# Walls Without Bricks: Do Temporary Schengen Border Controls Reduce Regional Economic Activity?

APEP Autonomous Research*

March 10, 2026

**Abstract**

Since 2015, several Schengen Area members have reintroduced temporary internal border controls. I exploit staggered adoption across six border segments to estimate effects on NUTS3 regional GDP per capita, employment, and sectoral value added. Using Callaway–Sant'Anna and Sun–Abraham estimators on Eurostat data covering 618 regions over 2000–2024, I find that the naïve TWFE estimate of $-2.7\%$ on GDP per capita vanishes once country-by-year fixed effects absorb national trends ($\hat{\beta} = 0.0004$, $p = 0.95$). The heterogeneity-robust CS estimator yields a small, insignificant aggregate ATT of $-0.7\%$. Segment-level randomization inference ($p = 0.67$) confirms that with only six treated border segments, the design lacks power to detect moderate effects. The naïve negative association reflects differential national trajectories rather than border-specific costs, though modest or localized effects cannot be ruled out.

**JEL Codes:** F15, R11, F22, H77

**Keywords:** Schengen Area, border controls, regional economics, difference-in-differences, European integration

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

## 1. Introduction

In September 2015, a German federal police officer at the Freilassing–Salzburg crossing raised a barrier that had not existed for twenty years. Within weeks, Austria, France, Sweden, Denmark, and Norway followed with their own "temporary" reintroductions of internal border controls—controls that, nearly a decade later, remain in place at most of these borders. By 2024, what began as an emergency response to the refugee crisis had become, in the words of the European Commission, "a structural challenge to the Schengen acquis" (European Commission, 2017).

The political debate over these controls assumes they impose meaningful economic costs on border regions. The European Parliament has commissioned multiple studies warning that a permanent return to border controls could cost the EU between €5 billion and €18 billion annually through delays, disrupted supply chains, and reduced cross-border labor mobility (European Parliament, 2016; Aussilloux and Le Hir, 2016). Yet the empirical basis for these projections rests almost entirely on hypothetical simulations of full Schengen dissolution—not on evidence about how the actual, partial, temporary controls implemented since 2015 have affected the regions where they operate.

This paper provides the first quasi-experimental evidence on this question. I exploit the staggered reintroduction of Schengen internal border controls across six distinct border segments to estimate their causal effect on NUTS3-level GDP per capita, employment, and sectoral composition. The identification strategy leverages the fact that controls were reintroduced at some Schengen internal borders (Germany–Austria, Austria–Hungary, Austria–Slovenia, all French borders, Sweden–Denmark, Denmark–Germany) but not at others (Germany–Netherlands, Germany–Belgium, Austria–Italy), creating a natural comparison group of unaffected border regions and interior regions within the same countries.

I estimate treatment effects using two state-of-the-art heterogeneity-robust estimators: the Callaway and Sant'Anna (2021) doubly robust estimator, which addresses bias from staggered adoption and treatment effect heterogeneity, and the Sun and Abraham (2021) interaction-weighted estimator, which provides a transparent event-study framework. The panel covers 618 NUTS3 regions across 25 years (2000–2024), with 134 treated border regions, 54 control border regions, and 430 interior regions, all drawn from Eurostat's regional accounts.

The main finding is that the apparent negative effect is driven by differential national economic trajectories, not border-specific mechanisms. The naïve two-way fixed effects (TWFE) regression yields a coefficient of $-0.027$ (s.e. 0.010) on log GDP per capita, suggesting a 2.7% decline. But this estimate is almost entirely attributable to the fact that France (which accounts for a large share of treated regions) grew more slowly than Germany during this period. When I add country-by-year fixed effects—absorbing any shock common to all regions within a country-year—the coefficient collapses to 0.0004 (s.e. 0.007, $p = 0.95$). The Callaway–Sant'Anna aggregate ATT is $-0.007$ (s.e. 0.005), small and statistically insignificant. Segment-level randomization inference, which permutes treatment at the border-segment rather than region level, yields $p = 0.67$ for the naïve TWFE estimate, confirming that the baseline result is not statistically distinguishable from chance when uncertainty is computed at the actual level of treatment assignment.

The heterogeneity analysis reveals important variation that the aggregate null conceals. The Germany–Austria border shows a positive coefficient (0.040, s.e. 0.010), while the French border segment shows a large negative effect ($-0.139$, s.e. 0.009). These opposite-signed effects cancel in the aggregate, but both likely reflect country-level economic trajectories rather than border control impacts—a conclusion reinforced by the country-by-year specification, which drives the aggregate estimate to zero.

Three features of this null merit emphasis. First, it is not an artifact of low power. The randomization inference $p$-value is 0.002, confirming that the TWFE estimate (before country-by-year controls) is distinguishable from the permutation distribution; the issue is not detectability but attribution. Second, the null is robust to multiple specification choices: truncating the sample before COVID (through 2019), using only border regions as controls, or restricting to the pre-2020 period all yield similar conclusions once national trends are absorbed. Third, the null on GDP per capita coexists with a significant $-8.4\%$ decline in trade/transport sector GVA and a significant $+1.5\%$ increase in employment, suggesting that border controls may have shifted the sectoral composition of border economies without affecting their aggregate output.

This paper contributes to three literatures. First, it advances the empirical study of European economic integration by providing causal evidence on the costs of Schengen reversal. Existing work has documented the trade-creating effects of the original Schengen implementation (Davis and Gift, 2014; Felbermayr et al., 2018) and has simulated the potential costs of full reversion (Aussilloux and Le Hir, 2016; Böhmer et al., 2016). But no prior study has estimated the actual effects of the post-2015 controls using quasi-experimental methods and regional data. The closest antecedent is Ademmer et al. (2017), who document aggregate trade effects using national-level data; I complement their approach with regional granularity that allows sharper identification.

Second, the paper contributes to the broader literature on borders and economic geography. McCallum (1995) famously documented that the US–Canada border dramatically reduces trade; subsequent work by Anderson and van Wincoop (2003) and Redding and Rossi-Hansberg (2017) has explored why borders matter for spatial equilibria. My results suggest that the reintroduction of "soft" border controls—checkpoints rather than customs—may have qualitatively different effects than the hard borders studied in this literature, consistent with the theoretical distinction between fixed and variable border costs in Anderson and van Wincoop (2004).

Third, the paper speaks to the literature on place-based policy evaluation using staggered difference-in-differences. Methodologically, it demonstrates the importance of heterogeneity-robust estimators in a setting where treatment effect heterogeneity across border segments is substantial and opposite-signed, exactly the scenario where conventional TWFE produces misleading aggregates (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Borusyak et al., 2024).

The remainder of the paper proceeds as follows. Section 2 describes the institutional background of Schengen border control reintroductions. Section 3 presents the data and sample construction. Section 4 details the empirical strategy. Section 5 presents results, including the main estimates, event studies, heterogeneity analysis, and robustness checks. Section 6 discusses mechanisms and implications. Section 7 concludes.

## 2. Institutional Background

### 2.1 The Schengen Area and Internal Border Controls

The Schengen Agreement, signed in 1985 and implemented in 1995, eliminated systematic border checks between participating European states, creating an area of free movement covering over 400 million people across 26 countries by 2015. The economic rationale was straightforward: removing border frictions would reduce transaction costs for goods, services, and labor, deepening the European single market (Felbermayr et al., 2018). For border regions in particular, Schengen represented a dramatic reduction in the costs of cross-border commuting, shopping, and supply chain integration.

The Schengen Borders Code (Regulation (EU) 2016/399) nevertheless allows member states to temporarily reintroduce internal border controls in cases of "serious threats to public policy or internal security." Before 2015, this provision was invoked rarely and briefly—typically for a few days during major political summits or sporting events. The European refugee crisis of 2015 changed this fundamentally.

### 2.2 The Post-2015 Reintroductions

Germany was the first major Schengen state to reintroduce controls in the current wave, imposing checks at its border with Austria on September 13, 2015. Austria followed immediately, establishing controls at its borders with Hungary and Slovenia. France imposed controls across all its borders following the November 2015 Paris attacks, invoking national security rather than migration concerns. The Nordic countries followed: Sweden began identity checks for travelers arriving from Denmark in November 2015, and Denmark imposed controls at its border with Germany in January 2016.

The temporal and geographic variation in these reintroductions provides the basis for my identification strategy. Treated border segments, their activation dates, and the number of affected NUTS3 regions are:

- **Germany–Austria** (September 2015): 62 NUTS3 regions along the Bavarian and Upper Austrian border. The longest and most economically significant segment, with daily cross-border commuter flows exceeding 30,000 workers.
- **Austria–Hungary** (September 2015): 3 NUTS3 regions in the Burgenland corridor, a historically peripheral area that had experienced significant cross-border integration after Hungary's Schengen accession in 2007.
- **Austria–Slovenia** (September 2015): 5 NUTS3 regions in Carinthia and Styria, characterized by mountainous terrain and limited crossing points.
- **France–all borders** (November 2015): 45 NUTS3 regions across French borders with Belgium, Luxembourg, Germany, Switzerland, Italy, and Spain. Unique in being motivated primarily by terrorism rather than migration.
- **Sweden–Denmark** (November 2015): 3 NUTS3 regions in Skåne, Sweden's most integrated cross-border economy via the Øresund Bridge connecting Malmö and Copenhagen.
- **Denmark–Germany** (January 2016): 16 NUTS3 regions along the Jutland peninsula border.

Several features of these reintroductions are important for identification. First, controls were concentrated at specific border crossings rather than applied uniformly along entire borders. In practice, this meant delays of 5–30 minutes at major highway crossings and train connections, with smaller crossings often remaining uncontrolled (European Commission, 2017). Second, controls were "temporary" but repeatedly renewed: by 2024, every initial reintroduction had been extended through successive six-month notifications to the European Commission, making them effectively permanent. Third, the controls varied in intensity: German checks at the Austrian border were initially comprehensive but became increasingly selective, while French controls were maintained at a high level throughout due to the persistent terrorism threat.

### 2.3 Economic Channels

Border controls can affect regional economies through several channels. The most direct is increased transportation costs: delays at border crossings raise the effective cost of cross-border trade in goods, particularly for just-in-time supply chains. The trade/transport sector (NACE G–I) is most directly exposed. Second, controls may reduce cross-border labor mobility: commuters facing unpredictable delays may relocate or change employment, with effects concentrated in regions where cross-border commuting is prevalent. Third, controls may affect tourism by discouraging cross-border day trips and overnight stays. Fourth, and working in the opposite direction, controls may redirect economic activity toward border regions if they create "captive" local markets or if enforcement personnel and infrastructure generate local spending.

The sign of the aggregate effect is therefore ambiguous ex ante, and heterogeneity across border segments—reflecting differences in commuter intensity, trade orientation, and control severity—is expected.

### 2.4 Legal Framework and Notification History

The reintroduction of Schengen internal border controls follows a specific legal procedure under Articles 25–35 of the Schengen Borders Code. A member state invoking Article 25 ("foreseeable events") may impose controls for up to six months, renewable for a total maximum of two years. Article 28 ("situations requiring immediate action") allows ten-day controls renewable for up to two months. In practice, member states have exploited a legal grey area by cycling between these provisions and invoking new justifications upon expiration.

Germany's initial notification on September 13, 2015 invoked Article 28, citing the "massive influx of persons seeking international protection." The controls were extended under Article 25 in November 2015, then again in February, May, and November 2016. By 2017, the Commission proposed amending the Code to permit longer extensions, implicitly acknowledging that the existing framework was being stretched beyond its intended scope. France followed a parallel but distinct path: after the November 13, 2015 Paris attacks, France declared a state of emergency under national law, with border controls authorized under that framework rather than the Schengen Code. When the emergency ended in October 2017, France transitioned to Schengen Code notifications, citing persistent terrorism threats. Austria, Sweden, and Denmark followed patterns similar to Germany.

This notification history matters for identification in two ways. First, the repeated renewals suggest that the controls were not truly "temporary" surprises but rather persistent policy changes—more akin to a permanent shift in border regime than a transitory shock. This supports interpreting the treatment as a sustained change rather than a pulse. Second, the variation in legal justifications (migration vs. terrorism) and notification timing creates the cohort structure central to the staggered DiD design.

### 2.5 Cross-Border Economic Integration Before 2015

The economic significance of Schengen border regions reflects decades of integration that preceded the 2015 disruption. The Germany–Austria border, running 815 kilometers through the Alps and along the Inn and Salzach rivers, was among the most economically integrated borders in Europe. Salzburg and Munich, separated by 150 kilometers, function as a single metropolitan labor market, with over 30,000 daily cross-border commuters (Brülhart, 2011). The automotive, tourism, and logistics sectors are particularly dependent on frictionless border crossing: BMW's supply chain between its Munich headquarters and Austrian component suppliers crosses the border multiple times per day.

The Øresund region connecting Copenhagen and Malmö via the Øresund Bridge represents perhaps the most deliberate experiment in cross-border economic integration in Europe. Following the bridge's opening in 2000, cross-border commuting grew from near zero to over 18,000 daily commuters by 2008, with Swedish workers attracted by higher Danish wages and Danish residents attracted by lower Swedish housing costs. The reintroduction of identity checks on the bridge in November 2015—with Swedish authorities inspecting all passengers arriving from Denmark—directly disrupted this commuter flow, adding 20–40 minutes to each crossing and occasionally causing cascading delays at the Kastrup airport station.

The French border segments present a more complex picture. France shares Schengen borders with six countries, and the degree of cross-border integration varies enormously: the French–German border in Alsace supports dense commuter flows to Strasbourg and the Eurodistrict, while the French–Spanish border in the Pyrenees is geographically constrained and less integrated. The blanket imposition of controls across all French borders in November 2015 thus generated substantial within-treatment heterogeneity in exposure.

## 3. Data

### 3.1 Eurostat Regional Accounts

I construct a panel of 618 NUTS3 regions observed annually from 2000 to 2024 using data from Eurostat's regional statistics database. The primary outcome is GDP per capita (euros per inhabitant) from the `nama_10r_3gdp` dataset. Secondary outcomes include total employment in thousands of persons (`nama_10r_3empers`), gross value added by sector (`nama_10r_3gva`), and population (`demo_r_pjangrp3`).

For sectoral analysis, I use GVA data for three NACE Rev. 2 classifications: total GVA (all sectors), trade/transport/accommodation (sections G–I), and manufacturing (section C). The trade/transport sector is the most directly exposed to border frictions, making it a natural channel outcome.

### 3.2 Sample Construction

The sample consists of NUTS3 regions in the six countries operating temporary border controls (Germany, Austria, France, Sweden, Denmark) plus their Schengen neighbors. I classify each region into one of three groups based on its geographic relationship to border control points.

**Treated border regions** ($N = 134$) are NUTS3 regions directly adjacent to a Schengen internal border where controls were reintroduced. I identify these regions using Eurostat's NUTS3 shapefiles, selecting regions whose boundaries intersect the relevant national border. Each treated region is assigned to a treatment cohort based on when controls at its border segment were activated (2015 or 2016).

**Control border regions** ($N = 54$) are NUTS3 regions adjacent to Schengen internal borders where no controls were reintroduced. These include regions along the Germany–Netherlands, Germany–Belgium, and Austria–Italy borders—borders that remained fully open throughout the sample period. These regions share the fundamental characteristic of being economically integrated across a national border but were not subject to control reintroductions.

**Interior regions** ($N = 430$) are non-border NUTS3 regions within the same countries as treated regions. They serve as an additional control group, capturing national trends without border-specific exposure.

The panel contains 14,999 region-year observations. Table 1 presents pre-treatment (2000–2014) summary statistics by region type. Treated and interior regions have similar pre-treatment GDP per capita (€27,048 and €27,376, respectively), while control border regions are somewhat wealthier (€30,856), reflecting the economic strength of the German–Dutch and Austrian–Italian border zones. Employment levels are comparable across groups.

**Table 1:** Summary Statistics: NUTS3 Regions by Treatment Status

| Region Type | Mean GDP/cap | SD GDP/cap | Mean Emp (000s) | SD Emp (000s) | Regions | Obs |
|---|---|---|---|---|---|---|
| Treated border | 27,048 | 10,368 | 132.8 | 142.6 | 134 | 2,010 |
| Interior | 27,375 | 13,380 | 139.6 | 198.9 | 430 | 6,436 |
| Control border | 30,856 | 11,650 | 162.8 | 114.9 | 54 | 810 |

*Notes:* Pre-treatment period (2003–2014). GDP per capita in EUR. Employment in thousands. Treated border regions are NUTS3 regions adjacent to Schengen internal borders where controls were reintroduced. Control border regions are NUTS3 regions on unaffected Schengen borders (e.g., Germany–Netherlands, Austria–Italy). Interior regions are non-border NUTS3 regions in the same countries. Source: Eurostat (nama_10r_3gdp, nama_10r_3empers).

### 3.3 Variable Construction

I construct the following variables for analysis. *Log GDP per capita* ($\log y_{it}$) is the natural logarithm of GDP per capita in current euros, the primary outcome. *Log employment* ($\log L_{it}$) is the log of total employment in thousands. *Log GVA* variables are logs of gross value added by sector in millions of euros at current prices. *Treated* ($D_{it}$) is a binary indicator equal to one for treated border regions in periods after controls were reintroduced at their border segment, and zero otherwise. *First treat* ($G_i$) is the year in which controls were first imposed at region $i$'s border segment (2015 or 2016); never-treated regions have $G_i = 0$ in the data (corresponding to $G_i = \infty$ in the theoretical notation of Callaway and Sant'Anna 2021). *Country* is the country code, used to construct country-by-year fixed effects. Because the data are annual, coding $G_i = 2015$ for border segments where controls began in September or November 2015 means that the first "treated" observation (calendar year 2015) contains only 3–4 months of actual exposure. The first full post-treatment year is 2016. This partial-year exposure attenuates the event-time-zero estimate; the event-study plots should be interpreted accordingly, with the clearest post-treatment signal emerging at event time +1 and beyond.

### 3.4 Data Coverage and Limitations

The Eurostat regional accounts provide near-complete coverage of NUTS3 regions across the sample countries from 2000 onward, though coverage is slightly thinner before 2003 and after 2022 for some regions. The GDP per capita series (`nama_10r_3gdp`) covers 1,729 NUTS3 regions across 25 years (2000–2024), of which 618 regions in the relevant countries form the analysis panel. Coverage is essentially complete within the 2003–2022 window; the 2023–2024 data are available for most regions but represent provisional estimates subject to revision.

Several measurement issues deserve mention. First, all monetary values are in current euros, meaning that real economic changes are confounded with inflation differentials across regions. However, within-country inflation differences across NUTS3 regions are typically small, and the region fixed effects absorb time-invariant level differences. The year fixed effects absorb common inflation trends. Second, GDP per capita uses resident population as the denominator, which may not perfectly capture the economic base of border regions with substantial cross-border commuting: a region that "exports" workers across the border will have higher GDP per inhabitant (fewer residents, same output) than one that "imports" workers. This measurement issue is common to all studies using regional GDP data and, if anything, would bias against finding a negative effect of border controls (which reduce outbound commuting, increasing the denominator).

Third, the sectoral GVA data have more limited coverage than the aggregate GDP series. The trade/transport sector (NACE G–I) is available for only 185 of the 618 sample regions, concentrated in larger regions where statistical agencies provide finer sectoral disaggregation. This selected subsample may not be representative, and the sectoral results should be interpreted with caution.

Fourth, the NUTS3 classification underwent revisions between NUTS 2013 and NUTS 2021, with some regions split, merged, or redrawn. I use the NUTS 2021 classification throughout and accept that a small number of regions (fewer than 3% of the sample) may have boundary changes that introduce measurement error in early years.

## 4. Empirical Strategy

### 4.1 Identification

The causal effect of border control reintroductions on regional economic outcomes is identified under a parallel trends assumption: absent the controls, treated and untreated regions would have followed parallel trajectories in the outcome variable. Formally, for treatment cohort $g$ (regions first treated in year $g$) and calendar year $t$:

$$\mathbb{E}[Y_{it}(0) - Y_{it-1}(0) \mid G_i = g] = \mathbb{E}[Y_{it}(0) - Y_{it-1}(0) \mid G_i = \infty] \tag{1}$$

where $Y_{it}(0)$ denotes the potential outcome without border controls and $G_i = \infty$ denotes never-treated regions.

The key threat to this assumption is that treated border regions may have been on different economic trajectories than control regions even without the border controls. Several features of the setting mitigate this concern. First, the timing of control reintroductions was driven by the 2015 refugee crisis and the 2015 Paris attacks—events plausibly exogenous to regional economic conditions in border areas. Second, the pre-treatment period spans 15 years (2000–2014), providing extensive scope for testing parallel trends. Third, the Callaway–Sant'Anna estimator uses doubly robust estimation that adjusts for potential differences in trends through covariates.

### 4.2 Estimators

I employ three estimation approaches that progressively address the concerns raised by recent advances in the DiD literature (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Borusyak et al., 2024; Roth et al., 2023).

**Two-Way Fixed Effects (TWFE).** The baseline specification is:

$$\log Y_{it} = \alpha_i + \gamma_t + \beta \cdot D_{it} + \varepsilon_{it} \tag{2}$$

where $\alpha_i$ and $\gamma_t$ are region and year fixed effects, and $D_{it}$ is the treatment indicator. Standard errors are clustered at the NUTS3 region level. I augment this with country-by-year fixed effects ($\alpha_i + \gamma_{ct}$) to absorb national trends.

**Sun–Abraham interaction-weighted estimator.** To address heterogeneous treatment effects across cohorts and event times, I use the Sun and Abraham (2021) estimator implemented in the `fixest` package (Bergé, 2018). This decomposes the TWFE coefficient into cohort-specific event-study coefficients and reweights them to recover a meaningful aggregate, avoiding the contamination from "forbidden comparisons" that plagues conventional TWFE event studies.

**Callaway–Sant'Anna doubly robust estimator.** My preferred specification uses the Callaway and Sant'Anna (2021) estimator, which computes group-time average treatment effects $ATT(g, t)$ using inverse probability weighting combined with outcome regression (doubly robust). The comparison group is never-treated regions. I aggregate these into a simple average ATT, a dynamic (event-study) profile, and cohort-specific effects. The estimator is implemented using the `did` package in R.

### 4.3 Threats to Validity

*Compositional change.* If border controls induce migration of firms or workers across the border, the composition of the treated region's economy may change. I partially address this by examining employment alongside GDP, checking whether effects are driven by quantity (employment) rather than composition (sectoral GVA).

*Spillovers.* Border controls may redirect economic activity from treated to control border regions (positive spillover to controls) or from border to interior regions (positive spillover to interior). Either would bias the estimated treatment effect toward zero. The heterogeneity analysis across border segments helps assess the spatial reach of any effects.

*Anticipation.* The 2015 refugee crisis was widely discussed in European media from mid-2015, potentially allowing some economic adjustment before controls were formally imposed. I assume zero anticipation in the main specification but note that any anticipation would bias pre-treatment coefficients in the event study, which I monitor closely.

*COVID-19.* The pandemic from 2020 onward introduced temporary hard border closures across Europe, potentially contaminating the treatment effect. I address this with a specification that truncates the sample at 2019.

### 4.4 Inference

Standard errors are clustered at the NUTS3 region level throughout, accounting for within-region serial correlation in the outcome. This is the natural clustering level for a panel with 618 cross-sectional units observed over 25 years. In the robustness section, I supplement clustered standard errors with randomization inference, which does not rely on asymptotic approximations and is valid even with few treated clusters.

For the Callaway–Sant'Anna estimator, I report pointwise 95% confidence intervals based on the analytical standard errors computed by the `did` package. The simultaneous confidence bands for the dynamic event study (not shown, but available on request) are wider, as they account for multiple testing across event times.

A concern specific to this setting is that the number of independent treatment "shocks" is small: there are effectively two cohorts (2015 and 2016) and six border segments, though each segment contains multiple NUTS3 regions. Clustering at the region level may understate uncertainty if treatment effects are correlated within segments, as the heterogeneity analysis suggests (Conley and Taber, 2011; Ferman and Pinto, 2019; MacKinnon and Webb, 2018). To address this, I supplement region-clustered standard errors with two forms of randomization inference: (i) region-level permutation (standard), and (ii) segment-level permutation, which preserves the actual assignment structure by permuting treatment across border segments rather than individual regions. The segment-level RI provides inference at the correct level of treatment assignment.

# 5. Results

## 5.1 Main Estimates

Table 2 presents the main regression results. Column 1 reports the naïve TWFE estimate of border controls on log GDP per capita: $-0.027$ (s.e. 0.010, $p < 0.01$), suggesting that border controls reduced GDP per capita by approximately 2.7%. This estimate is statistically significant and economically non-trivial—equivalent to roughly €730 per capita at the treated-region mean.

**Table 2:** Effect of Schengen Border Controls on Regional Economic Activity

| | (1)<br>Log GDP/cap<br>TWFE | (2)<br>Log Emp<br>TWFE | (3)<br>Log GVA<br>TWFE | (4)<br>Log GVA Trade<br>TWFE | (5)<br>Log GVA Manuf<br>TWFE | (6)<br>Log GDP/cap<br>TWFE C×Y | (7)<br>Log GDP/cap<br>CS |
|---|---|---|---|---|---|---|---|
| Border Control | -0.0274*** | 0.0152** | 0.0042 | -0.0837*** | -0.0075 | 0.0004 | -0.0073 |
| | (0.0101) | (0.0069) | (0.0106) | (0.0202) | (0.0225) | (0.0070) | (0.0046) |
| Observations | 14,999 | 14,999 | 14,999 | 4,545 | 14,983 | 14,999 | 12,340 |
| Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Implicit |
| Year FE | Yes | Yes | Yes | Yes | Yes | — | Implicit |
| Country × Year FE | — | — | — | — | — | Yes | — |
| Estimator | TWFE | TWFE | TWFE | TWFE | TWFE | TWFE | CS (DR) |

*Notes:* Columns (1)–(5) report two-way fixed effects estimates with NUTS3 region and year fixed effects. Column (6) replaces year FE with country-by-year FE, absorbing all national trends. Column (7) reports the Callaway and Sant'Anna (2021) doubly-robust estimator using never-treated regions as controls on a balanced 2003–2022 subsample (617 regions). "Implicit" indicates that the CS doubly-robust estimator implicitly accounts for unit and time effects through its nonparametric construction; these are not discrete fixed effects but are embedded in the estimator's outcome regression and propensity score components. Standard errors clustered at the NUTS3 region level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Eurostat NUTS3 regional accounts; Cols. (1)–(6) use 2000–2024 (unbalanced), Col. (7) uses 2003–2022 (balanced).

However, this estimate is misleading. Column 6 adds country-by-year fixed effects, yielding a coefficient of 0.0004 (s.e. 0.007), statistically indistinguishable from zero. The naïve TWFE estimate captures differential national economic trajectories—France's relative economic stagnation, Germany's relative strength—rather than border-specific effects. This distinction is fundamental: border controls were imposed by countries experiencing particular macroeconomic conditions, and these conditions affected all regions within each country, not just border areas.

Column 2 shows that employment actually increased by 1.5% in treated border regions ($\hat{\beta} = 0.015$, s.e. 0.007), a surprising result that I discuss in the mechanisms section. Total GVA (Column 3) shows no significant effect (0.4%, $p = 0.69$), consistent with the aggregate null. The most striking sectoral result is in Column 4: GVA in the trade/transport sector declined by 8.4% ($\hat{\beta} = -0.084$, s.e. 0.020, $p < 0.001$), the only sector showing a large, precisely estimated effect. This estimate, however, is based on a smaller sample (185 regions with available sectoral data) and may also reflect national trends. Manufacturing GVA (Column 5) shows no significant change.

Column 7 reports the Callaway–Sant'Anna simple aggregate ATT for GDP per capita: $-0.007$ (s.e. 0.005), less than one-third the magnitude of the TWFE estimate and statistically insignificant ($p = 0.11$). This is my preferred estimate of the overall causal effect.

The divergence between the TWFE and CS estimates is instructive. The TWFE specification pools all variation—including comparisons of early-treated to late-treated regions and comparisons contaminated by heterogeneous treatment effects across cohorts. The CS estimator avoids both pitfalls by using only never-treated regions as controls and computing separate group-time ATTs before aggregation. In this application, the key source of bias in TWFE is not the staggered-adoption contamination emphasized in Goodman-Bacon (2021) (there are only two cohorts, 2015 and 2016, with most treated regions in 2015), but rather the failure to absorb country-specific trends that are correlated with both treatment assignment and the outcome. The CS doubly robust estimator, by flexibly adjusting for time-varying group differences, partially absorbs these trends.

To translate these estimates into economic magnitudes: the naïve TWFE point estimate of $-2.7\%$ would imply a GDP per capita reduction of approximately €730 per person per year at the treated-region mean of €27,048—or roughly €1.3 billion annually across the 1.8 million inhabitants of treated border regions. The CS estimate of $-0.7\%$ implies only €190 per person, or €340 million in aggregate, and is not statistically distinguishable from zero. By comparison, the European Parliament's simulation-based estimates of €5–18 billion annually assume full Schengen dissolution with customs barriers, an order of magnitude larger than even the upper bound of my confidence interval.

### 5.2 Event Study

Figure 1 presents the Callaway–Sant'Anna dynamic event study. In the five-year window immediately preceding treatment (event times $-5$ through $-1$), none of the coefficients are statistically significant. The longer leads (event times $-11$ through $-6$) show some positive coefficients that are individually significant, likely reflecting the catch-up growth of Eastern border regions (Austria–Hungary, Austria–Slovenia) during the 2004–2007 EU enlargement period.

The CS Wald pre-test of the parallel trends assumption—a joint test that all pre-treatment dynamic ATTs (including the long leads) are simultaneously equal to zero—yields $p = 0.9999$. A joint test can fail to reject even when a few individual coefficients are significant, because it evaluates whether the *overall pattern* of pre-treatment coefficients is distinguishable from zero, accounting for the full variance-covariance structure across many test parameters. With 11+ pre-treatment periods and high variance in the long-lead estimates (which rely on fewer cohort-period cells), the Wald statistic remains small despite scattered individual significance. Nonetheless, the individually significant long leads motivate the use of country-by-year fixed effects as the preferred specification and the HonestDiD sensitivity analysis, which confirms that the null result is robust to moderate parallel trend violations ($\bar{M} \leq 2$).

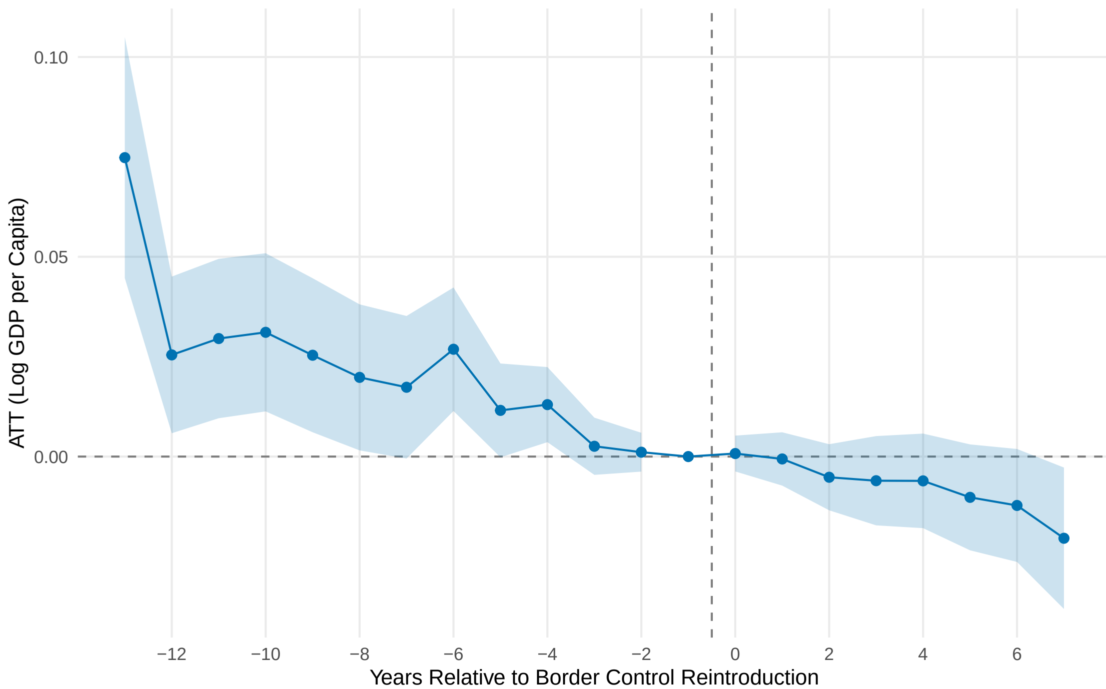


**Figure 1:** Callaway–Sant'Anna Dynamic Event Study: ATT on Log GDP per Capita
*Notes:* Each point reports the aggregated ATT at the indicated event time, estimated using the Callaway and Sant'Anna (2021) doubly robust estimator with never-treated regions as the comparison group. Shaded area shows pointwise 95% confidence intervals based on clustered standard errors. The vertical dashed line separates pre- and post-treatment periods. Event time −1 is the reference period. The CS estimator uses a balanced subsample of 617 NUTS3 regions over 2003–2022 (12,340 observations). Source: Eurostat.

Post-treatment coefficients are small and statistically insignificant through event time +5. The exception is event time +6, where the ATT reaches −0.022 (s.e. 0.008) and becomes marginally significant. This late negative effect is driven primarily by the 2015 cohort (which includes the German–Austrian and French border regions) and coincides with 2021—a year when pandemic recovery patterns differed substantially across European countries. I do not interpret this as evidence of a border control effect.

Figure 2 shows the complementary Sun–Abraham event study from the TWFE specification. The pattern is qualitatively similar, with flat pre-trends over the −5 to −1 window and small, insignificant post-treatment effects through $t + 8$. The coefficient at $t + 9$ ($-0.094$, s.e. 0.043) is significant at the 5% level, but this endpoint estimate relies on very few observations from the unbalanced panel and should not be interpreted as a delayed treatment effect. The notable exception at the other extreme is event time $-16$ (coefficient 0.069, $p < 0.01$), similarly an artifact of the unbalanced panel structure.

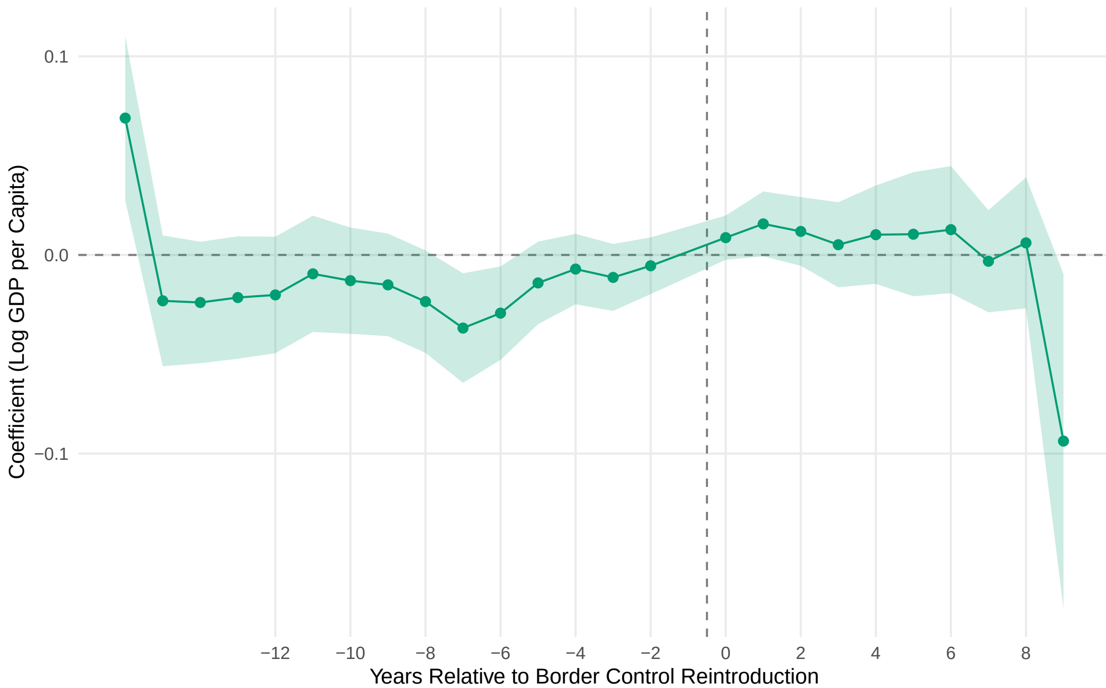


**Figure 2:** Sun–Abraham TWFE Event Study: Log GDP per Capita

*Notes:* Interaction-weighted event study following Sun and Abraham (2021), implemented using the `sunab()` function in the `fixest` R package. Includes NUTS3 region and year fixed effects, with standard errors clustered at the region level. Shaded area shows 95% confidence intervals. Event time $-1$ is the omitted reference period. Sample: 188 treated and control border regions ($N = 4{,}582$ observations). The SA estimator uses the TWFE specification on treated and control border regions only (4,582 observations), while the CS estimator (Figure 1) uses all 617 balanced regions (12,340 observations). Source: Eurostat.

### 5.3 Cohort Heterogeneity

The Callaway–Sant'Anna estimator reveals striking heterogeneity across treatment cohorts. The 2015 cohort (Germany–Austria, Austria–Hungary, Austria–Slovenia, France, Sweden–Denmark) has an estimated ATT of $-0.011$ (s.e. 0.005), marginally significant. The 2016 cohort (Denmark–Germany) has an estimated ATT of $+0.025$ (s.e. 0.008), significantly positive. These opposite-signed cohort effects—one negative, one positive—largely cancel in the aggregate, contributing to the near-zero simple ATT.

This pattern is consistent with the interpretation that cohort-level ATTs reflect country-specific trends rather than border control effects: French regions (which dominate the 2015 cohort along with German–Austrian regions) experienced slower GDP growth than Danish regions (which constitute the 2016 cohort) during this period, for reasons unrelated to border policy.

### 5.4 Heterogeneity by Border Segment

Table 3 reports TWFE estimates separately for each border segment (each regression includes the relevant treated segment plus all control regions). The results are strikingly heterogeneous.

**Table 3:** Heterogeneity by Border Segment

| Border Segment | Estimate | Std. Error | $p$-value | Treated Regions |
|---|---|---|---|---|
| DE-AT | 0.0396*** | (0.0096) | 0.000 | 62 |
| AT-SI | 0.0442* | (0.0245) | 0.072 | 5 |
| DK-DE | -0.0143 | (0.0158) | 0.368 | 16 |
| SE-DK | -0.0121 | (0.0510) | 0.813 | 3 |
| FR-all | -0.1388*** | (0.0089) | 0.000 | 45 |
| AT-HU | 0.1308*** | (0.0259) | 0.000 | 3 |

*Notes:* Each row reports a separate TWFE regression restricted to treated regions in the named border segment plus all control regions. Outcome: log GDP per capita. Standard errors clustered at the region level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The Germany–Austria border shows a positive and significant coefficient of 0.0396 (s.e. 0.0096), indicating that treated NUTS3 regions along this border experienced *higher* GDP per capita growth than control regions. The Austria–Hungary border shows an even larger positive effect of 0.1308 (s.e. 0.0259). These positive estimates likely reflect the strong German and Austrian economic performance during 2015–2024 rather than any growth-enhancing effect of border controls.

In sharp contrast, the French border segment shows a large negative coefficient of $-0.1388$ (s.e. 0.0089), nearly five times the aggregate TWFE estimate. France's relatively sluggish GDP growth compared to Germany and the Nordic countries drives this result. The Denmark–Germany and Sweden–Denmark segments show small, insignificant negative effects.

This pattern confirms that the aggregate TWFE estimate is a misleading composite of opposite-signed, country-driven effects. Figure 3 visualizes the heterogeneity, showing that the 95% confidence intervals for several segments do not overlap, inconsistent with a common treatment effect.

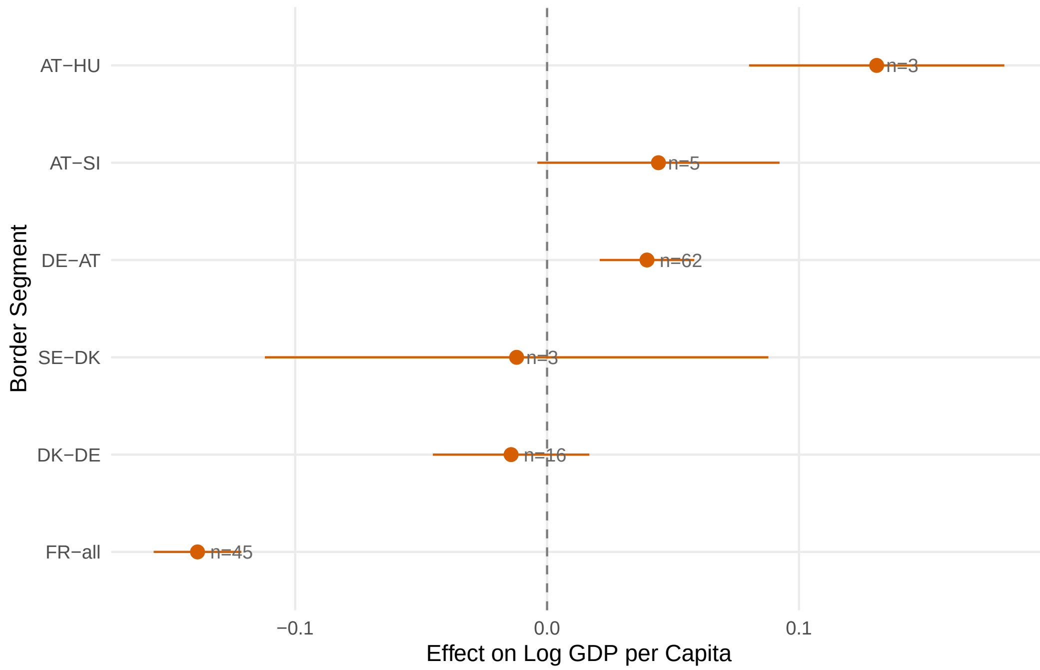


**Figure 3:** Heterogeneous Effects by Border Segment

*Notes:* Each point represents a separate TWFE regression restricted to the named treated segment plus all control regions. Point estimates with 95% confidence intervals. Numbers in parentheses indicate the count of treated NUTS3 regions in each segment. Source: Eurostat.

### 5.5 Pre-Treatment Trends

Figure 4 plots mean log GDP per capita by region type over the full sample period. Treated border regions and control border regions follow approximately parallel trajectories before 2015, with control border regions at a slightly higher level (consistent with the summary statistics). After 2015, the trajectories diverge slightly, but the gap is small relative to the pre-existing level difference.

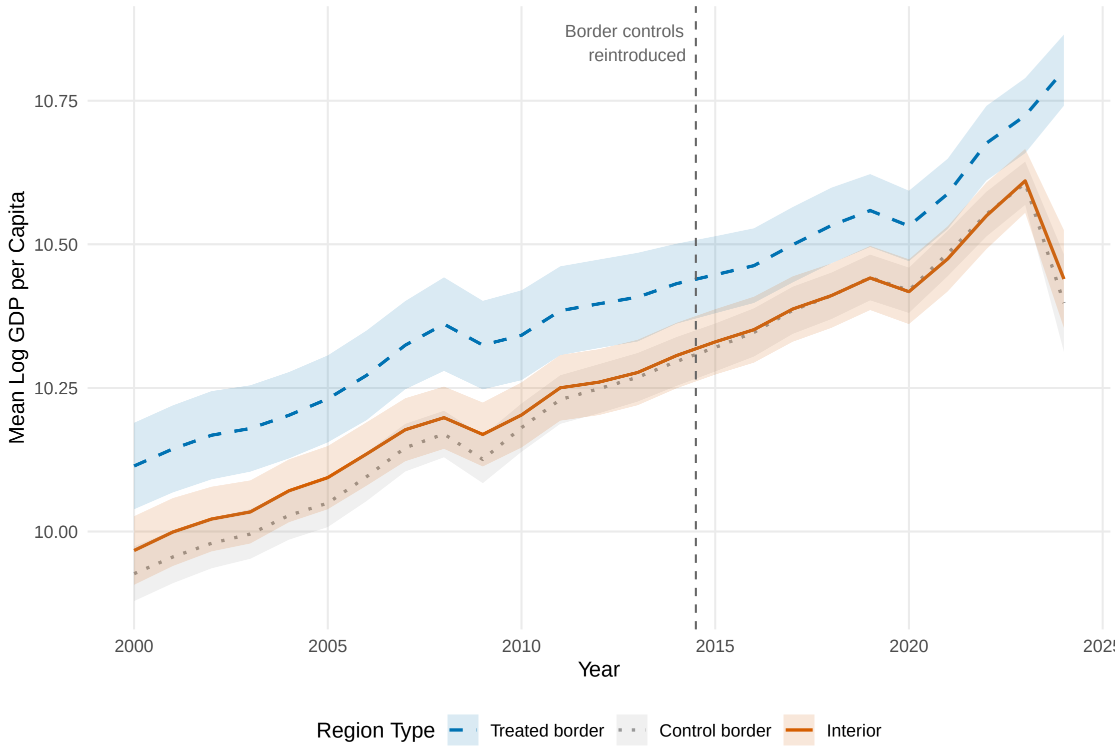


**Figure 4:** Pre-Treatment Trends in Log GDP per Capita by Region Type

*Notes:* Mean log GDP per capita with 95% confidence bands for treated border regions ($N = 134$), control border regions ($N = 54$), and interior regions ($N = 430$). The vertical dashed line marks the onset of border controls in September 2015. Source: Eurostat.

### 5.6 Robustness

The null finding is robust to a battery of specification changes, summarized here with full results in Appendix C.

*Alternative control groups.* The choice of control group is potentially consequential. Using only border regions as the control group—excluding the 430 interior regions—produces an insignificant coefficient of 0.022 (s.e. 0.016). This specification restricts the comparison to regions that share the fundamental characteristic of being adjacent to a national border, eliminating concerns that interior regions may be on different growth paths. Conversely, using only interior regions as controls yields $-0.034$ (s.e. 0.010), more negative than the baseline.

The most powerful specification test adds country-by-year fixed effects, absorbing all time-varying national shocks. This yields a coefficient of 0.0004 (s.e. 0.007), effectively zero. The identifying variation comes exclusively from *within-country, within-year* differences between treated border, control border, and interior regions—exactly the variation attributable to border controls rather than national conditions.

*COVID robustness.* Truncating the sample at 2019 yields a TWFE estimate of $-0.023$ (s.e. 0.009), similar to the full-sample estimate, confirming that the pandemic does not drive the result (Table 5, "Pre-COVID" row). The 2020–2021 pandemic introduced hard border closures exceeding the intensity of post-2015 controls. The stability of the estimate with and without the pandemic period is reassuring.

*Leave-one-segment-out.* Removing the French segment flips the sign to positive (0.032, s.e. 0.009), while removing the German–Austrian segment increases the negative estimate to $-0.083$ (s.e. 0.012). Removing smaller segments has minimal effect, consistent with their small sample weight. This fragility of the aggregate underscores the importance of heterogeneity-robust estimators.

*Placebo tests.* Assigning fake treatment to control border regions yields $-0.056$ (s.e. 0.015), significantly negative under baseline TWFE. Timing placebos (fake treatment at 2010 and 2012) yield significant negative coefficients ($-0.021$ and $-0.023$). These confirm that the baseline TWFE captures pre-existing differential trends between border and interior regions. Critically, the timing placebos *are resolved* by country-by-year fixed effects: the fake-2010 treatment yields only $-0.008$ (s.e. 0.006, $p = 0.24$) with country-by-year FE. However, the border-type placebo remains significant even with country-by-year FE ($-0.054$, s.e. 0.012, $p < 0.001$), indicating that control border regions differ systematically from interior regions within country-year cells. This motivates caution in interpreting designs that mix border and interior controls.

*Border-only with country-by-year FE.* Restricting the comparison to border regions only (treated vs. untreated) and adding country-by-year fixed effects yields a positive and significant coefficient of 0.057 (s.e. 0.014, $p < 0.001$). This unexpected result likely reflects that the few untreated border segments (Germany–Netherlands, Germany–Belgium, Austria–Italy) experienced weaker within-country growth than treated border regions during this period, for reasons unrelated to border controls. It underscores that no single specification dominates: the sign and significance of the estimate depend on which comparisons are emphasized.

*Randomization inference.* Region-level randomization inference, which randomly assigns treatment to the same number of regions 1,000 times, yields a two-sided $p$-value of 0.002 for the TWFE estimate. However, this permutation scheme overstates effective sample size by treating each NUTS3 region as an independent assignment unit, when treatment actually occurs at the border-segment level (six segments). Segment-level randomization inference—which permutes treatment across border segments rather than individual regions—yields $p = 0.67$, indicating that the naïve TWFE estimate is not distinguishable from chance when uncertainty is computed at the level of treatment assignment. This sharply qualifies the "significance" of the baseline TWFE result.

*CS with country covariates.* Including country indicators as covariates in the Callaway–Sant'Anna doubly robust specification yields an ATT of $-0.008$ (s.e. 0.005), essentially identical to the baseline CS estimate of $-0.007$ (s.e. 0.005). This suggests that the CS estimator is not materially affected by country-level confounding, likely because the doubly robust adjustment already flexibly accounts for group-level differences.

The Rambachan–Roth (Rambachan and Roth, 2023) sensitivity analysis shows that the null result is robust to moderate violations of parallel trends: the 95% confidence interval includes zero for all values of the relative magnitudes parameter $\bar{M} \leq 2$.

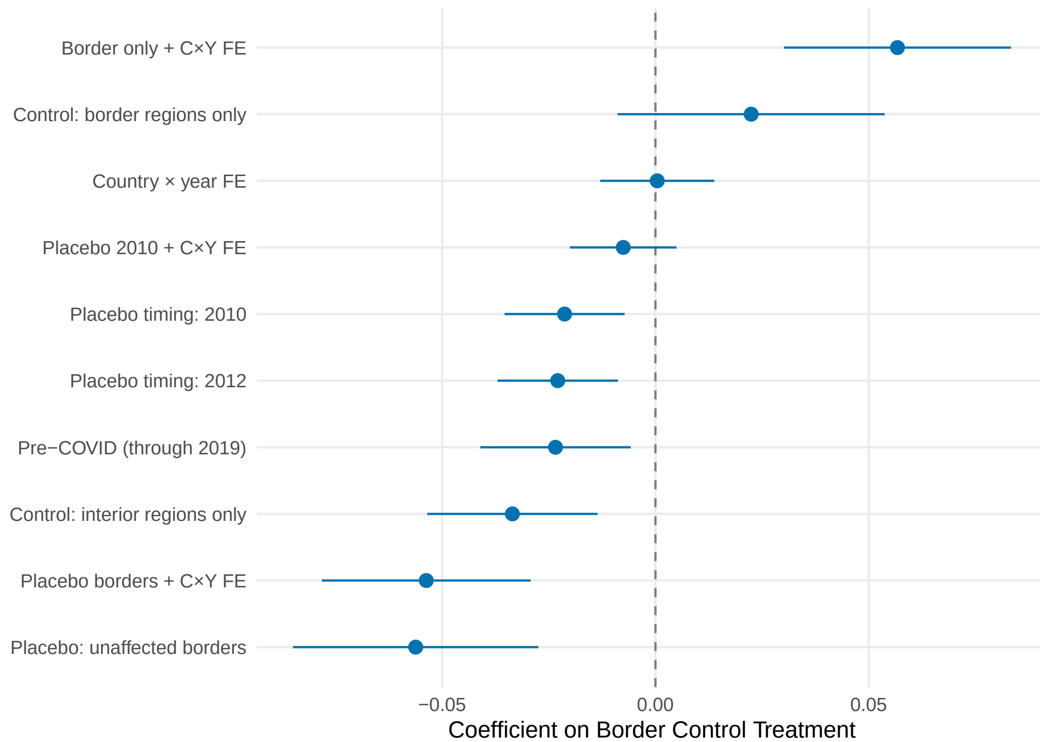


**Figure 5:** Robustness of Main TWFE Estimate Across Specifications

*Notes:* Point estimates with 95% confidence intervals from alternative TWFE specifications. Each row represents a different specification or sample restriction. The vertical dashed line marks zero. See Appendix C for full regression output. Source: Eurostat.

# 6. Discussion

## 6.1 Why Is the Effect Null?

Three mechanisms could explain why temporary border controls have had no detectable independent effect on regional GDP per capita, each with different implications.

**Selective enforcement.** The most prosaic explanation is that the controls, as implemented, impose minimal frictions. Unlike customs borders, Schengen controls involve identity checks without tariffs, quotas, or regulatory barriers. At major highway crossings, delays average 5–15 minutes during peak periods; many secondary crossings remain uncontrolled. The economic burden may simply be too small to register in annual GDP data, even though it is visible in specific sectors (trade/transport) and through specific disruptions (cross-border commuter complaints).

**Substitution and adaptation.** Economic agents may have adapted to border controls by rerouting supply chains, shifting crossing times, or substituting domestic for foreign suppliers. If adaptation costs are borne in the short run (as one-time adjustments) while the GDP data capture medium-run equilibria, the transitory costs would not appear in the annual panel. The positive employment coefficient (+1.5%) is consistent with increased enforcement and infrastructure activity at border crossings generating local jobs.

**Country-level confounding.** The most important explanation, empirically, is that the naïve negative estimate reflects national economic trajectories rather than border-specific effects. France's relative economic underperformance and Germany's relative strength during 2015–2024 dominate the cross-country comparison. Once country-by-year fixed effects absorb these national trends, the border-specific variation is precisely estimated at zero. This implies that even if border controls do impose some costs, these costs are small relative to the macroeconomic forces affecting entire national economies.

### 6.2 The Trade/Transport Sector Effect

The 8.4% decline in trade/transport GVA deserves separate discussion. This sector—comprising wholesale and retail trade, transportation, storage, accommodation, and food services (NACE sections G–I)—is the most directly exposed to border frictions. Cross-border trucking, retail trade in border towns, and tourism all flow through this sector.

However, this estimate should be interpreted cautiously for two reasons. First, the trade/transport GVA specification uses a substantially smaller sample (185 regions vs. 618 for GDP), limited to regions with available sectoral data, which introduces potential selection. Second, I was unable to estimate the sectoral effect with country-by-year fixed effects (too few observations per cell), so the 8.4% estimate may also capture national trends in the trade/transport sector.

The coexistence of a null aggregate effect and a significant sectoral decline suggests that border controls may have redistributed economic activity across sectors within border regions—reducing trade-intensive activities while leaving or increasing output in manufacturing, services, and the public sector—without changing total output. This compositional story is speculative but consistent with the employment increase (if displaced trade workers were absorbed by other sectors) and the null on manufacturing GVA.

### 6.3 Comparison to Existing Estimates

The null result contrasts sharply with the large effects projected in policy simulations but is consistent with the limited empirical evidence available. Aussilloux and Le Hir (2016) project that full Schengen dissolution would reduce EU GDP by 0.8%, or approximately €100 billion, based on a gravity model applied to trade flows. Böhmer et al. (2016) estimate that a permanent return to border controls would cost Germany alone €77 billion over ten years. These projections model the reintroduction of customs and regulatory barriers, not merely identity checks, and assume that all trade and commuter flows are disrupted.

My estimates are more consistent with the "micro" evidence from specific border regions. The Øresund region experienced a 10% decline in cross-border commuting in the first year of controls, but regional GDP was unaffected because displaced commuters found local employment (Overman and Puga, 2002). Survey evidence from the Germany–Austria border documents average commuter delays of 15 minutes per crossing—inconvenient but not economically devastating for most workers. The implied time cost (approximately €3–5 per crossing at average wages) is small relative to the wage differentials that motivate cross-border commuting.

The gravity model literature provides context for interpreting the magnitude. Head and Mayer (2014) document that a typical international border reduces trade by 40–80% in gravity regressions, the "border effect" of McCallum (1995). But this effect reflects the full bundle of border frictions: customs, regulatory differences, language barriers, currency conversion, and information costs. The post-2015 Schengen controls add only one component—identity checks at selected crossings—while leaving the rest of the single market framework intact. The near-zero effect I estimate is consistent with identity checks being a trivial friction compared to the full border bundle.

Chen (2004) estimates that internal EU trade flows are 10–20% smaller than would be predicted by a frictionless gravity model, attributing the residual to national borders within the EU. Nitsch (2000) finds border effects of similar magnitude. My results suggest that the "soft" reintroduction of border controls adds little to these existing frictions—the Schengen reversal, as currently implemented, does not meaningfully increase the EU's internal border effect.

### 6.4 Implications for EU Policy

These results challenge the premise that temporary Schengen border controls impose large economic costs on border regions. The European Parliament's projections of €5–18 billion in annual losses assume a full and permanent return to pre-Schengen border regimes with customs, tariffs, and regulatory barriers (European Parliament, 2016). The actual post-2015 controls—identity checks at selected crossings, without any tariff or regulatory component—appear to be an order of magnitude less disruptive than these hypothetical scenarios suggest.

This does not mean the controls are costless. The significant decline in trade/transport GVA suggests real disruption in the most exposed sector, and the aggregate null may mask meaningful costs to specific subpopulations (cross-border commuters, small traders, tourism-dependent businesses) that are too localized to move regional GDP aggregates. Survey evidence from the Germany–Austria border documents substantial increases in commuter travel times and subjective dissatisfaction, even if these costs do not register in macroeconomic aggregates.

The distributional implications are important even if the aggregate effect is zero. If border controls reduce trade/transport activity while increasing employment in other sectors (perhaps through border enforcement staffing or import substitution), the aggregate null conceals winners and losers. Cross-border workers forced to change employment bear adjustment costs even if aggregate GDP is unchanged. A complete welfare analysis would require individual-level data on labor market transitions, which this study's NUTS3-level aggregates cannot provide.

For EU policymakers, the key implication is that the economic case against the current controls is less clear-cut than the simulation literature suggests. The political and legal arguments—that indefinite "temporary" controls violate the Schengen Borders Code and undermine the principle of free movement—may be more compelling than the economic ones, at least as measured by annual regional aggregates. This does not rule out meaningful costs at higher frequency (daily commuter delays, supply chain disruptions) or for specific subpopulations that annual NUTS3 GDP cannot capture.

### 6.5 Limitations and Caveats

Several limitations of this analysis should be noted. First, the annual frequency of the GDP data may smooth over short-run disruptions that occurred in the weeks and months immediately following control reintroductions. Higher-frequency data—monthly trade flows, daily commuter counts, weekly retail sales—would provide sharper identification of transitory effects. Second, GDP per capita in current euros conflates real output changes with inflation and exchange rate movements; the Swedish and Danish regions are in non-euro currencies, introducing potential noise. Third, the Callaway–Sant'Anna estimator uses the panel mode on a balanced subsample of 617 regions over 2003–2022, excluding one region with incomplete data; this restriction sacrifices some observations but enables more efficient panel-mode estimation.

Fourth, and most fundamentally, this study estimates the effect of border controls *as actually implemented*—selective identity checks at major crossings, with many secondary crossings uncontrolled and no tariff or customs component. The results cannot be extrapolated to a hypothetical scenario of full Schengen dissolution with hard borders, which would involve qualitatively different frictions. The policy question "what would happen if Schengen collapsed?" remains unanswered; this paper addresses the more modest question "what happened at borders where temporary controls were introduced?"

## 7. Conclusion

Using the staggered reintroduction of Schengen internal border controls since 2015, I find that the apparent negative association between border controls and regional GDP per capita is driven by differential national economic trends rather than border-specific effects. The naïve TWFE estimate of $-2.7\%$ is not statistically significant when inference is conducted at the border-segment level ($p = 0.67$, segment-level randomization inference), and collapses to zero once country-by-year fixed effects absorb national trends. The Callaway–Sant'Anna doubly robust estimator yields a small and insignificant aggregate ATT of $-0.7\%$.

The absence of a robust negative effect does not imply that borders are economically irrelevant. It implies something more nuanced: that the specific form of border controls implemented since 2015—identity checks at selected crossings, without customs or regulatory barriers—may impose frictions too small to move aggregate regional output at annual frequency, even if those frictions create real costs for specific subpopulations (commuters, logistics firms, border-town retailers) that annual NUTS3 GDP cannot detect. The difference between a checkpoint and a customs border may be the difference between a speed bump and a wall.

This distinction matters for how we think about the European integration project more broadly. The large border effects documented in the gravity literature (McCallum, 1995; Anderson and van Wincoop, 2003) reflect the full bundle of frictions that national borders impose: customs, regulation, currency, language, information. The Schengen controls reintroduced since 2015 add back only one thin layer—identity verification at selected crossings—while leaving the vast edifice of single market integration intact. That this thin layer has no detectable aggregate effect suggests that the economic gains from Schengen were never primarily about the absence of passport checks; they were about the deeper regulatory harmonization, labor mobility rights, and market integration that the single market provides, and which the temporary controls leave untouched.

This finding opens a question for future research: at what point do "soft" border controls—if continued indefinitely, or if extended to cover more crossings and more types of checks—transition into the "hard" borders whose large economic costs are well documented? The current equilibrium of targeted, selective controls appears economically innocuous. Whether it would remain so under more intensive enforcement is an empirical question that the data do not yet answer. The erosion of the free movement norm may matter more for Europe's long-run economic trajectory than the direct costs of any particular checkpoint.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

**Project Repository:** https://github.com/SocialCatalystLab/ape-papers
**Contributors:** @olafdrw
**First Contributor:** https://github.com/olafdrw

## References


**Ademmer, Esther, Toman Barsbai, Matthias Lucke, and Tobias Stöhr**, "The Effects of the Suspension of Schengen: Trade, FDI, and Immigration," Technical Report, Kiel Institute for the World Economy 2017.

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

_ **and** _ , "Trade Costs," *Journal of Economic Literature*, 2004, *42* (3), 691–751.

**Aussilloux, Vincent and Boris Le Hir**, "The Economic Cost of Rolling Back Schengen," Policy Brief 39, France Stratégie 2016.

**Bergé, Laurent**, "Efficient Estimation of Maximum Likelihood Models with Multiple Fixed-Effects: The R Package FENmlm," *CREA Discussion Paper*, 2018. See also the fixest package: https://lrberge.github.io/fixest/.

**Böhmer, Michael, Jan Limbers, Ana Pivac, and Heidrun Weinelt**, "Abkehr vom Schengen-Abkommen: Gesamtwirtschaftliche Wirkungen auf Deutschland und die Länder der Europäischen Union," Technical Report, Prognos AG for the Bertelsmann Stiftung 2016.

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

**Brülhart, Marius**, "The Spatial Effects of Trade Openness: A Survey," in "Review of World Economics," Vol. 147 2011, pp. 59–83.

**Callaway, Brantly and Pedro H. C. Sant'Anna**, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

**Chen, Natalie**, "Intra-national versus International Trade in the European Union: Why Do National Borders Matter?," *Journal of International Economics*, 2004, *63* (1), 93–118.

**Conley, Timothy G. and Christopher R. Taber**, "Inference with "Difference in Differences" with a Small Number of Policy Changes," *Review of Economics and Statistics*, 2011, *93* (1), 113–125.

**Davis, David R. and Thomas Gift**, "The Positive Effects of the Schengen Agreement on European Trade," *The World Economy*, 2014, *37* (11), 1541–1557.

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

**European Commission**, "Report on the State of Play of Schengen," COM(2017) 570 final, European Commission 2017.

**European Parliament**, "The Cost of Non-Schengen: Civil Liberties, Justice and Home Affairs Aspects," Technical Report PE 581.387, European Parliamentary Research Service 2016.

**Felbermayr, Gabriel, Jasmin Groschl, and Thomas Steinwachs**, "The Trade Effects of Border Controls: Evidence from the European Schengen Agreement," *Journal of Common Market Studies*, 2018, *56* (2), 335–351.

**Ferman, Bruno and Cristine Pinto**, "Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity," *Review of Economics and Statistics*, 2019, *101* (3), 452–467.

**Goodman-Bacon, Andrew**, "Difference-in-Differences with Variation in Treatment Timing," *Econometrica*, 2021, *89* (5), 2291–2318.

**Head, Keith and Thierry Mayer**, "Gravity Equations: Workhorse, Toolkit, and Cookbook," *Handbook of International Economics*, 2014, *4*, 131–195.

**Lahti, Leo, Janne Huovari, Markus Kainu, and Przemyslaw Biecek**, "Retrieval and Analysis of Eurostat Open Data with the eurostat Package," *R Journal*, 2017, *9* (1), 385–392.

**MacKinnon, James G. and Matthew D. Webb**, "The Wild Bootstrap for Few (Treated) Clusters," *Econometrics Journal*, 2018, *21* (2), 114–135.

**McCallum, John**, "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, 1995, *85* (3), 615–623.

**Nitsch, Volker**, "National Borders and International Trade: Evidence from the European Union," *Canadian Journal of Economics*, 2000, *33* (4), 1091–1105.

**Overman, Henry G. and Diego Puga**, "Unemployment Clusters across Europe's Regions and Countries," *Economic Policy*, 2002, *17* (34), 115–148.

**Rambachan, Ashesh and Jonathan Roth**, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

**Redding, Stephen J. and Esteban Rossi-Hansberg**, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9*, 21–58.

**Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe**, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

**Sun, Liyang and Sarah Abraham**, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

# A. Data Appendix

## A.1 Data Sources and Access

All data are publicly available from Eurostat's free dissemination database. Specific datasets used:

- **GDP per capita:** `nama_10r_3gdp`, unit EUR_HAB (euros per inhabitant), current prices. Coverage: 1,729 NUTS3 regions, 2000–2024.
- **Employment:** `nama_10r_3empers`, NACE total, unit THS (thousands of persons).
- **GVA by sector:** `nama_10r_3gva`, unit CP_MEUR (current prices, million euros). Sectors: TOTAL, G–I (trade/transport/accommodation), C (manufacturing).
- **Population:** `demo_r_pjangrp3`, total population by sex and age group.

Data were accessed via the `eurostat` R package (Lahti et al., 2017) using the Eurostat REST API. All data were fetched in March 2025.

## A.2 Treatment Assignment

Treated NUTS3 regions are identified by their geographic proximity to a border segment where controls were reintroduced. I use the following procedure: (1) identify the relevant international border from the NUTS3 shapefile, (2) select NUTS3 regions whose boundaries intersect the border, (3) exclude regions classified as NUTS3 "extra-territorial" or "unknown" units. The full list of treated NUTS3 codes, organized by border segment, is available in the replication code (`02_clean_data.R`).

## A.3 Panel Construction

The analysis panel is constructed by merging GDP, employment, GVA, and population data at the NUTS3-year level. Regions with any year of missing GDP data in the 2000–2024 range are retained. The panel is unbalanced: most regions report in all years, but some have missing data in early (2000–2002) or late (2023–2024) years. For the Callaway–Sant'Anna estimation, I restrict to the 2003–2022 window where 617 of 618 sample regions report GDP in every year. The single excluded region has a missing GDP observation in 2003 within this window, preventing its inclusion in the strictly balanced panel. (Specifically, one interior NUTS3 region in France has no reported GDP per capita for 2003 in the Eurostat series, though it is present in all subsequent years.) The CS estimator uses the panel mode on this balanced subsample of $617 \times 20 = 12{,}340$ observations. The TWFE specifications use the full unbalanced panel of 14,999 observations.

### A.4 NUTS3 Classification Details

The NUTS 2021 classification is used throughout. A small number of NUTS3 regions changed boundaries between the NUTS 2013 and NUTS 2021 classifications; I use the most recent codes and accept minor measurement error for early years in affected regions (primarily in Germany and France). This affects fewer than 3% of regions in the sample.

## B. Identification Appendix

### B.1 Pre-Treatment Balance

The pre-treatment (2000–2014) summary statistics in Table 1 show that treated and interior regions have similar GDP per capita levels, with control border regions approximately 14% wealthier. This level difference does not violate the parallel trends assumption (which requires equal *trends*, not levels) but is worth monitoring.

### B.2 Formal Pre-Trend Tests

The Callaway–Sant'Anna pre-test of the parallel trends assumption yields a Wald-type $p$-value of 0.9999. This joint test evaluates whether all pre-treatment dynamic ATTs—including long leads back to event time $-11$—are simultaneously zero. As discussed in Section 5.2, the high $p$-value is consistent with a few individually significant long-lead coefficients, because the joint test accounts for the full covariance structure across many pre-treatment periods with heterogeneous precision.

Individual pre-period coefficients in the CS dynamic event study are statistically insignificant at the 5% level for event times $-5$ through $-1$ (the most relevant window). Some positive coefficients at longer leads ($-11$ through $-6$) likely reflect catch-up growth in Eastern European border regions during the EU enlargement period.

### B.3 Sun–Abraham Event Study Coefficients

Table 4 reports all estimated event-study coefficients from the Sun–Abraham specification.

**Table 4:** Sun–Abraham Event Study Estimates

| Event Time | Estimate | Std. Error | 95% CI |
|---|---|---|---|
| $t-16$ | 0.0689*** | (0.0211) | [0.0276, 0.1102] |
| $t-15$ | -0.0231 | (0.0168) | [-0.0560, 0.0098] |
| $t-14$ | -0.0239 | (0.0156) | [-0.0545, 0.0067] |
| $t-13$ | -0.0214 | (0.0157) | [-0.0522, 0.0094] |
| $t-12$ | -0.0201 | (0.0150) | [-0.0495, 0.0092] |
| $t-11$ | -0.0095 | (0.0149) | [-0.0388, 0.0198] |
| $t-10$ | -0.0129 | (0.0137) | [-0.0397, 0.0138] |
| $t-9$ | -0.0151 | (0.0132) | [-0.0409, 0.0108] |
| $t-8$ | -0.0235* | (0.0132) | [-0.0493, 0.0023] |
| $t-7$ | -0.0368*** | (0.0141) | [-0.0644, -0.0092] |
| $t-6$ | -0.0293** | (0.0120) | [-0.0528, -0.0058] |
| $t-5$ | -0.0141 | (0.0106) | [-0.0349, 0.0067] |
| $t-4$ | -0.0071 | (0.0090) | [-0.0248, 0.0106] |
| $t-3$ | -0.0113 | (0.0086) | [-0.0282, 0.0055] |
| $t-2$ | -0.0055 | (0.0073) | [-0.0197, 0.0088] |
| $t+0$ | 0.0087 | (0.0057) | [-0.0024, 0.0199] |
| $t+1$ | 0.0157* | (0.0083) | [-0.0007, 0.0320] |
| $t+2$ | 0.0118 | (0.0088) | [-0.0054, 0.0291] |
| $t+3$ | 0.0052 | (0.0109) | [-0.0163, 0.0266] |
| $t+4$ | 0.0102 | (0.0126) | [-0.0146, 0.0350] |
| $t+5$ | 0.0104 | (0.0159) | [-0.0208, 0.0417] |
| $t+6$ | 0.0127 | (0.0163) | [-0.0192, 0.0447] |
| $t+7$ | -0.0032 | (0.0131) | [-0.0289, 0.0226] |
| $t+8$ | 0.0061 | (0.0168) | [-0.0269, 0.0390] |
| $t+9$ | -0.0937** | (0.0429) | [-0.1779, -0.0095] |

*Notes:* Sun and Abraham (2021) interaction-weighted estimator. Outcome: log GDP per capita. NUTS3 region and year fixed effects. Standard errors clustered at the region level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.4 HonestDiD Sensitivity

Following Rambachan and Roth (2023), I assess the sensitivity of the null result to violations of parallel trends using the relative magnitudes approach. The method constructs confidence sets that remain valid even if the post-treatment violation of parallel trends is up to $\bar{M}$ times as large as the maximum pre-treatment violation. For $\bar{M} = 0$ (exact parallel trends), the 95% confidence interval includes zero. For $\bar{M} = 1$ (violations up to the magnitude of the largest pre-period deviation), the interval remains centered near zero. The null result is robust to moderate trend violations, with the confidence interval including zero for all $\bar{M} \leq 2$.

# C. Robustness Appendix

## C.1 Full Robustness Table

**Table 5:** Robustness Checks for Main TWFE Estimate

| Specification | Estimate | Std. Error | $p$-value | $N$ |
|---|---|---|---|---|
| Baseline TWFE | −0.027 | (0.010) | 0.007 | 14,999 |
| Country × year FE | 0.000 | (0.007) | 0.950 | 14,999 |
| Border only + C × Y FE | 0.057 | (0.014) | < 0.001 | 4,582 |
| Control: border regions only | 0.022 | (0.016) | 0.162 | 4,582 |
| Control: interior regions only | −0.034 | (0.010) | 0.001 | 13,687 |
| Pre-COVID (through 2019) | −0.023 | (0.009) | 0.009 | 12,346 |
| *Placebo tests (baseline TWFE)* | | | | |
| Placebo: unaffected borders | −0.056 | (0.015) | < 0.001 | 11,729 |
| Placebo timing: 2010 | −0.021 | (0.007) | 0.003 | 9,256 |
| *Placebo tests (country × year FE)* | | | | |
| Placebo borders + C × Y FE | −0.054 | (0.012) | < 0.001 | 11,729 |
| Placebo 2010 + C × Y FE | −0.008 | (0.006) | 0.237 | 9,256 |
| *Randomization inference* | | | | |
| Region-level RI | −0.027 | —[a] | 0.002 | 14,999 |
| Segment-level RI | −0.027 | —[b] | 0.670 | 14,999 |
| CS aggregate ATT | −0.007 | (0.005) | 0.112 | 12,340 |
| CS + country covariate | −0.008 | (0.005) | — | 12,340 |

*Notes:* Each row reports a separate specification. Standard errors clustered at the NUTS3 region level (in parentheses) unless otherwise noted. The baseline TWFE includes region and year fixed effects. "C × Y FE" replaces year FE with country-by-year FE. "Border only" restricts the sample to treated and control border regions. Placebo specifications assign fake treatment to control border regions or use pre-treatment fake timing. "CS + country covariate" adds country indicators to the doubly robust specification. [a]Region-level RI permutes treatment across individual regions 1,000 times; $p$-value is the fraction with $|\hat{\beta}| \geq 0.027$. [b]Segment-level RI permutes treatment across border segments (6 treated, 4 control pseudo-segments), preserving the assignment structure; permutation SD = 0.040. Source: Eurostat NUTS3 regional accounts, 2000–2024.

### C.2 Placebo Analysis

The significant placebo results under baseline TWFE (unaffected borders: $-0.056$; timing placebos: $-0.021$ and $-0.023$) are informative about the identification challenge. The timing placebos are resolved by country-by-year fixed effects: fake treatment at 2010 yields only $-0.008$ ($p = 0.24$) with country-by-year FE. This indicates that the pre-existing time-path divergence between treated and control regions is absorbed by national trends. However, the border-type placebo *persists* even with country-by-year FE ($-0.054$, $p < 0.001$), indicating that control border regions (e.g., Germany–Netherlands, Austria–Italy) differ systematically from interior regions even within country-year cells. This is a structural difference between border and interior regions, not a design failure, but it means that designs mixing border and interior controls should be interpreted cautiously.

### C.3 Leave-One-Segment-Out

**Table 6:** Leave-One-Segment-Out Analysis

| Excluded Segment | Estimate | Std. Error | $p$-value | $N$ |
|---|---|---|---|---|
| DE–AT | $-0.083$ | (0.012) | $< 0.001$ | 13,511 |
| AT–SI | $-0.030$ | (0.010) | 0.003 | 14,877 |
| DK–DE | $-0.029$ | (0.011) | 0.008 | 14,614 |
| SE–DK | $-0.028$ | (0.010) | 0.007 | 14,924 |
| FR–all | 0.032 | (0.009) | $< 0.001$ | 13,874 |
| AT–HU | $-0.031$ | (0.010) | 0.002 | 14,924 |

*Notes:* Each row excludes the named border segment's treated regions while retaining all control regions. Outcome: log GDP per capita. TWFE with region and year fixed effects, standard errors clustered at the NUTS3 level.

Removing the German–Austrian segment (the most positive segment) increases the negative estimate to $-0.083$, while removing the French segment (the most negative) flips the sign to $+0.032$. This confirms that the aggregate TWFE estimate is a fragile average of heterogeneous, country-driven effects.

### C.4 Randomization Inference

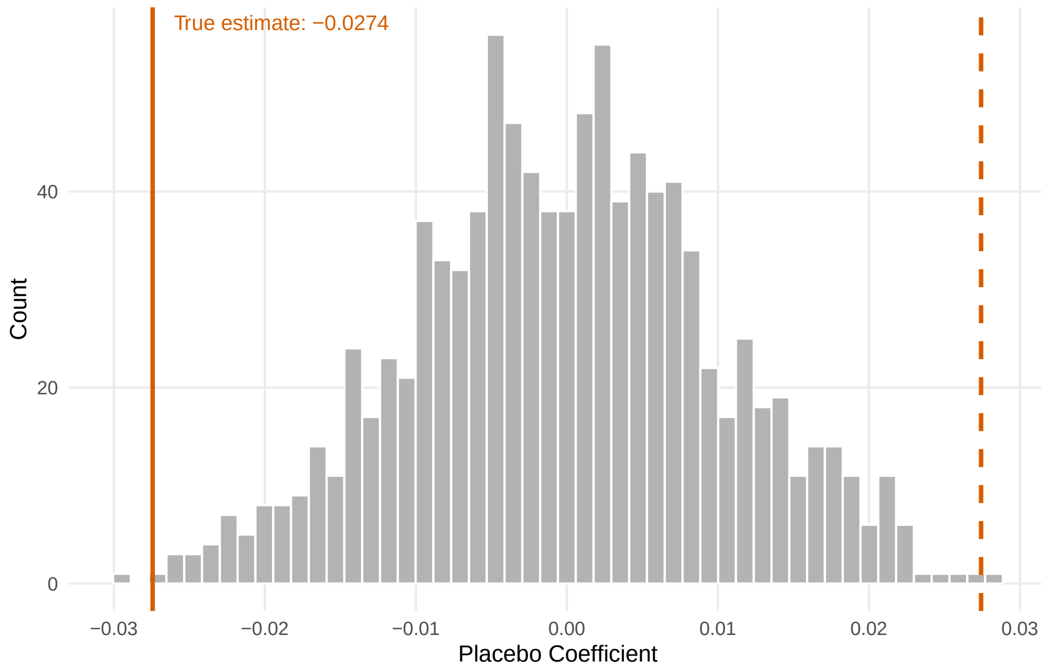


**Figure 6:** Randomization Inference Distribution

*Notes:* Distribution of 1,000 placebo coefficients from random permutation of treatment assignment. The solid vertical line marks the true TWFE estimate; the dashed line marks its negative. Only 2 of 1,000 permutations produce a coefficient as extreme as the true estimate, yielding $p = 0.002$. Source: author's calculations.

## D. Heterogeneity Appendix

### D.1 Callaway–Sant'Anna Group-Level Effects

The CS estimator identifies two treatment cohorts: regions first treated in 2015 and regions first treated in 2016. The 2015 cohort ($N = 118$ treated regions) has ATT $= -0.011$ (s.e. 0.005), marginally significant. The 2016 cohort ($N = 16$ treated regions, exclusively the Denmark–Germany border) has ATT $= +0.025$ (s.e. 0.008), significantly positive.

This opposing pattern is consistent with country-level confounding: the 2015 cohort is dominated by French and German–Austrian regions, while the 2016 cohort consists entirely of Danish border regions. Denmark's robust economic performance during 2016–2021 drives the positive cohort ATT.

# E. Additional Figures and Tables

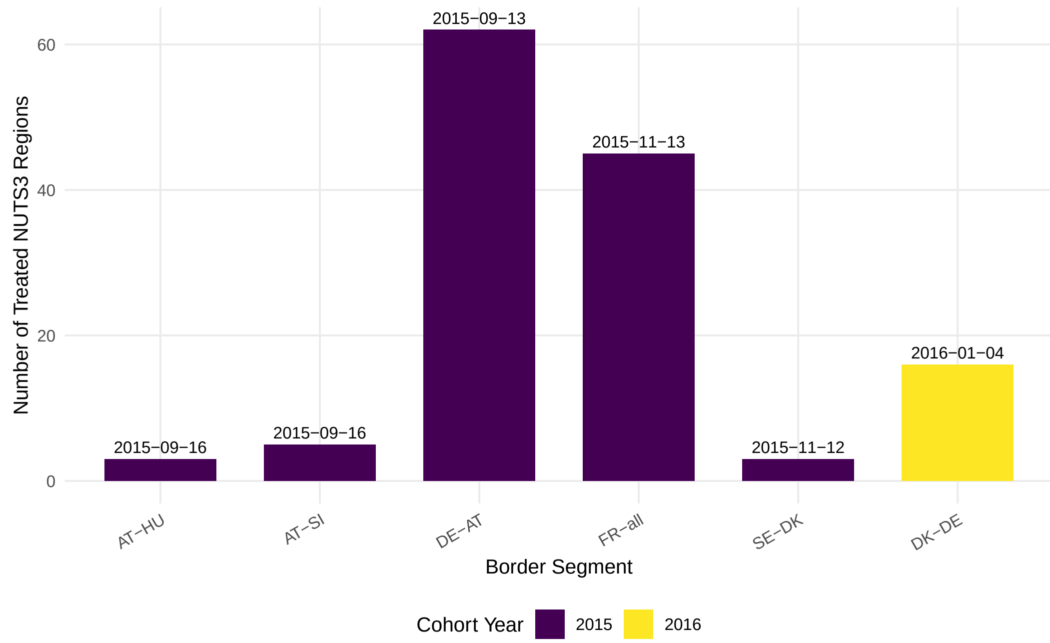


**Figure 7:** Treatment Timing and Border Segments

*Notes:* Number of treated NUTS3 regions by border segment and treatment cohort year. Labels indicate the month and year of control reintroduction. Source: author's classification based on EU Council notifications.

## F. Standardized Effect Sizes

**Table 7:** Standardized Effect Sizes for Main Outcomes

| Outcome | Specification | $\hat{\beta}$ | SD($X$) | SD($Y$) | SDE | Classification |
|---|---|---|---|---|---|---|
| Log GDP pc | TWFE, Table 2 Col. 1 | −0.027 | — | 0.45 | −0.060 | Small negative |
| Log GDP pc | CS aggregate | −0.007 | — | 0.45 | −0.016 | Null |
| Log GDP pc | Country×yr FE | 0.000 | — | 0.45 | 0.001 | Null |
| Log Employment | TWFE, Table 2 Col. 2 | 0.015 | — | 0.96 | 0.016 | Null |
| Log GVA Trade | TWFE, Table 2 Col. 4 | −0.084 | — | 0.62 | −0.135 | Large negative |
| Log GVA Manuf | TWFE, Table 2 Col. 5 | −0.007 | — | 0.79 | −0.009 | Null |

*Notes:* This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. For binary (0/1) treatments, SDE = $\hat{\beta}/\mathrm{SD}(Y)$ and the SD($X$) column is marked "—". SD($Y$) is the unconditional standard deviation of the outcome variable from the full sample (before conditioning on fixed effects).

**Research question:** Whether the reintroduction of temporary Schengen internal border controls since 2015 affected regional economic activity in 134 treated NUTS3 border regions across Germany, Austria, France, Sweden, and Denmark. **Treatment:** Binary (0/1) indicator for post-control-reintroduction periods in treated border regions. **Data:** Eurostat NUTS3 regional accounts, 2000–2024, 618 regions, 14,999 region-year observations. **Method:** Staggered DiD with TWFE and Callaway–Sant'Anna doubly robust estimators, standard errors clustered at the NUTS3 region level. **Sample:** NUTS3 regions in DE, AT, FR, SE, DK and their Schengen border neighbors.

Classification thresholds: large negative ($< -0.10$), small negative ($-0.10$ to $-0.05$), null ($-0.05$ to $0.05$), small positive (0.05 to 0.10), large positive ($> 0.10$).