely factor loading dependence, inducing within-origin correlation; pairs sharing a destination share importer-side general-equilibrium adjustments. Following [Cameron et al. \(2011\)](#) (CGM), we use two-way clustering by exporter and importer, the standard approach for bilateral data: $\text{Var}_{\text{CGM}} = \text{Var}_{\text{exporter}} + \text{Var}_{\text{importer}} - \text{Var}_{\text{pair}}$, with $t(\min(G_{\text{ex}}, G_{\text{im}}) - 1)$ critical values, where $\text{Var}_{\text{exporter}}$ clusters by origin, $\text{Var}_{\text{importer}}$ clusters by destination, and Var_{pair} is the pair-level (robust) variance subtracted to avoid double-counting. The use of $t(\min(G_{\text{ex}}, G_{\text{im}}) - 1)$ critical values corrects for the smaller cluster dimension, following the recommendation for few-cluster settings. The relevant dimension is the number of treated exporters G , not the number of treated pairs: pairs sharing an exporter are not independent observations, so the effective cross-section for inference is G and the aggregate converges at rate $\sqrt{G}$ (Appendix B, Remark 1). In the empirical application, we additionally verify every baseline result against the more conservative exporter cluster bootstrap, and significance is robust to either choice (Section 5.2).⁶

⁵We use the term ATET for the rreg aggregation, implying it targets the mean; however, this is only correct under a random coefficient assumption: $\theta_{ij} = \theta^{\text{ATET}} + v_{ij}$, with $\mathbb{E}[v_{ij}] = 0$ and $\mathbb{E}[v_{ij}^2] = \sigma_v^2 < \infty$. More generally, rreg targets a robust M-functional (Tukey-biweight location) of pair-level effects, which coincides with the mean only under symmetry of e_{ij} ; under right-skew it lies below the mean, so the simulation evaluated against a mean target is, if anything, a conservative test of unbiasedness. We use rreg for robustness to the heavy pair-level tails and report that it is the location statistic throughout. See also Appendix B, Assumption B.1.

⁶In factor-augmented models, common factors induce residual cross-cluster dependence that no standard clustering method fully captures. CGM two-way clustering provides the best available approximation — see Appendix B, Assumption B.2, for a description of the dependence structure.

3.2 Identification

Factor Spanning. Our core identification assumption replaces parallel trends in conventional het-DID with a weaker factor spanning condition.

Assumption 1 (Factor Spanning). *Let $\bar{\gamma}_{ctrl}^{z_m} = N_c^{-1} \sum_{c \in \mathcal{C}} \gamma_c^{z_m}$ denote the average $K \times 1$ factor loading vector of the m -th CA observable across never-treated countries. Stack these into the $M \times K$ matrix $\bar{\Gamma} = [\bar{\gamma}_{ctrl}^{z_1}, \dots, \bar{\gamma}_{ctrl}^{z_M}]'$. The factor spanning condition requires: $\text{rank}(\bar{\Gamma}) = K$, where M is the number of distinct CA variables and K is the number of common factors.*

This requires that the CAs from never-treated countries collectively capture all K factors driving outcomes, demanding $M \geq K$ with non-collinear average loading vectors. Deterministic regressors count towards spanning: a pair-specific linear trend spans smooth drifting components of the factor space, so in the application a single income average together with the trend covers a two-dimensional structure, one stochastic factor and one drift, a configuration examined in the simulations of Section 4.3. An additional unspanned factor dimension is harmless provided it is uncorrelated with treatment, condition (b) in the following paragraph; Section 4.4 quantifies this case. Unlike parallel trends, which requires homogeneous factor loadings, factor spanning permits each pair its own sensitivity to the common factors, identifying the treatment effect from residual variation after removing pair-specific factor exposure.

Factor spanning fails only when all three of the following conditions hold simultaneously: (a) a factor $F_{k,t}^*$ loads exclusively on treated countries ($\gamma_{c,k}^* = 0$ for all $c \in \mathcal{C}$); (b) this factor is correlated with the treatment indicator conditional on the spanned factors ($\text{Cov}(F_{k,t}^*, D_{it} \mid \mathbf{F}_t) \neq 0$); and (c) this factor has non-trivial time variation ($\text{Var}(F_{k,t}^*) > 0$). If any one condition fails, the estimator remains consistent. Condition (a) is the most demanding: it requires that a shock affects only treated countries, with no control country sharing any exposure. Global business cycles, for instance, affect all countries, so CAs capture them regardless. Spanning fails only for shocks that load *exclusively* on treated countries, a much more specific failure mode than the general heterogeneous-loadings violation that invalidates parallel trends.

Additional assumptions. Our approach imposes four additional assumptions. First, sufficient cross-sectional units in the control sample ($N_c \rightarrow \infty$) ensure that CAs are reliable factor proxies. Second, sufficient time periods ($\sqrt{N_{tr}}/T \rightarrow 0$) are needed to estimate pair-specific parameters with adequate precision. Third, the observable variables used for CA construction must load on the relevant factors; in practice, gravity observables such as income are natural candidates since

they enter the structural model directly. Fourth, treatment must not be perfectly correlated with unspanned factors. The ΔCSD diagnostic (Section 4.4) provides a one-sided symptom check on whether the latter two conditions are met in practice: it can flag a treated-specific unspanned factor, but a clean pass is necessary rather than sufficient for spanning. Full sufficiency, as with parallel trends, cannot be certified by any single test; confidence is built by enriching the averaged observables until the diagnostic is stable, and in our application by the external agreement of the border exercise (Section 5.7). In our setting, these steps are available and they agree, so we regard the identifying assumption as credible.

On the choice of CA variables. The CCE methodology imposes weaker requirements on the variables used to construct cross-sectional averages than structural exogeneity in the sense of a particular economic model. What is required is that the chosen variables (a) load on the underlying common factors, and (b) are not contaminated by treated-country treatments to first order. Income in the never-treated countries, which we use in the empirical application, satisfies both. The two requirements pull in different directions without being in conflict: condition (b) demands exogeneity with respect to the treatment, which the second-order dilution detailed below delivers, while condition (a) is served best by endogeneity with respect to the system, since equilibrium outcomes respond to the common shocks both directly and through the network’s adjustment to them. We highlight this because income is endogenously determined in many structural frameworks (in particular structural gravity, where it is solved jointly with the price system as part of the equilibrium), which might appear to disqualify it as a factor proxy. It does not, because the CCE conditions are statistical, not structural: what matters is whether control-country income loads on common shocks and whether it is materially contaminated by treatments elsewhere in the panel. A variable can be endogenously determined within any structural model and still satisfy both conditions.

The contamination can be bounded directly, and generically for dyadic flows. A never-treated unit’s income, or any country-level observable used for a CA, responds to treatments elsewhere only through the network: when a treated unit’s primitives change, its demand shifts and its partners’ flows and incomes move. The magnitude of this feedback on the control cross-sectional average is the product of three small quantities. First, the shock reaches a control unit c in proportion to c ’s exposure to the treated bloc, $\sum_{k \in \mathcal{T}} s_{ck}$, where s_{ck} is the share of c ’s income accounted for by its flow with k ; with openness ω (the share of income that crosses borders, one minus the domestic share) and a treated share N_{tr}/N , this exposure is of order $\omega \cdot (N_{\text{tr}}/N)$.

Second, the pass-through of a treated unit's own shock to that flow is of order θ , the treatment's effect on the primitive. Third, the contamination that survives is *systematic*: every control country shares roughly the same exposure to the treated bloc, so it does not cancel in the average, but neither does averaging over the N_c controls shrink it. The combined order is

$$O(\omega \cdot (N_{tr}/N) \cdot \theta), \quad (11)$$

the product of openness, the treated share, and the shock size. In our application ($\omega \approx 0.3$, $N_{tr}/N \approx 0.2$, $\theta \approx 0.2$) this is about 0.01 log points. Two features make it harmless without any appeal to the size of the panel. It is small, by the product of three small factors; and it is *smooth*, co-moving with the gradual roll-out of treatment, so the pair-specific linear trend absorbs it, spanning smooth drifting components exactly as the discussion of Assumption 1 permits. That the residual is both small and smooth is why modelling the control observables as clean factor proxies, rather than solving them from a full equilibrium, costs essentially nothing: the term such a solution would add is one the trend already removes. The bound is generic to dyadic flows and *tightens* where the outcome and the observable are less tightly linked than trade and income; trade, in which income is the accounting sum of the very flows studied, is the adverse case, and migration or capital flows are milder. By contrast, contamination of constructed equilibrium-adjustment terms (multilateral resistances, or averages of the bilateral flows themselves) is *first* order: each treated unit enters them with weight $1/N$ through the equilibrium, not through a small pass-through, and the contamination is not smooth, so no trend removes it. This is why country-level averages of never-treated units are admissible controls while constructed equilibrium terms are not. Section 4.3 provides direct simulation evidence.

The estimand. The CCE-DID estimator recovers the total reduced-form ATET: the effect of a country's own treatment on its bilateral exports inclusive of every equilibrium adjustment it induces, the multilateral resistances and the treated country's own income effect. It is a general-equilibrium object, but not the income-fixed conditional GE of the structural gravity literature. Because the control-country averages span only the common factors, the treated country's own induced income response is idiosyncratic to it, absent from the control average, and collinear with D_{it} within the pair, so it stays in the residual and loads onto the coefficient; Section 4.2 demonstrates this in the simulation. The border exercise of Section 5.7 recovers the income-fixed conditional GE separately, as the trade-cost component of this total, and the gap

between the two is the induced income channel. Under Assumption 1 and as $N_{\text{tr}}, T \rightarrow \infty$ with $\sqrt{N_{\text{tr}}}/T \rightarrow 0$, the MG estimator is consistent for

$$\theta^{\text{ATE}} = \frac{1}{|\mathcal{M}|} \sum_{(i,j) \in \mathcal{M}} \theta_{ij}, \quad \theta_{ij} = \ln Y_{ij,t}^{\text{tr}} - \ln Y_{ij,t}^{\text{cf}}, \quad (12)$$

where $\mathcal{M}$ is the set of treated pairs, Y^{tr} is the observed outcome, and Y^{cf} is the counterfactual outcome absent treatment, the log difference being time-constant under the maintained assumption. This is an observational rather than a structural parameter: it is defined over the treated pairs and the treatment history realised in the specific sample at hand, so a different window or set of countries defines a different average.

Relationship to the existing het-DID literature. The CCE-DID framework subsumes standard het-DID concerns while addressing additional challenges specific to dyadic data. In place of group-time ATTs (Callaway & Sant’Anna 2021), we estimate pair-specific treatment effects θ_{ij} , disaggregating below the cohort level to accommodate arbitrary heterogeneity across destination partners. The principle of using never-treated controls carries over directly: control-only CAs serve the same role as never-treated comparison groups. We do not use not-yet-treated controls, because including future-treated countries in the CAs introduces contamination analogous to the forbidden comparisons identified by de Chaisemartin & D’Haultfœuille (2020).

Both the CCE-DID and imputation-based estimators (Borusyak et al. 2024) construct counterfactuals from control-group information, but the CCE approach uses a factor model rather than direct outcome imputation, requiring the weaker factor-spanning assumption. Whereas extended two-way fixed effects (ETWFE) (Wooldridge 2023, Nagengast & Yotov 2025) handles bilateral treatments varying at the pair level, CCE-DID handles monadic treatments varying at the origin level. A monadic treatment is perfectly collinear with the directional-time fixed effects, so in the absence of a domestic-versus-international contrast a two-way fixed-effects design cannot identify it. The border exercise supplies that contrast with domestic flows, and there we apply ETWFE (Section 5.7), while CCE-DID identifies the effect from international flows alone through control-country cross-sectional averages. Negative weighting (Goodman-Bacon 2021) is avoided by construction: each pair’s effect is estimated from its own data only, without cross-pair comparisons. Event-study estimation follows the same logic as in Sun & Abraham (2021), applied at the pair level and aggregated via MG.

4 Finite-Sample Performance and Diagnostics

We examine the finite-sample properties of the CCE-DID estimator in a structural-gravity simulation. Bilateral flows solve the [Anderson & Van Wincoop \(2003\)](#) system exactly, so the estimator is judged against an exact general-equilibrium counterfactual; income enters as a factor-loaded fundamental that the estimator conditions on, as in conditional-gravity estimation. The design isolates the identifying problem, separating a stochastic common factor and a smooth drift from the treatment, and lets the quality of the control averages be varied directly.

4.1 Data-Generating Process

A conditional-gravity economy. The DGP is single-sector structural gravity with N countries, T periods, and elasticity of substitution $\sigma = 5$. A single stochastic common factor F_t , an AR(1) process with coefficient 0.7, drives the country-level primitives alongside country-specific linear drifts: the factor is the stochastic confounder, the drifts the smooth one. Bilateral costs load on the factor and carry the monadic treatment,

$$\ln \tau_{ij,t} = \ln \bar{\tau}_{ij} + \gamma_i^T F_t + \gamma_j^T F_t + \frac{\theta^{\text{cost}}}{1 - \sigma} D_{it}, \quad \theta^{\text{cost}} = 0.36, \quad (13)$$

mapping into a structural log-trade (partial-equilibrium) effect of 0.36, calibrated to the border estimate of [Section 5.7](#). Income is a factor-loaded fundamental, the simulation counterpart of the country-level observable of [Section 2](#),

$$\ln Y_{it} = a_i + \gamma_i^Y F_t + \theta^{\text{gdp}} D_{it} + \kappa_i t + u_{it}, \quad \theta^{\text{gdp}} = 0.13, \quad (14)$$

so democracy raises a country's own income (the income channel) and the income of every never-treated country carries the same factor with its own factor loading. Given income and costs, the multilateral-resistance system is solved exactly. This is the conditional-gravity system rather than the full Anderson–van Wincoop equilibrium, in which income would itself be solved from market clearing. The conditional and total effects below are two counterfactual equilibrium computations of this system, income held fixed in one and allowed to adjust in the other; the partial effect is not a re-solved equilibrium but the structural cost coefficient, read by holding the resistances and income at their actual values.

Why income enters as a fundamental. This is a conditional-gravity design: income is an observed fundamental the estimator conditions on through the control-country average, as gravity is conventionally estimated (Head & Mayer 2014), rather than an object solved from the wage system. The cross-section average stands in for the directional-time effects a monadic treatment rules out, so, unlike fixed-effects gravity in which income is simply absorbed and its determination irrelevant, here the average must carry the common factor (and it matters how income evolves). This evolution is a property of the simulated income, not of the estimating equation. Income solved from market clearing is normalised against a world aggregate, so it carries no common movement and cannot stand in for a global cycle; what survives such a normalisation is a drift, and a simple pair-specific trend already spans a drift. Observed income is not normalised in that way and does co-move with global shocks. Static structural gravity contains no such global factor, so unless income carries it the simulation would omit the very confounder the estimator exists to remove; positing income as a factor-loaded fundamental supplies that feature while leaving the trade block fully structural. For the estimating equation the requirement is different and weaker: not structural exogeneity, but the statistical conditions of Section 3.2, that control income loads on the factors and be first-order uncontaminated by treatment. The design also lets us set the factor loadings, and with them the spanning parameter s , directly, and it extends to dyadic settings where the observable and the outcome are less tightly linked than income and trade (e.g., migration, capital flows), where the contamination bound (11) is smaller still.

Selection and spanning failure. Treatment onset is staggered, its timing tracking the country's own income path, its exposure to the stochastic factor and its drift, so that both confounders bear on selection: countries tend to democratise during global expansions (Milner & Mukherjee 2009, Gunitsky 2014). Spanning failure is parameterised by $s \in [0, 1]$, which dampens the control countries' loadings on the factor (in costs and income) while leaving treated loadings, and hence all true effects, unchanged. So s controls only how well the control averages proxy the confounder: at $s = 0$ they carry its full variation, at $s = 1$ they carry none of it.

4.2 Two Estimands: Partial and General Equilibrium

Because the true effect is known exactly, the simulation can distinguish the two objects an applied reader might have in mind. Writing the gravity equation in logs,

$$\ln Y_{ij} = \underbrace{(1 - \sigma) \ln \tau_{ij}}_{\text{trade cost}} + \underbrace{(1 - \sigma) \ln c_i}_{\text{own cost}} - \underbrace{(1 - \sigma) \ln P_j}_{\text{partner MR}} + \underbrace{\ln E_j}_{\text{partner income}},$$

Treatment enters this decomposition only through the first term, by cutting $\ln \tau_{ij}$, and the structural coefficient on that cut is θ^{cost} . The *partial-equilibrium* (PE) effect holds the own cost c_i , the partner multilateral resistance P_j , and partner income E_j at their factual values and removes the cost cut: it is then exactly $\theta^{\text{cost}} = 0.36$. Two general-equilibrium objects sit below it. The *conditional* GE holds the treated country's own income fixed and lets the multilateral resistances adjust; the resistance feedback pulls the effect well below the PE, to 0.16. The *total* GE additionally lets the treated country's own income respond to its democratisation, the channel $\theta^{\text{gdp}} D_{it}$ of equation (14), which raises the effect back to 0.26. This *total* is not the *full* equilibrium in which income is itself solved from the market-clearing wage system: income follows the fundamental of equation (14) throughout, and the conditional counterfactual switches off only its democratisation response, not the whole system. We keep *total* for the object the estimator targets and reserve *full* GE for the market-clearing solution, which the conditional-gravity design does not compute. Because the estimator is staggered and heterogeneous, each country's effect is identified off its own switch, so the object it targets is this total, marginal (own-switch) effect. The control-country averages span only the common factor, so the treated country's own income response is idiosyncratic to it, is not netted out, and loads onto the coefficient, which is why the estimator recovers the income-adjusting total, not the income-fixed conditional GE. The border exercise of Section 5.7 recovers that conditional GE separately, as the trade-cost component of the total, and the gap between it and our estimate is the induced income channel. A reduced-form regression cannot strip the multilateral resistances or the own-income response out of the coefficient; only an oracle that offsets $(1 - \sigma) \ln c_i$, $-(1 - \sigma) \ln P_j$, and $\ln E_j$ at their true values recovers the PE of 0.36, which we implement as an infeasible reference.

4.3 Key Results

Table 1 reports the estimator at $N = 60$, $T = 65$, and 100 replications, medians across replications (the per-pair estimates are heavy-tailed), with onset timing tracking the common

factor, which concentrates onsets in the interior of the panel rather than spreading them evenly across it; the specification includes the destination treatment D_{jt} and aggregates by robust Mean Group, as in Section 3.1. The estimand is the total, income-adjusting general-equilibrium effect, whose value here is 0.26; the partial-equilibrium cost parameter is $\theta^{\text{cost}} = 0.36$ and the income-fixed conditional GE is 0.16 (Section 4.2). The table offers four insights.

Table 1: Simulation Results: Two Confounders, Two Controls

specification	bias vs. total GE
<i>Panel A. The two controls do distinct jobs ($s = 0$)</i>	
no control	−0.079
pair trend only	−0.020
control-income CA only	−0.034
control-income CA + pair trend	−0.000
<i>Panel B. Which average (pair trend held fixed throughout)</i>	
income, never-treated only	−0.000
income, all countries	+0.070
bilateral flows, never-treated only	−0.064
bilateral flows, all countries	−0.108

Notes: Median bias across 100 replications, $N = 60$, $T = 65$, $\sigma = 5$, onsets timed on the common factor, measured against the total (income-adjusting) general-equilibrium target of 0.26; the inter-quartile range is ≈ 0.18 , so each cell carries a Monte-Carlo error of about ± 0.02 . *Panel A* varies the controls: neither the pair trend alone (which spans the smooth drift and roll-out) nor the control-income average alone (which spans the stochastic factor) is unbiased; the two together recover the target. *Panel B* holds the trend fixed and varies the averaged observable: income over never-treated countries is the only clean control; averaging income over all countries, or averaging the dependent variable (bilateral flows) at either sample, re-introduces treatment contamination, as equation (6) and the bound (11) anticipate.

First, the two controls do distinct jobs, and both are needed (Panel A). Neither alone is unbiased: the pair-specific trend, which spans the smooth drift and the treatment roll-out, leaves the stochastic factor and a bias of -0.020 ; the control-income average, which spans the stochastic factor, leaves the smooth component and a bias of -0.034 ; with no control the bias is -0.079 . Only the two together recover the target (bias -0.000). This is the identification argument of Section 3.2 made visible: one stochastic factor, spanned by the control average, and one smooth drift, spanned by the deterministic trend.

Second, the average has to be the right one (Panel B), and the choice is settled by the causal hierarchy of Section 3.2. Holding the trend fixed and varying the averaged observable, income over the never-treated countries is the only clean control (the specification of Panel A, bias -0.000). Averaging income over all countries, treated included, contaminates it ($+0.070$), the first-order term of equation (6). And averaging the dependent variable itself, the bilateral flows,

fails at either sample (-0.064 over controls, -0.108 over all countries), because constructed equilibrium objects inherit treatment at first order, not through the small pass-through of the bound (11). Only a country-level income average over never-treated units keeps the control both spanning and uncontaminated.

Third, the partial-equilibrium cost parameter is recoverable only by an infeasible oracle. An analyst who offsets the multilateral resistances at their true values backs out $\theta^{\text{cost}} = 0.36$ exactly; every feasible specification instead targets the total effect of 0.26: the reduced-form coefficient includes the equilibrium response rather than netting it out. The gap between them, which the border exercise of Section 5.7 quantifies in the data, is the general-equilibrium incidence of the treatment.

Fourth, because the estimator is heterogeneous, a separate regression for each pair, then Mean Group aggregation, there are no already-treated-as-control comparisons, and it is immune by construction to the negative-weighting problem of staggered two-way fixed effects (Goodman-Bacon 2021, de Chaisemartin & D’Haultfœuille 2020).

Table 1 holds $s = 0$ throughout, so the control averages span the factor in every row. The next section lets s rise and shows that the ΔCSD diagnostic detects the resulting spanning failure, with power that rises as the averages lose the factor.

4.4 Factor Spanning Diagnostics: ΔCSD

We develop a diagnostic to assess whether Assumption 1 holds in a given application, building on differences in average residual correlations between treated and control samples (ΔCSD). The idea is a simple implication of spanning: if the cross-sectional averages absorb the common factors for treated and control pairs alike, the two sets of residuals have similar cross-sectional dependence (CSD), so their difference ΔCSD is close to zero. If an unspanned factor loads on treated countries only and is correlated with treatment, it generates systematic residual structure among treated pairs that is absent among control pairs, leading to $\Delta\text{CSD} > 0$.

Test statistic. Let $\hat{e}_{ij,t}$ denote the Pearson residual from the pair-level PPML regression (8). Define the mean absolute pairwise correlation of residuals within group g :

$$\overline{|\rho|}_g = \frac{2}{n_g(n_g - 1)} \sum_{\substack{(i,j) < (k,l) \\ (i,j),(k,l) \in g}} |\text{corr}(\hat{e}_{ij,\cdot}, \hat{e}_{kl,\cdot})|, \quad g \in \{\text{treated}, \text{control}\} \quad (15)$$

The CSD diff statistic is then:

$$\Delta\text{CSD} = \overline{|\rho|}_{\text{treated}} - \overline{|\rho|}_{\text{control}} \quad (16)$$

Null hypothesis. Under H_0 Assumption 1 on adequate spanning holds and, when treated and control labels are exchangeable, $\mathbb{E}[\Delta\text{CSD}] = 0$. Under the alternative (unspanned factors load on treated pairs), treated-pair residuals retain systematic factor structure, yielding $\Delta\text{CSD} > 0$. Selecting treated countries on their factor exposure does not by itself raise the statistic: when the averages successfully span the factors, the diagnostic behaves the same way under selection as under random assignment, as the simulation below confirms.

Permutation inference. Exact p -values for the test statistic are obtained by permuting origin treated/control labels $B = 999$ times and recomputing ΔCSD under the null, following the $(b+1)/(B+1)$ formula of Phipson & Smyth (2010). The permutation is performed at the origin level (not the pair level), preserving the within-origin correlation structure and ensuring valid inference despite the dyadic dependence.

Simulation performance. Table 2 evaluates the diagnostic as a test on the conditional-gravity DGP of Section 4.1. Each replication computes ΔCSD on the per-pair PPML residuals of the estimating specification, together with its permutation p -value. Take size first, at $s = 0$, where the averages span the factor. With the treatment effects switched off, so that treated and control pairs are exchangeable, the permutation test is exactly sized: it rejects 5.0 percent of the time under both random and selection assignment (this figure is reported in the note to Table 2). With the treatment effects present, the rejection rate rises to 9–10 percent, similarly under both schemes. This rise reflects the treatment itself: each pair’s treatment dummy removes the average effect, but over a finite sample a small treatment-related component remains in the treatment regression residuals, and the diagnostic, which compares treated with control residual correlation, registers that remainder. Spanning is not the cause, since the averages span at $s = 0$. Because the test is one-sided, its practical effect is to flag a little more readily than a 5 percent test would. Power against a true spanning failure is high and rises quickly: 88 percent already at $s = 0.25$ and close to one for $s \geq 0.5$.

Two features matter in practice. First, the result does not depend on how treatment is assigned. Drawing the treated countries at random and selecting the most factor-exposed countries give

the same size and almost the same power (Table 2). Selection on factor exposure is the case in which an unspanned factor would most threaten identification, and even here the statistic is no larger than under random assignment. Second, the diagnostic depends on the controls in the regression. At a moderate spanning failure ($s = 0.5$) the test rejects almost always when the control averages are included, but only about half the time when they are dropped, because without them the unspanned factor raises residual correlation among treated and control pairs alike and the difference between the two shrinks. A clean pass therefore reflects the specification the estimator actually uses. The diagnostic checks whether spanning holds for the chosen controls; it does not rank one control set against another.

Table 2: Size and Power of the CSD Diff Permutation Test (Conditional-Gravity DGP)

s	0	0.25	0.50	0.75	1
<i>Random assignment (size benchmark)</i>					
ΔCSD (median)	0.010	0.145	0.269	0.336	0.164
rejection rate (%)	10.0	88.0	100.0	100.0	98.0
<i>Selection on factor exposure (the design of Table 1)</i>					
ΔCSD (median)	0.016	0.152	0.275	0.333	0.167
rejection rate (%)	9.0	89.0	100.0	100.0	98.0

Notes: Conditional-gravity DGP ($N = 60$, $T = 65$, 100 replications), onset dates drawn uniformly over the sample interior under random assignment, and timed on the common factor, which concentrates them, under selection. ΔCSD , equation (16), is computed on per-pair PPML Pearson residuals of the estimating specification (income CA and pair trend); rejection rates are at the 5% level, using one-sided p -values from 999 origin-label permutations. Random assignment draws treated countries at random, so labels are exchangeable under the null and the rejection rate at $s = 0$ is the test’s size. Selection on factor exposure uses an alternative DGP, where countries most exposed to the confounder are treated: this is the design of the simulations in Table 1. With the treatment effects removed the test rejects 5.0% at $s = 0$ under both schemes, so the 9–10% recorded with treatment present comes from the treatment rather than from a size distortion. Dropping the control averages from the regression lowers the power at $s = 0.5$ from 100% to 44–53%, so the diagnostic should be computed on the residuals from the specification the estimator uses.

4.5 Treatment Reversal and Carryover

Most DID estimators assume an absorbing treatment, with the notable exceptions of [Liu et al. \(2024\)](#) and [de Chaisemartin & D’Haultfoeuille \(2026\)](#). Yet political treatments reverse, countries which experienced democratic regime change at times revert to autocracy,⁷ and so we ask what happens when the treatment indicator switches on and off, on the same conditional-gravity DGP as above ($N = 60$, $T = 65$). In a *pure cycling* variant, half of the treated countries cycle

⁷Naturally, it is also of interest to study the treatment effects of temporary, ‘episodal’ treatments, such as democratic backsliding (e.g. [Boese-Schlosser et al. 2025](#)).

through on/off/on episodes and the primitives (the cost cut and the income boost of treatment) genuinely revert when $D_{it} = 0$, with the multilateral resistances re-solved each period. In the *carryover* variants, the same cycling indicator is observed, but the effective treatment persists after a reversal, decaying at rate ρ : trade costs and income respond to the persistent effective treatment while the estimator only sees the binary D_{it} .

Table 3: Treatment Reversal and Carryover (Conditional-Gravity DGP)

	(1) income+trend	(2) true effect
pure cycling (effects revert, $\rho = 0$)	0.231	0.250
carryover $\rho = 0.3$	0.228	0.222
carryover $\rho = 0.7$	0.215	0.228
carryover $\rho = 1.0$ (full persistence)	0.154	0.255

Notes: Median estimates, $N = 60$, 100 replications. Conditional-gravity DGP with half of the treated countries cycling (two additional transitions, twelve-year episodes). The true effect is the exact counterfactual log trade change averaged over periods with positive effective treatment. Under full persistence the binary indicator is decoupled from the effective treatment and the estimate is biased towards zero.

Table 3 shows two things. First, cycling on its own is harmless: when effects genuinely revert after treatment ends, the binary indicator tracks the truth, the on/off contrasts add within-pair variation, and the estimator stays near the target (0.231, against a true cycling effect of 0.250). The low-persistence cells ($\rho \leq 0.7$) are all near-unbiased, their deviations from the truth small, inconsistent in sign, and within Monte Carlo error, an order of magnitude below the full-persistence bias. Only strong persistence biases the estimator systematically: at full persistence ($\rho = 1$) the indicator becomes uninformative about the effective treatment and the estimate drops to 0.154 against a true parameter of 0.255, biased towards zero by about two-fifths. The choice between coding the treatment as its current state (incidence) and as an absorbing onset therefore matters, and the simulation results illustrate how: once effects persist through reversals, the toggling indicator under-estimates the effect and an absorbing onset is the appropriate coding. Section 5.6 returns to this choice in the real data.

5 Application: Democracy and Bilateral Exports

5.1 Background

Democratic regime change is a prominent example of a monadic treatment with expected effects on bilateral flows. Within the political theory of trade, a ‘direct democracy’ model suggests that democratic regimes act, under majority rule, in agreement with the trade preferences of the median voter (Mayer 1984, Dutt & Mitra 2002). In a standard Heckscher-Ohlin framework, assuming that countries undergoing democratisation are labour-abundant, capital-poor workers gain from trade liberalisation; extension of the franchise implies that the median voter is a worker, and therefore greater trade openness is adopted by the newly elected government. A ‘political support’ or ‘protection for sale’ approach (Grossman & Helpman 1994, Rodrik 1995) would also allow existing interest groups to lobby for their preferred trade policies, but their political influence is likely to be substantially eroded in the newly democratised regime (Milner & Mukherjee 2009).

In addition, democratisation can have indirect economic effects by promoting legal and economic reforms which support all businesses (Rode & Gwartney 2012, Giuliano et al. 2013). The indirect income channel, whereby democracy causes economic growth (Acemoglu et al. 2019), generates higher bilateral flows through the gravity mechanism. Democracy’s growth effect emerges in stages (Boese-Schlosser & Eberhardt 2024), suggesting dynamic treatment effects that our event-study approach is well suited to capture.

The empirical literature on the effects of democratisation on trade is limited and typically deviates from the prescribed model specification of structural gravity. It generally reports a positive association, though what is measured varies: some studies concern trade policy and openness (Milner & Kubota 2005, Tavares 2008), others bilateral flows (Kono 2008, Aidt & Gassebner 2010, Yu 2010), and Eichengreen & Leblang (2008) document a two-way relationship between democracy and trade, a reverse-causality concern we return to in Section 5.5. We provide among the first credible causal estimates of the monadic effect on bilateral flows, by addressing the identification challenges of monadic treatments in dyadic data.

5.2 Data and Treatment

Trade data and estimator. We use the Head–Mayer trade matrix (Head & Mayer 2021), bilateral goods trade for 208 origins over 1960–2018, which includes domestic (internal) trade

flows. Our estimator uses international flows only; the domestic rows are needed only by the border benchmark exercise in Section 5.7. Flows enter in levels, including zeros, via Poisson pseudo-maximum likelihood (PPML) (Santos Silva & Tenreyro 2006). After requiring sufficient within-pair variation, the estimation sample contains about 27,000 international pairs. The specification follows Section 3.1: pair-level PPML of

$$\mathbb{E}[Y_{ij,t} \mid \cdot] = \exp(\alpha_{ij} + \theta_{ij} \cdot D_{it} + \delta_{ij} \cdot D_{jt} + \kappa_{ij} t + \mathbf{X}'_{ij,t} \boldsymbol{\beta}_{ij} + \phi'_{ij} \bar{\mathbf{z}}_t^{\text{ctrl}}), \quad (17)$$

with $\bar{\mathbf{z}}_t^{\text{ctrl}}$ a control-country cross-sectional average of log GDP and a pair-specific linear trend $\kappa_{ij} t$ (i.e. income+trend). We also include dyadic controls $\mathbf{X}$ for free trade agreements (FTAs) and currency unions (CUs).

The pair-specific treatment effects are aggregated by robust Mean Group regression (rreg), with two-way Cameron et al. (2011) (CGM) standard errors using $t(\min(G_{ex}, G_{im}) - 1)$ critical values. Because the treated-exporter dimension is modest (roughly 40 to 70 clusters depending on the measure), we verify all baseline results against the more conservative exporter cluster bootstrap (999 draws); we report the CGM standard errors throughout; because the bootstrap standard errors are larger by a quarter to a half, requiring significance to survive them is the more demanding test. However, every baseline result passes this challenge.⁸

Democracy measures. We use four measures of democratisation. (i) V-Dem’s *liberal democracy* (Coppedge et al. 2021): a binary indicator for the liberal democracy index crossing a fixed threshold of 0.30, the cross-country mean.⁹ (ii) V-Dem’s *Regimes of the World* (RoW): V-Dem’s validated four-category regime classification (Lührmann et al. 2018), with democracy defined as electoral democracy or above ($v2x_regime \geq 2$), a multi-criteria decision rule (*de facto* multiparty, free and fair elections) rather than an index cutoff. (iii) V-Dem’s *Episodes of Regime Transformation* (ERT): episode-validated democratisation (Maerz et al. 2024), dated at the start of the democratisation episode, on average two years before the regime-change crossing itself, so an ERT onset marks the beginning of a sustained liberalisation episode rather than its outcome. (iv) *Treisman by-mistake*: democratisations coded by Treisman (2020) as unintended by the incumbent, so their timing is set by the incumbent’s misjudgement rather than by a democratic outcome; we use his vetted set with a ‘mistake rating’ of at least four

⁸The exporter cluster bootstrap standard errors are 0.065, 0.071, 0.090, and 0.098 for libdem, RoW, ERT, and Treisman r45, against CGM two-way SEs of 0.053, 0.056, 0.065, and 0.067; all remain statistically significant.

⁹Not to be confused with the highest category of the RoW classification, also termed ‘liberal democracy’.

out of five (labelled r45). Both the countries and the dates are Treisman’s own: the onset is the year of the episode he identifies, which is read from his case histories and makes no use of V-Dem. His dating agrees with the V-Dem codings about as often as they agree with one another (median gap of one to two years; correlations of 0.65 to 0.80, against 0.64 to 0.88 among the three V-Dem measures), so the four measures date the same transitions independently rather than in lockstep; Table A-1 lists every country and year. The three V-Dem measures date democratisation from institutional criteria; Treisman dates it from the incumbent’s misjudgement rather than from a democratic outcome, which makes it a useful cross-check; it is less exposed to the reverse-causality reading than the outcome-based measures, though we do not treat it as a clean natural experiment.

Coding. We code treatment as an absorbing onset, switched on from the first in-sample democratisation. Because (as we will see) the trade gains persist through later reversals, a country that reverts to autocracy keeps most of its gains, so a coding that switched treatment off on reversal would *understate* the effect (Section 5.6). Three of the four measures (libdem, RoW, ERT) share the V-Dem coding backbone and differ in threshold, decision rule, and episode validation, so what they deliver is robustness to how democratisation is coded and dated within one data source rather than across several. Treisman is constructed independently of V-Dem, in both the countries he selects and the years he assigns, and so supplies a comparison that does not share that backbone. The four accordingly span two kinds of variation, and the spread of their estimates in Section 5.3 reflects both.

Coverage. Table 4 reports how each measure’s democratisations translate into estimation coverage. The measures define between 53 and 115 onsets, but only a subset can identify anything. A country enters the treated group only if it is an autocracy in its first observed trade year, and its onset is then its first democratisation after that year. A country already democratic when the panel opens is, under the absorbing coding above, already treated: D_{it} never varies and no difference-in-differences comparison exists for it. This is what separates columns (2) and (3), and it falls hardest on RoW and ERT, whose codings run back to 1900 and so record many democratisations that *predate* our trade data by decades. It also excludes the countries that were democracies in 1960, reverted, and later democratised again, such as Chile in 1990 and Uruguay in 1985: the absorbing coding treats their trade gains as already banked, so their second transition is not a fresh onset. Of the onsets that survive, the estimator uses every one.

Column (4) equals column (3) for all four measures, so the per-pair requirements, at least 12 observations and within-pair variation in D_{it} , drop out no treated exporter, and the ATET is not identified off a filtered subset of transitions.

Table 4: Democratisation Onsets and Estimation Coverage

measure	(1) onsets	(2) in trade panel	(3) identifiable	(4) exporters (G_{ex})	(5) pairs
V-Dem libdem	87	87	72	72	7,949
V-Dem RoW	115	108	70	70	7,873
V-Dem ERT	110	104	63	63	6,775
Treisman r45	53	53	43	43	4,822

Notes: Column (1) counts the democratisation onsets each measure defines. Column (2) restricts to onsets whose country appears as an origin in the estimation sample (197 origins, 1960–2018). Column (3) further requires the country to be an autocracy ($v2x_regime < 2$) in its first observed trade year, with the onset dated at its first democratisation thereafter; a country already democratic when its trade data begin has $D_{it} = 1$ throughout and admits no treatment variation. Regime status is taken from $v2x_regime$ for all four measures, so the at-risk set and the control pool are common across them. Column (4) is the number of treated exporters contributing at least one pair estimate to the Mean Group, the G_{ex} used for inference and the figure reported as N_{tr} in Tables 5 and 8; it equals column (3) for every measure. Column (5) counts the estimated pairs, which meet the requirements of at least 12 observations and within-pair variation in D_{it} .

5.3 Average Treatment Effects

Table 5 reports the estimator on all four measures, each coded as an absorbing treatment onset, for all onset exporters.¹⁰ Every measure returns a positive effect significant at the 1% level, and the four estimates lie between 0.17 and 0.29 log points. We do not read the four as interchangeable. They disagree by a factor of about 1.7, which is what one should expect of codings that differ in the threshold applied (libdem’s index cutoff), in the decision rule (RoW’s multi-criteria classification), in episode validation and the resulting earlier dating (ERT), and in selection and dating together (Treisman, who picks the democratisations he judges unintended and dates them from his own case histories, with no V-Dem input). What is common to all four is the sign, the significance, and the order of magnitude; what varies is how large the effect is, and the spread bounds how much that answer depends on coding choices the literature makes

¹⁰Dropping both the income average and the pair-wise trend inflates the estimate to +1.5–1.7 log points for every democracy measure: with no factor control the per-pair regressions attribute the secular growth of world trade, or a country-specific income growth trend, to democratisation. Either control alone nets most of this out and leaves every measure positive; for RoW the pair trend alone gives +0.11, the income average alone +0.24, and the two together the reported +0.17.

differently. No coding overturns the result. Restricting to the countries that never revert to autocracy leaves the estimates larger, though less precisely estimated, as Section 5.6 shows.

Table 5: Democracy and Trade, by Democracy Measure

measure	construction	(1) ATET	(2) N_{tr}	(3) ΔCSD	(4) $\text{mean} \rho $
V-Dem libdem	index cutoff	0.221*** (0.053)	72	−0.020	0.173
V-Dem RoW	validated categorical	0.169*** (0.056)	70	−0.020	0.176
Treisman r45	by mistake	0.290*** (0.067)	43	−0.025	0.182
V-Dem ERT	validated episode	0.262*** (0.065)	63	−0.007	0.184

Notes: Head–Mayer data, income+trend specification, robust Mean-Group (rreg) aggregation on the full (unbalanced) sample, in contrast to the balanced $[-5, 15]$ window of the event study in Table 6; CGM two-way clustered standard errors in parentheses, with $t(\min(G) - 1)$ critical values; ** $p < 0.05$, *** $p < 0.01$. ‘ N_{tr} ’ reports the number of treated exporters contributing to the Mean Group, the G_{ex} on which inference is based; Table 4 reports the onset counts these are drawn from. Columns (3)–(4) report the spanning diagnostic of Section 4.4 on the per-pair residuals: ΔCSD is the treated-minus-control difference in mean absolute pairwise residual correlation and $\text{mean}|\rho|$ is the treated-group value (the control-group value is ΔCSD below it, near 0.20 throughout). $\Delta\text{CSD} < 0$ for every measure, so the one-sided permutation test never rejects the spanning null (999 exporter-label draws).

5.4 Diagnostics

Factor spanning. We compute the ΔCSD diagnostic of Section 4.4 on the per-pair residuals of the estimating specification. For every measure, the test never rejects, suggesting the absence of a treated-specific unspanned factor. These statistics appear alongside the estimates in Table 5, columns (4)–(5).

5.5 Dynamic Treatment Effects

Table 6 provides an event study. It reports the CCE-DID estimator on a balanced $[-5, 15]$ window for the four measures, with a three-year pre-trend bin $[-5, -3]$ tested against the reference $[-2, -1]$ (the five-year pre-window split into a three-year bin and a two-year reference) and two near-decadal post-treatment bins, alongside the ATET on the full sample and the balanced panel. The post-treatment effects are similar across measures: each gives a significant rise of between 0.15 and 0.25 log points after democratisation, and for the three codings that date the regime change (libdem, RoW and Treisman), the two post-treatment decades are close to

each other. We estimate the event study on the balanced sample because fitting many event-time bins on a changing set of pairs distorts the comparison across bins; the single-coefficient static ATET is not affected, so Table 5 above uses the full sample, reprinted in column (4). The two static ATET columns differ by up to 0.05 log points, in both directions across measures, which reflects which onsets survive the balanced window rather than a change in the effect itself. The post-treatment path is flat rather than still rising, so the effect is a realised level rather than an increment climbing towards a larger long-run value.

Table 6: The Effect of Democratisation on Trade: CCE-MG Event Study

	(1) pre-trend	(2) post-treatment	(3)	(4) static ATET	(5)
democracy measure (onset)	$[-5, -3]$	$[0, 7]$	$[8, 15]$	full	balanced
V-Dem liberal democracy	-0.131***	+0.161***	+0.145**	+0.221***	+0.172***
V-Dem RoW electoral democracy	-0.129***	+0.190***	+0.193***	+0.169***	+0.188***
Treisman r45 ('by mistake')	-0.127**	+0.247***	+0.215***	+0.290***	+0.268***
V-Dem ERT transition start (<i>early onset</i>)	-0.056	+0.218***	+0.155**	+0.262***	+0.226***

Notes: Per-pair het-DID PPML (control-country income cross-sectional averages + pair trend), robust Mean-Group aggregation; balanced $[-5, 15]$ event window, reference $[-2, -1]$; static ATET is the single post-onset indicator on the full sample and on the balanced window. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. ERT dates the onset at the start of the democratisation episode, about two years before the other three codings date the regime change; Section 5.5 reads the resulting flat ERT pre-trend, and the run-up in the other three, as evidence on when the effect begins.

The run-up. A run-up precedes onset in three of the four codings, except for the ERT measure, which dates the beginning of the democratisation process ('episode') on average two years before the regime change itself. This suggests that dating the democratic transition matters, as shown in Table 7. Holding the countries and the selection rule fixed and mechanically moving the onset two years earlier flattens the run-up and raises the average treatment effect for every coding that dates the regime change: RoW's pre-coefficient falls from -0.129 to -0.073 and loses significance, as does Treisman's which moves from -0.127 to -0.054 . The crude libdem index cutoff, in contrast, remains significant at the 5% level at -0.099 . The ATETs for the re-coding rise by 0.06 to 0.10 log points, the largest rise being libdem's (0.221 to 0.324). ERT, which already dates the episode start about two years early, barely moves.

The run-up therefore reflects the effect beginning before the coded onset, not a pre-existing confound. It sits in the pre-window because the outcome-based codings date the transition at its

completion, and a post coefficient cannot count what the coding has placed before the onset.

Table 7: Dating the Transition: Moving the Onset Two Years Earlier

democracy measure (onset)	pre-trend $[-5, -3]$		static ATET	
	as coded	2 years earlier	as coded	2 years earlier
V-Dem liberal democracy	-0.131*** (0.043)	-0.099** (0.038)	+0.221*** (0.053)	+0.324*** (0.057)
V-Dem RoW electoral democracy	-0.129*** (0.045)	-0.073 (0.044)	+0.169*** (0.056)	+0.225*** (0.061)
Treisman r45 ('by mistake')	-0.127** (0.054)	-0.054 (0.047)	+0.290*** (0.067)	+0.355*** (0.068)
V-Dem ERT transition start (<i>early onset</i>)	-0.056 (0.037)	-0.041 (0.044)	+0.262*** (0.065)	+0.286*** (0.072)

Notes: Each measure's own onsets, re-dated two years earlier. The at-risk set is determined on the original onset year, so the countries and the selection rule are held fixed and only the dating moves; the shift nonetheless pushes one to four countries out of the estimation window (treated exporters 72/70/63/43 as coded against 70/66/60/42 shifted). Pre-trend from the balanced $[-5, 15]$ event study, static ATET from the full sample, as in Tables 5 and 6; CGM two-way clustered standard errors in parentheses with $t(\min(G) - 1)$ critical values. ** $p < 0.05$, *** $p < 0.01$. Moving the date earlier flattens the run-up and raises the effect for the three codings that date the regime change, and leaves ERT, which already dates the episode start about two years early, close to where it was.

5.6 Onset Coding and Reversals

Democracy can reverse, which poses a coding choice: is treatment the absorbing fact of having democratised (an onset), or the country's current democratic status, switching off when it reverts? We code regime change as an absorbing onset, one from the first in-sample democratisation, and the carryover simulation of Section 4.5 shows why. Because the trade flows built up after democratisation persist through later political reversals, a country that returns to autocracy keeps most of its gains; coding its post-reversal years as untreated would discard treatment that is still active and bias the estimate towards zero.

Reversals are common but do not drive our results. Table 8 restricts each measure to the origins still meeting the democracy criterion in their last year of trade data (no reversal) and re-estimates the treatment effect. Between a fifth and a third of the treated exporters revert, and dropping them raises the estimate for every measure, but merely by 0.02 to 0.05 log points. Precision falls with the smaller treated sample, since the standard error is driven by the number of exporter clusters, so the effects are larger but less precisely estimated; all four remain significant at the 1% level. Coding reverters as permanently treated therefore does not inflate the onset estimate; dropping them raises it, so holding reverters treated makes the reported effect if

anything conservative. The small size of the change is consistent with the carryover simulation, in which the trade gains largely survive reversal.

Table 8: Reversal Robustness: The Onset Estimate Is Not Driven by the Reverters

democracy measure (onset)	(1) full sample	(2) never reverts	(3) N_{tr}
V-Dem liberal democracy	+0.221*** (0.053)	+0.271*** (0.065)	72/48
V-Dem RoW electoral democracy	+0.169*** (0.056)	+0.191*** (0.061)	70/54
Treisman r45	+0.290*** (0.067)	+0.328*** (0.079)	43/30
V-Dem ERT	+0.262*** (0.065)	+0.286*** (0.075)	63/49

Notes: Static ATET (Head–Mayer data, income+trend specification, robust Mean-Group aggregation, CGM two-way clustered standard errors with $t(\min G - 1)$ critical values); *** $p < 0.01$. Column (1) is the full onset sample, as in Table 5; column (2) restricts to origins still meeting the democracy criterion in their last year of trade data. N_{tr} reports treated exporters contributing to the Mean Group (full/never-reverts); see Table 4. Dropping the reverters raises the effect for every measure while widening its standard error, since the CGM standard error depends on the number of exporter clusters and between a fifth and a third of them are lost; all four remain significant at the 1% level. The onset estimate is therefore not driven by coding reverters as permanently treated.

5.7 The Border Benchmark: Decomposing the Total Effect

What object do these estimates measure? The answer from our simulation in Section 4.2, that the estimator recovers the general-equilibrium (GE) effect of a country’s own switch rather than the partial-equilibrium cost parameter, can be taken to the data, because the Head–Mayer trade matrix contains domestic (internal) trade and therefore supports the one empirical design that can identify the partial effect directly.

The border exercise. With domestic flows, the [Heid et al. \(2021\)](#) border regression estimates

$$\mathbb{E}[Y_{ij,t} \mid \cdot] = \exp(a_{it} + g_{jt} + m_{ij} + \beta [D_{it} \times \text{border}_{ij}]), \quad (18)$$

where border_{ij} indicates international flows. The directional-time fixed effects absorb the exporter’s entire trajectory, including the multilateral resistances and any income response, so β is a partial-equilibrium object: the effect of democratisation D on international relative to domestic trade, holding the equilibrium fixed. It requires domestic trade data, which is what

the general monadic problem commonly lacks; here it serves as a benchmark. We estimate it in two ways. The first is a simple pooled two-way fixed-effects het-DID, which we label the *clean pooled TWFE*: the estimation sample holds only democratisers and countries that are autocratic in every year of the sample, so the untreated comparison group consists of *bona fide* always-autocracies rather than partial democracies or reverts. ‘Clean’ refers to this comparison group, not to the weighting: not-yet-treated democratisers still enter the pooled comparison, and with staggered adoption and heterogeneous effects pooled TWFE combines cohorts with weights that need not be non-negative (de Chaisemartin & D’Haultfœuille 2020, Goodman-Bacon 2021). The second estimator addresses exactly that, the cohort-robust extended two-way fixed-effects (ETWFE) estimator of Wooldridge (2023), which we implement for the Poisson case and which aggregates cohort-specific effects with non-negative weights. We again adopt a balanced panel.

Table 9: The Pooled Border is Sensitive to the Secular Control; the CCE-MG is Not

measure	estimator	(1)	(2)	(3)	(4)	(5)	(6)
		no control		+ border×trend		+ decade×border	
		pre	post	pre	post	pre	post
V-Dem libdem	TWFE	−0.10	+0.84***	+0.01	+0.27***	−0.04	+0.11
	ETWFE	−0.10**	+0.81***	+0.02	+0.25***	−0.05	+0.07
V-Dem RoW	TWFE	−0.01	+0.85***	+0.09*	+0.36***	+0.01	+0.12
	ETWFE	−0.02	+0.83***	+0.09*	+0.34***	−0.01	+0.08
Treisman r45	TWFE	−0.24***	+1.00***	−0.13*	+0.52***	−0.10	+0.39***
	ETWFE	−0.22***	+0.95***	−0.11**	+0.47***	−0.08	+0.30***
V-Dem ERT	TWFE	−0.01	+0.79***	+0.11**	+0.33***	+0.21***	+0.17*
	ETWFE	−0.03	+0.83***	+0.08*	+0.38***	+0.16***	+0.22**

Notes: Pooled border event-study coefficients (Heid–Larch–Yotov international-vs-domestic design), balanced $[-5, 15]$ window, pre-bin $[-5, -3]$ and post-bin $[0, 15]$ against reference $[-2, -1]$; simple TWFE and cohort-robust ETWFE (Wooldridge 2023); three secular controls of increasing flexibility (pair-clustered standard errors). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The border post effect is large and positive for every measure with no control (+0.81 to +1.00) and shrinks by a factor of two to ten as the secular control tightens (over-netting, because the democratisation effect and globalisation unfold over the same decades) while staying positive throughout; under the most aggressive decade×border control it remains significant only for the ERT and Treisman codings. The CCE-MG estimator (Table 6) requires no secular control, since factor spanning absorbs the common trend, and is not subject to this attenuation.

Results That the two estimators agree closely throughout Table 9 indicates the cohort weighting is not distorting the results. Without a secular border control the border effect is large and positive for every measure (+0.81 to +1.00 log points). Because international trade

rose relative to domestic trade for all countries over this period, the coefficient shrinks as this common decade $\times$ border trend is netted, to $+0.07$ – $+0.30$; it stays positive throughout, and significant for the ERT and Treisman codings. The border therefore corroborates a positive effect but is sensitive to how the common globalisation trend is netted, because the democratisation effect and globalisation unfold over the same decades and a single common control cannot fully separate them. Its pre-trends behave the same way. Across the four measures and three secular controls they range from -0.22 to $+0.22$, changing sign and significance from cell to cell, with no pattern across either dimension, including for the early-onset ERT measure. We read the border as a raw-level corroboration of the direction and rough magnitude of the effect rather than as a precise benchmark.

Decomposing the total effect, structurally. Our estimate is the total GE effect; the border coefficient is the pure partial-equilibrium trade-cost effect. The two are reconciled by exact hat algebra on the same Head–Mayer trade matrix, in two steps that also decompose our estimate. First, we back out the cost shock $\Delta \ln \tau = \beta^{\text{PE}}/(1 - \sigma)$, apply it to the democratisers’ international links, and re-solve the multilateral-resistance system in changes, with incomes held fixed. This maps the border PE into the income-fixed conditional GE, the trade-cost component of the total, which multilateral-resistance feedback pulls well below the PE (by 56 to 59% for every measure). Second, the treated countries’ own induced income gains, fed through the same system, add the income channel on top. The CCE-DID estimator carries both together; the border exercise and the hat algebra separate them. Because a staggered Mean-Group estimator identifies each country off its own switch, every object here is the marginal (own-switch) counterfactual, to match the estimand.

Table 10 reports the structural result, which decomposes our estimate rather than validating a separate object. For RoW, multilateral-resistance feedback pulls the border PE (0.34) down to an income-fixed conditional GE of 0.15, which is already most of the total effect our estimator recovers directly (0.17). The estimator exceeds the conditional GE for all four measures. Read against a benchmark stated at the horizon the estimator measures (0.07–0.10, being three-fifths of the twenty-five-year income gain in [Acemoglu et al. \(2019\)](#) times a marginal income elasticity of 0.57), the libdem, ERT and Treisman channels ($+0.11$, $+0.10$, $+0.10$) fall in the band, while the RoW channel ($+0.02$) falls short of it: RoW combines the smallest total effect with one of the largest border PEs, so its estimate is almost entirely the trade-cost component. RoW is also the one measure whose static ATET does not rise when the estimation window is lengthened,

even though its event-study bins do, which suggests the pair-specific trend absorbs more of its effect than of the others'; its small income channel should be read in that light. The attenuation from PE to conditional GE is 56–59% for every measure and is invariant to the trade elasticity, so it is a property of the trade network rather than a fitted quantity. That two designs built on different variation (the international-versus-domestic border contrast and the trend-controlled CCE-DID) meet at the same total effect once the equilibrium is solved is not a decisive test, since its power against alternatives that predict a similar attenuation is limited. But a pre-trend artefact in our estimator would not, on its own, reproduce a border coefficient obtained from a different design and then decompose into trade-cost and income components of the right structural magnitudes.

Table 10: Decomposing the Total Effect: Border PE, Conditional GE, and the Income Channel

	V-Dem libdem	V-Dem RoW	V-Dem ERT	Treisman r45
border PE (marginal; MR & income fixed)	+0.25	+0.34	+0.38	+0.47
conditional GE (MR adjusts; income <i>fixed</i>)	+0.11	+0.15	+0.16	+0.19
CCE-DID estimator = total GE	+0.22	+0.17	+0.26	+0.29
implied income channel (total – conditional)	+0.11	+0.02	+0.10	+0.10

Notes: Every object is the marginal (own-switch) counterfactual, to match the staggered Mean-Group estimand. All border PE and CCE-DID estimates are statistically significant at the 1% level. The conditional GE is computed from the border PE by exact hat algebra on the Head–Mayer matrix (cost shock $\beta^{\text{PE}}/(1 - \sigma)$ applied to the democratiser's international links, multilateral resistances re-solved, income fixed); the attenuation from the PE is invariant to the trade elasticity (identical for $\sigma \in \{4, 5, 6.03\}$), set by the treated origin's international trade share. The CCE-DID coefficient is always the static ATET of Table 5, so the decomposition applies to the estimates the paper reports rather than to a separate quantity; it is the total GE and includes the treated country's own induced income response, and it exceeds the conditional GE for every measure. The structural benchmark must be stated at the horizon the estimator measures: Acemoglu et al. (2019) report a 0.20–0.30 log income gain in the long run, which they date at roughly twenty-five years, so the first fifteen years carry about three-fifths of it, and a marginal income elasticity of ≈ 0.57 implies a channel of roughly 0.07–0.10 rather than the 0.11–0.18 the long-run figure would give. The border PE is estimated on the balanced $[-5, 15]$ window while the total is the full-sample ATET, so the two ends of the decomposition are not matched on horizon; on a matched balanced window the RoW channel is +0.04 and the other three +0.06 to +0.07. The border PE uses a linear border $\times$ trend secular control aggregated cohort-robustly (ETWFE); the more aggressive decade $\times$ border control yields a lower bound of 0.07–0.30 on the same estimator. The conditional GE and the income channel carry no standard errors here, but they are not free of sampling error: the hat algebra is a deterministic map applied to the *estimated* border PE, so the conditional GE inherits that coefficient's uncertainty, and the income channel, being the difference between the estimated total GE and that conditional GE, inherits both. The decomposition should be read as a point accounting of where the estimate comes from, not as a set of separately testable quantities.

The income channel. The income channel in Table 10 is a structural back-of-the-envelope calculation, and it is already inside our estimate rather than an increment to it. Acemoglu

et al. (2019) put the income gain of democracy at +0.20 to +0.30 log points over roughly two decades. Feeding such a gain through the trade matrix raises international exports less than one-for-one: total sales scale with income by accounting identity, but a part accrues to domestic trade, and the marginal income elasticity of international exports (about 0.57 on these data, in the range of standard structural-gravity values) is below the unit elasticity of total sales. That figure is a long-run one, however, and our static coefficient is dominated by the first fifteen years after onset, which carry roughly three-fifths of it. The horizon-matched channel is then about $0.57 \times (0.12-0.18) = 0.07-0.10$, which is the order of the gap between our estimate and the conditional GE for libdem, ERT and Treisman (+0.11, +0.10, +0.10); the RoW gap (+0.02) is smaller. This elasticity is distinct from the multilateral-resistance attenuation in the first step of the decomposition, which is set by the trade share; the two should not be conflated. Because our static coefficient spans the post-onset horizon over which income adjusts, it already reflects this channel in full. The fully income-adjusted total is what the estimator recovers (0.17 to 0.29 across the measures), not a separate larger number. Both the conditional GE and this total lie well below the reduced-form event-study path at long horizons, which is inflated by survivorship, since only the democracies that endure are observed decades after onset.

6 Concluding Remarks

A large class of policy-relevant questions asks how a country-level treatment affects its bilateral flows with all partners: does democratisation expand trade, do capital controls redirect investment, does emigration reform reshape a country’s migration patterns? These questions have remained difficult to answer credibly because the standard econometric toolkit for bilateral flows cannot identify monadic treatment effects, and the recent het-DID revolution has not addressed the dyadic setting. Our paper fills this gap.

The CCE-DID estimator resolves the identification challenges that arise when a monadic treatment meets dyadic data: collinearity with directional-time fixed effects, mediator bias from network-wide equilibrium propagation, and heterogeneous pair-level shock exposure. Cross-sectional averages of fundamental variables from never-treated countries, entered with pair-specific coefficients alongside pairwise linear trend terms, model the network’s common-factor structure in the tradition of factor models for dyadic data (Chen et al. 2021); these coefficients absorb the heterogeneous factor loadings that the network structure generates, and Mean Group aggregation recovers the ATET without negative weighting. The identifying

assumption, factor spanning, is weaker than parallel trends in permitting heterogeneous factor loadings, at the cost of requiring observable averages that span the factor space, and is partially testable through the CSD diff symptom check we introduce.

Our application demonstrates the estimator's value and decomposes its estimand in the data. Cleanly coded democratisation onset raises bilateral exports by 0.17–0.29 log points across four alternative democracy measures, an effect existing methods cannot recover due to the constraints of the empirical gravity model. This is the total general-equilibrium effect of a country's own democratisation, inclusive of the induced income response. Because the trade data we use include domestic flows, a border exercise identifies the partial-equilibrium trade-cost effect independently, and exact hat algebra decomposes our estimate: multilateral-resistance feedback maps that partial effect down to an income-fixed conditional GE (a 56 to 59% attenuation, set by the trade network rather than the trade elasticity), and the treated countries' own income gains add the remainder, reconstructing what the estimator recovers. Income is thus already part of what we estimate. Fed through the same structural map at the horizon our estimator measures, the [Acemoglu et al. \(2019\)](#) income gains account for about +0.10 to +0.11 log points of the estimate for the libdem, ERT and Treisman codings (and +0.02 for RoW), linking the democracy-growth and democracy-trade literatures.

Several directions for future work present themselves: extension to continuous treatments, adaptation to hybrid monadic-dyadic treatment structures (such as unilateral sanctions), formal asymptotic theory for the CCE-DID estimator in the dyadic setting, and structural interpretation of the reduced-form estimates through sufficient statistics. The requirement of a sufficient pool of never-treated countries is a practical limitation; in applications where few countries remain untreated, alternative approaches using not-yet-treated countries with appropriate bias corrections warrant closer investigation.

The CCE-DID estimator is broadly applicable to any setting where a country-level binary treatment affects bilateral flows. By providing both the estimator and accompanying diagnostics, we offer a practical toolkit that enables credible causal inference for reduced form monadic treatment effects in dyadic data.

References

- Acemoglu, D., Naidu, S., Restrepo, P. & Robinson, J. A. (2019), ‘Democracy does cause growth’, *Journal of Political Economy* **127**(1), 47–100.
- Aidt, T. S. & Gassebner, M. (2010), ‘Do autocratic states trade less?’, *World Bank Economic Review* **24**(1), 38–76.
- Anderson, J. E. & Van Wincoop, E. (2003), ‘Gravity with gravitas: A solution to the border puzzle’, *American Economic Review* **93**(1), 170–192.
- Arellano, C., Bai, Y. & Lizarazo, S. V. (2017), Sovereign risk contagion, NBER Working Paper Series 24031, National Bureau of Economic Research.
- Arezki, R., Ramey, V. A. & Sheng, L. (2017), ‘News shocks in open economies: Evidence from giant oil discoveries’, *Quarterly Journal of Economics* **132**(1), 103–155.
- Bai, J. (2009), ‘Panel data models with interactive fixed effects’, *Econometrica* **77**(4), 1229–1279.
- Beverelli, C., Keck, A., Larch, M. & Yotov, Y. V. (2024), ‘Institutions, trade, and development: Identifying the impact of country-specific characteristics on international trade’, *Oxford Economic Papers* **76**(2), 469–494.
- Boese-Schlusser, V. A. & Eberhardt, M. (2024), ‘Democracy doesn’t always happen over night: Regime change in stages and economic growth’, *Review of Economics and Statistics* . forthcoming.
- Boese-Schlusser, V., Desbordes, R., Eberhardt, M. & Larch, M. (2025), The economic consequences of democratic backsliding: Evidence from U.S. States, CEPR Discussion Paper DP20874, Centre for Economic Policy Research.
- Boneva, L. & Linton, O. (2017), ‘A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance’, *Journal of Applied Econometrics* **32**(7), 1226–1243.
- Borusyak, K., Jaravel, X. & Spiess, J. (2024), ‘Revisiting event-study designs: Robust and efficient estimation’, *Review of Economic Studies* **91**(6), 3253–3285.

- Callaway, B. & Sant'Anna, P. H. C. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.
- Cameron, A. C., Gelbach, J. B. & Miller, D. L. (2011), 'Robust inference with multiway clustering', *Journal of Business & Economic Statistics* **29**(2), 238–249.
- Cesa-Bianchi, A., Eguren Martin, F. & Thwaites, G. (2019), 'Foreign booms, domestic busts: The global dimension of banking crises', *Journal of Financial Intermediation* **37**, 58–74.
- Chan, M. K. & Kwok, S. S. (2022), 'The PCDID approach: Difference-in-differences when trends are potentially unparallel and stochastic', *Journal of Business and Economic Statistics* **40**(3), 1216–1233.
- Chen, M., Fernández-Val, I. & Weidner, M. (2021), 'Nonlinear factor models for network and panel data', *Journal of Econometrics* **220**(2), 296–324.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J. et al. (2021), 'V-dem dataset v11', *Varieties of Democracy (V-Dem) Project*. Available at <https://www.v-dem.net>.
- Cotet, A. M. & Tsui, K. K. (2013), 'Oil and conflict: What does the cross country evidence really show?', *American Economic Journal: Macroeconomics* **5**(1), 49–80.
- de Chaisemartin, C. & D'Haultfœuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–2996.
- de Chaisemartin, C. & D'Haultfœuille, X. (2026), 'Difference-in-differences estimators of intertemporal treatment effects', *Review of Economics and Statistics* **108**(4), 863–880.
- Dutt, P. & Mitra, D. (2002), 'Endogenous trade policy through majority voting: An empirical investigation', *Journal of International Economics* **58**(1), 107–133.
- Eichengreen, B. & Leblang, D. (2008), 'Democracy and globalization', *Economics and Politics* **20**(3), 289–334.
- Freeman, R., Larch, M., Theodorakopoulos, A. & Yotov, Y. V. (2025), 'Unlocking new methods to estimate country-specific effects and trade elasticities with the structural gravity model', *Journal of Applied Econometrics* **40**(6), 669–684.

- Giuliano, P., Mishra, P. & Spilimbergo, A. (2013), ‘Democracy and reforms: Evidence from a new dataset’, *American Economic Journal: Macroeconomics* **5**(4), 179–204.
- Gobillon, L. & Magnac, T. (2016), ‘Regional policy evaluation: Interactive fixed effects and synthetic controls’, *Review of Economics and Statistics* **98**(3), 535–551.
- Goodman-Bacon, A. (2021), ‘Difference-in-differences with variation in treatment timing’, *Journal of Econometrics* **225**(2), 254–277.
- Graham, B. S. (2020), Network data, in S. N. Durlauf, L. P. Hansen, J. J. Heckman & R. L. Matzkin, eds, ‘Handbook of Econometrics’, Vol. 7A, Elsevier, pp. 111–218.
- Grossman, G. M. & Helpman, E. (1994), ‘Protection for sale’, *American Economic Review* **84**(4), 833–850.
- Gunitsky, S. (2014), ‘From shocks to waves: Hegemonic transitions and democratization in the twentieth century’, *International Organization* **68**(3), 561–597.
- Head, K. & Mayer, T. (2014), Gravity Equations: Workhorse, Toolkit, and Cookbook, in G. Gopinath, E. Helpman & K. Rogoff, eds, ‘Handbook of International Economics’, Vol. 4, Elsevier, chapter 3, pp. 131–195.
- Head, K. & Mayer, T. (2021), ‘The United States of Europe: A gravity model evaluation of the four freedoms’, *Journal of Economic Perspectives* **35**(2), 23–48.
- Heid, B., Larch, M. & Yotov, Y. V. (2021), ‘Estimating the effects of non-discriminatory trade policies within structural gravity models’, *Canadian Journal of Economics* **54**(1), 376–409.
- Juodis, A. & Reese, S. (2026), ‘Five lessons for applied researchers from twenty years of common correlated effects estimation’, *Journal of Econometrics* **253**, 106120.
- Kono, D. Y. (2008), ‘Democracy and trade discrimination’, *Journal of Politics* **70**(4), 942–955.
- Liu, L., Wang, Y. & Xu, Y. (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.
- Lührmann, A., Tannenberg, M. & Lindberg, S. I. (2018), ‘Regimes of the World (RoW): Opening New Avenues for the Comparative Study of Political Regimes’, *Politics and Governance* **6**(1), 60–77.

- Maerz, S. F., Edgell, A. B., Wilson, M. C., Hellmeier, S. & Lindberg, S. I. (2024), 'Episodes of regime transformation', *Journal of Peace Research* **61**(6), 967–984.
- Mayer, W. (1984), 'Endogenous tariff formation', *American Economic Review* **74**(5), 970–985.
- Milner, H. V. & Kubota, K. (2005), 'Why the move to free trade? democracy and trade policy in the developing countries', *International Organization* **59**(1), 107–143.
- Milner, H. V. & Mukherjee, B. (2009), 'Democratization and economic globalization', *Annual Review of Political Science* **12**, 163–181.
- Nagengast, A. J. & Yotov, Y. V. (2025), 'Staggered difference-in-differences in gravity settings: Revisiting the effects of trade agreements', *American Economic Journal: Applied Economics* **17**(1), 271–296.
- Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.
- Phipson, B. & Smyth, G. K. (2010), 'Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn', *Statistical Applications in Genetics and Molecular Biology* **9**(1), Article 39.
- Rode, M. & Gwartney, J. D. (2012), 'Does democratization facilitate economic liberalization?', *European Journal of Political Economy* **28**(4), 607–619.
- Rodrik, D. (1995), Political economy of trade policy, in G. M. Grossman & K. Rogoff, eds, 'Handbook of International Economics', Vol. 3, Elsevier, chapter 28, pp. 1457–1494.
- Roth, J., Sant'Anna, P. H. C., Bilinski, A. & Poe, J. (2023), 'What's trending in difference-in-differences? A synthesis of the recent econometrics literature', *Journal of Econometrics* **235**(2), 2218–2244.
- Santos Silva, J. a. M. & Tenreyro, S. (2006), 'The log of gravity', *The Review of Economics and Statistics* **88**(4), 641–658.
- Sun, L. & Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.
- Tavares, J. (2008), 'Trade, factor proportions, and political rights', *The Review of Economics and Statistics* **90**(1), 163–168.

- Treisman, D. (2020), ‘Democracy by mistake: How the errors of autocrats trigger transitions to freer government’, *American Political Science Review* **114**(3), 792–810.
- Wooldridge, J. M. (2023), ‘Simple approaches to nonlinear difference-in-differences with panel data’, *The Econometrics Journal* **26**(3), C31–C66.
- Wooldridge, J. M. (2025), ‘Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators’, *Empirical Economics* **69**(5), 2545–2587.
- Xu, Y. (2017), ‘Generalized synthetic control method: Causal inference with interactive fixed effects models’, *Political Analysis* **25**(1), 57–76.
- Yu, M. (2010), ‘Trade, democracy, and the gravity equation’, *Journal of Development Economics* **91**(2), 289–300.

Online Appendix – Not Intended for Publication

A Democratisation Dates by Measure

Table A-1 lists, for every country entering the estimation sample, the year each measure dates its democratisation. Blank cells are countries the measure does not code as an at-risk in-sample onset. The three V-Dem codings agree with one another about as often as they agree with Treisman's independent dating, which is what one expects of four readings of the same events: pairwise median gaps run from zero to four years and correlations from 0.64 to 0.88.

Table A-1: Year of Democratisation, by Country and Measure

country	V-Dem libdem	V-Dem RoW	V-Dem ERT	Treisman r45
Albania	1992	2005	1998	1990
Argentina	1964	1964	1963	–
Bangladesh	–	1992	1986	–
Benin	1991	1992	1990	1990
Bhutan	2008	2009	2007	–
Bolivia	1985	1986	1982	1982
Bosnia and Herzegovina	1997	1997	1996	–
Brazil	1986	1987	1979	1979
Bulgaria	1990	1990	1990	1990
Burkina Faso	1999	2000	2016	–
Cape Verde	1991	1991	1980	–
Colombia	1975	1991	1990	–
Comoros	2014	–	–	–
Croatia	2000	2000	–	1999
Cyprus	–	1975	1970	–
Dominican Republic	1982	1982	1978	1978
Ecuador	1980	1980	1978	1979
El Salvador	1999	1999	1992	1979
Fiji	1967	1970	1963	1990
Georgia	2004	2004	2004	2004
Ghana	1970	1980	1992	1969
Greece	1975	1975	1974	–
Guatemala	1999	2000	1984	–
Guinea-Bissau	–	2015	2014	–
Guyana	1993	1998	1986	1992
Honduras	1995	1994	1990	–
Hong Kong	1992	–	–	–
Hungary	1990	1990	1988	1988
Indonesia	1999	1999	1998	–
Ivory Coast	2002	2016	2010	–

Table A-1 continued

country	V-Dem libdem	V-Dem RoW	V-Dem ERT	Treisman r45
Kenya	2003	2003	2010	2002
Kyrgyzstan	2011	–	–	2010
Lebanon	2011	–	–	–
Lesotho	1993	1993	1990	1993
Liberia	2006	2006	2005	2006
Libya	2013	2013	–	2012
Madagascar	1994	1994	1990	1991
Malawi	1995	1995	1992	1993
Maldives	2009	2009	2005	2009
Mali	1992	1993	1992	1991
Mexico	1997	1996	1988	1997
Mongolia	1991	1991	1990	1990
Mozambique	1995	–	–	–
Namibia	1990	1990	–	–
Nepal	1991	2009	2014	1990
Nicaragua	1990	1990	1984	1990
Niger	1993	1993	1993	1991
Nigeria	2000	2012	2010	–
North Macedonia	1995	1998	1994	–
Panama	1991	1991	1979	1989
Papua New Guinea	1972	1973	1970	–
Paraguay	1992	1993	1989	1989
Peru	1964	1981	1978	–
Philippines	1988	1988	1986	1986
Poland	1990	1990	1989	1989
Portugal	1976	1976	1970	1974
Republic of the Congo	1993	–	–	–
Romania	–	1990	1990	–
Russia	1995	–	–	–
Sao Tome and Principe	1991	1991	1987	1991
Senegal	1978	1990	1990	2000
Seychelles	2003	2015	2013	–
Sierra Leone	2003	2003	1999	–
Singapore	2000	–	–	–
Slovakia	–	1994	–	–
Solomon Islands	1978	1979	2007	–
South Africa	1995	1995	1989	1994
South Korea	1988	1988	1981	1987
Spain	1978	1978	1976	1975
Taiwan	1996	1996	1986	1991
Tanzania	1992	–	–	–
Thailand	1993	1998	1992	1992

Table A-1 continued

country	V-Dem libdem	V-Dem RoW	V-Dem ERT	Treisman r45
The Gambia	1967	2018	2017	–
Trinidad and Tobago	–	1962	–	–
Tunisia	2012	2012	2011	2011
Türkiye	1966	1966	1962	–
Ukraine	2006	1994	2002	1994
Vanuatu	–	1980	–	–
Zambia	1992	2007	–	1991
<i>countries coded</i>	72	70	63	43

B Robust Regression Aggregation

Assumption B.1 (Random Coefficients). *The pair-specific treatment effects satisfy*

$$\theta_{ij} = \theta^{ATET} + v_{ij}, \quad \mathbb{E}[v_{ij}] = 0, \quad \mathbb{E}[v_{ij}^2] = \sigma_v^2 < \infty.$$

The distribution of v_{ij} is symmetric about zero, and the estimation errors $\hat{\theta}_{ij} - \theta_{ij}$ are mean-zero conditional on the treatment effects. Two conditions ensure the rreg location targets the equal-weighted θ^{ATET} . First, pair-level estimation precision is independent of θ_{ij} given the factors, so that the biweight’s downweighting of poorly-identified pairs does not tilt the estimand toward precisely-estimated pairs; absent this, rreg targets a precision-weighted ATET. Second, exact centring requires the distribution of $\hat{\theta}_{ij} = \theta^{ATET} + v_{ij} + (\hat{\theta}_{ij} - \theta_{ij})$ to be symmetric, hence symmetry of the estimation error as well as of v_{ij} ; finite- T PPML errors are skewed, so centring holds in the $T \rightarrow \infty$ limit, and finite-sample skew is why we report medians (Section 4.3).

Assumption B.2 (Two-Way Cluster Dependence). *Let $\mathcal{M}_i = \{(i, j) : (i, j) \in \mathcal{M}\}$ and $\mathcal{M}_j = \{(i', j) : (i', j) \in \mathcal{M}\}$ denote the sets of treated pairs sharing exporter i and importer j , respectively, with G treated exporters in total. The pair-level estimation errors $(\hat{\theta}_{ij} - \theta_{ij})$ exhibit two sources of within-sample dependence:*

- *Exporter clustering: pairs sharing exporter i are correlated because they share D_{it} and exporter-level shocks.*
- *Importer clustering: pairs sharing importer j are correlated because they share D_{jt} and importer-level shocks.*

Conditional on $\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_T)'$, pairs that share neither exporter nor importer are independent up to equilibrium spillovers of order $O(N^{-1})$; this weak network dependence does not affect the CLT (cf. [Graham 2020](#), on conditionally independent dyads). The two-way-

clustered estimation errors, rescaled by their $O_p(T^{-1/2})$ rate, satisfy

$$\sqrt{\frac{T}{G}} \sum_{i=1}^G \bar{e}_i \xrightarrow{d} \mathcal{N}(0, \sigma_e^2) \quad \text{as } G, T \rightarrow \infty,$$

where $\bar{e}_i = |\mathcal{M}_i|^{-1} \sum_{(i,j) \in \mathcal{M}_i} (\hat{\theta}_{ij} - \theta_{ij}) = O_p(T^{-1/2})$ and $\sigma_e^2 = \lim \text{Var}(\sqrt{T} \bar{e}_i)$ incorporates both within-exporter and within-importer correlation. The number of partners $J_i = |\mathcal{M}_i|$ satisfies $\sup_i J_i / \bar{J} < \infty$, where $\bar{J} = G^{-1} \sum_i J_i$, so that no single exporter dominates the aggregate.

Remark 1 (Effective Cross-Section). Although the *rreg* estimator operates on $|\mathcal{M}|$ pair-level estimates, the effective cross-section for inference is G (treated exporters), not $|\mathcal{M}|$ (treated pairs). Pairs sharing an exporter are correlated through D_{it} and exporter-level shocks; pairs sharing an importer are correlated through D_{jt} and importer-level shocks. Two-way clustering by exporter and importer ([Cameron et al. 2011](#)) accounts for both channels of dependence. Importer-side clustering does not further reduce the effective dimension below G : the two-way clustered variance decomposes as $\hat{V}_{exp} + \hat{V}_{imp} - \hat{V}_{pair}$, which inflates the variance relative to exporter-only clustering but does not change the number of independent units driving the $\sqrt{G}$ convergence rate. The rate of convergence is therefore $\sqrt{G}$, requiring $G \rightarrow \infty$ and $\sqrt{G}/T \rightarrow 0$.