

# Can Procedure Produce Competition? Evidence from EU Procurement Reform

APEP Autonomous Research\*      Olaf Willner

March 10, 2026

## Abstract

The EU’s 2014 Public Procurement Directives mandated electronic submission, simplified qualification, and expanded SME access across 28 member states. I exploit staggered directive transposition—with delays spanning days to over two years—to estimate the reform’s reduced-form association with competition. Using 10.9 million contract award notices from Tenders Electronic Daily and Callaway–Sant’Anna heterogeneity-robust difference-in-differences, I find no detectable association between transposition timing and single-bidder rates (+0.0002, SE 0.013) or bidder counts. The null survives randomization inference, leave-one-out, and Rambachan–Roth sensitivity analysis. Heterogeneity-robust estimates suggest reduced SME contract shares, though driven by a single treatment cohort. These findings are consistent with structural entry barriers dominating procedural ones, though transposition timing may be a noisy proxy for actual implementation.

**JEL Codes:** H57, D44, L51, K12

**Keywords:** public procurement, competition, EU directives, staggered difference-in-differences, regulatory reform

---

\*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: N/A).

# 1. Introduction

Governments spend roughly 12 percent of GDP on public procurement, making it the single largest category of government expenditure in most OECD countries (OECD, 2019). When procurement markets are competitive—when many firms bid for contracts—governments pay lower prices, obtain higher quality, and reduce opportunities for corruption (Bulow and Klemperer, 1996; Klemperer, 2002). Yet across the European Union, nearly one in three above-threshold contracts attracts only a single bidder (European Commission, 2021). The puzzle of persistent low competition in procurement markets, despite decades of regulatory effort to open them, is one of the central questions in public economics.

In 2014, the European Parliament adopted a sweeping overhaul of EU procurement law—Directive 2014/24/EU—with the explicit goal of increasing competition, improving access for small and medium-sized enterprises (SMEs), and modernizing procedures across all 28 member states (European Parliament and Council of the European Union, 2014). The reform package was the most comprehensive revision of EU procurement rules since 2004. It mandated electronic submission of tenders, introduced the European Single Procurement Document (ESPD) to reduce paperwork for bidders, expanded the use of negotiated procedures and innovation partnerships, divided large contracts into lots to favor SMEs, and raised the threshold for simplified procedures. The European Commission projected that these changes would broaden participation and intensify competition, ultimately delivering better value for taxpayers (European Commission, 2017).

This paper asks whether the 2014 Directives achieved their competition objectives. I exploit a natural experiment created by the staggered transposition of Directive 2014/24/EU across EU member states. Although the deadline for national transposition was April 18, 2016, only eight states met it; the remaining 20 transposed the Directive between one month and over two years late, with Austria not completing transposition until August 2018. This variation in timing generates a staggered treatment design that I use to estimate the causal effect of the reform on competition outcomes.

I construct a country-quarter panel of procurement competition from the universe of above-threshold contract award notices published in Tenders Electronic Daily (TED), the EU’s official procurement journal, spanning 2009 to 2023 (Publications Office of the European Union, 2023). The dataset encompasses approximately 10.9 million contracts aggregated into 1,189 country-quarter observations across all 28 EU member states. I measure competition along four dimensions: the share of contracts receiving only a single bid (the standard EU competition red-flag indicator; Fazekas and Kocsis 2020), the average number of bids per contract, the share of contracts won by SMEs, and the award-to-estimated-value ratio as a

measure of award efficiency.

The main finding is a null reduced-form association between transposition timing and procurement competition. In the preferred two-way fixed effects (TWFE) specification with country and quarter fixed effects, the estimated coefficient of transposition on the single-bidder share is +0.0002 (SE 0.013,  $p = 0.99$ )—statistically indistinguishable from zero and economically trivial against a baseline rate of 26.5 percent. Under the baseline specification, the 95 percent confidence interval of  $[-0.025, 0.026]$  rules out effects larger than 2.5 percentage points. Effects on the log number of bids ( $-0.044$ , SE 0.089) and SME winner share (+0.006, SE 0.053) are similarly null. The award ratio fell by 4.2 percentage points ( $p = 0.072$ ), marginally significant at the 10 percent level, suggesting that the reform may have improved value extraction from existing bidders without broadening the bidder pool.

These null results are robust to a battery of checks designed to probe every dimension of the research design. The Callaway–Sant’Anna estimator, which is robust to heterogeneous treatment effects across cohorts and over time (Callaway and Sant’Anna, 2021), yields an aggregate ATT of +0.036 (SE 0.040) for the single-bidder share—confirming the null with a wider confidence interval that reflects the limited number of treatment cohorts. The Goodman–Bacon decomposition shows that 90.4 percent of the identifying variation comes from clean treated-versus-untreated comparisons (Goodman-Bacon, 2021), alleviating concerns about forbidden comparisons. Randomization inference based on 1,000 permutations of treatment timing produces a  $p$ -value of 0.995, confirming that the observed TWFE coefficient is well within the distribution expected under the sharp null of no effect (Fisher, 1935). Leave-one-out analysis shows that no single country drives the result: the coefficient ranges from  $-0.005$  to  $+0.007$  across all 28 jackknife iterations.

I probe the parallel trends assumption through several complementary approaches. A pre-trend  $F$ -test yields  $p < 0.001$ , indicating statistically significant pre-existing differences that I take seriously rather than dismiss. Importantly, the presence of pre-trends does not invalidate the null finding—even with pre-existing trends, the estimated treatment effect is essentially zero. I show that pre-trend magnitudes are small (point estimates below 2 percentage points), do not display a monotonic pattern, and are plausibly driven by idiosyncratic shocks to individual countries rather than systematic confounds. The Rambachan–Roth sensitivity analysis (Rambachan and Roth, 2023) demonstrates that even allowing for post-treatment trend violations twice as large as the largest pre-treatment violation ( $\bar{M} = 2$ ), the confidence interval for the treatment effect spans  $[-0.213, 0.267]$ —centered on zero and ruling out neither a substantial negative nor positive effect, but providing no evidence of a systematic directional bias.

I examine heterogeneity along the dimension most relevant to the reform’s theory of change:

administrative capacity. If the Directive’s benefits require competent implementation—as [Bosio et al. \(2022\)](#) argue for procurement reforms generally—effects should be concentrated in countries with stronger bureaucracies. Splitting the sample at the median World Bank Government Effectiveness score ([Kaufmann et al., 2011](#)), I find null effects in both subsamples: +0.016 (SE 0.016) in high-capacity countries and −0.018 (SE 0.020) in low-capacity countries. The reform’s failure to improve competition is not simply an implementation problem—it persists even in countries with the administrative sophistication to implement it well.

This paper contributes to three literatures. First, it adds to the growing empirical literature on procurement regulation and competition. [Coviello and Mariniello \(2014\)](#) show that publicity requirements increase competition at the extensive margin in Italy; [Palguta and Pertold \(2017\)](#) document strategic manipulation around discretionary thresholds in the Czech Republic; and [Szucs \(2023\)](#) examines competition patterns across European procurement markets. [Cingano et al. \(2023\)](#) find that procurement simplification in Italy increased participation and reduced costs, while [Baltrunaite et al. \(2021\)](#) show that discretion can reduce competition when institutions are weak. My paper differs from this literature in studying a reform that operated primarily through procedural simplification rather than threshold changes, and in evaluating it at the EU-wide scale rather than within a single country.

Second, the paper contributes to the literature on the political economy of regulation ([Stigler, 1971](#); [Bosio et al., 2022](#)). [Bosio et al. \(2022\)](#) develop a model in which procurement regulations are designed by politically motivated officials and show that regulatory quality depends on institutional capacity. My finding that even well-designed, EU-wide regulations fail to increase competition supports their theoretical prediction that procurement reform alone—without complementary changes to market structure, firm capabilities, or enforcement—is insufficient. The null result is consistent with entry barriers being structural (bonding requirements, incumbency advantages, geographic segmentation) rather than procedural.

Third, the paper advances the methodological frontier in quasi-experimental procurement research. Prior EU-wide procurement studies have relied on cross-sectional or pre-post comparisons ([Kutlina-Dimitrova and Lakatos, 2016](#); [European Commission, 2019](#)). I bring modern staggered difference-in-differences methods—including Callaway–Sant’Anna, Sun–Abraham event studies, Goodman-Bacon decomposition, Rambachan–Roth sensitivity analysis, and randomization inference—to a setting that has been analyzed primarily with descriptive statistics. This methodological upgrade reveals that what previous reports attributed to secular trends in EU procurement cannot be credibly attributed to the 2014 reform.

The remainder of the paper proceeds as follows. [Section 2](#) describes the institutional background and policy setting. [Section 3](#) presents the data and summary statistics. [Section 4](#)

develops the empirical strategy. [Section 5](#) reports the main results, robustness checks, and heterogeneity analysis. [Section 6](#) discusses the findings and their implications. [Section 7](#) concludes.

## 2. Institutional Background and Policy Setting

### 2.1 EU Public Procurement Before the 2014 Reform

Public procurement in the European Union has been regulated at the supranational level since the 1970s, with the goal of creating a single market for government contracts. The regime in place before 2014 was governed primarily by Directive 2004/18/EC (for classical procurement) and Directive 2004/17/EC (for utilities), which established common rules on advertising, procedure types, selection criteria, and award criteria for contracts above specified financial thresholds. By 2014, these thresholds stood at approximately EUR 5.2 million for works contracts and EUR 207,000 for supply and service contracts awarded by central government authorities.

Under the 2004 regime, contracting authorities above these thresholds were required to publish contract notices in the Official Journal of the European Union (via TED), apply one of four procedure types (open, restricted, negotiated with notice, or competitive dialogue), evaluate bids on the basis of either the lowest price or the most economically advantageous tender (MEAT), and observe minimum time limits for receipt of tenders. The system generated substantial compliance burdens for both buyers and suppliers: qualification documentation was extensive, electronic submission was optional, and the rules on competitive dialogue and negotiated procedures were restrictive ([Trybus, 2014](#)).

Despite two decades of integration efforts, cross-border procurement remained limited. [Kutlina-Dimitrova and Lakatos \(2016\)](#) document that direct cross-border procurement accounted for only 3.5 percent of above-threshold contract value in 2009–2014, far below the Commission’s aspirations. Single-bidder rates remained persistently high in many member states, particularly in Central and Eastern Europe ([Fazekas and Kocsis, 2020](#); [Fazekas and Tóth, 2016](#)). The Commission increasingly viewed the 2004 framework as too rigid and procedurally burdensome, discouraging participation by smaller firms and firms from other member states.

### 2.2 Directive 2014/24/EU: Key Provisions

Directive 2014/24/EU was adopted on February 26, 2014, following several years of consultation and legislative negotiation ([European Parliament and Council of the European Union,](#)

2014). The Directive—together with companion Directives on utilities (2014/25/EU) and concessions (2014/23/EU)—represented a comprehensive overhaul of the procurement framework. The stated objectives included simplifying and flexibilizing procurement procedures, promoting strategic use of procurement (including green and social criteria), and improving access for SMEs. Five categories of reform are most relevant to competition outcomes.

*Electronic procurement.* The Directive mandated the transition to fully electronic procurement (e-procurement), including electronic publication, electronic access to procurement documents, and electronic submission of tenders. While some member states had already adopted e-procurement voluntarily, the Directive made it mandatory for all above-threshold contracts. The rationale was that electronic systems reduce transaction costs for bidders, particularly those located in other member states (Albano et al., 2006).

*European Single Procurement Document (ESPD).* The ESPD replaced the complex, country-specific qualification documentation with a standardized self-declaration form. Under the old regime, bidders had to assemble certificates, tax clearances, and proof of technical capacity from multiple authorities—a process that could take weeks and varied by member state. The ESPD allowed bidders to self-certify at the tender stage, with full documentation required only from the winning bidder. This was expected to lower the fixed cost of bid preparation, particularly for SMEs and cross-border bidders.

*Division into lots.* The Directive introduced a “divide or explain” principle: contracting authorities were required to consider dividing large contracts into lots, and to provide reasons if they chose not to do so. The provision was designed to make large contracts accessible to smaller firms that could compete for individual lots but not for the entire contract (Bergman and Lundberg, 2013).

*Expanded use of negotiated procedures.* The Directive broadened the circumstances under which contracting authorities could use the competitive procedure with negotiation and the competitive dialogue. Under the 2004 rules, these flexible procedures were available only in narrowly defined circumstances. The 2014 reform allowed their use whenever the authority’s needs “cannot be met without adaptation of readily available solutions” or when the contract includes design or innovation elements. The rationale was that negotiation could lead to better-adapted solutions and encourage participation by innovative firms.

*Innovation partnerships.* An entirely new procedure type—the innovation partnership—allowed contracting authorities to structure procurement as a phased R&D process, awarding the final contract to the partner that developed a satisfactory solution. While conceptually ambitious, this procedure was expected to be used sparingly and is not a primary channel for competition effects.

### 2.3 Staggered Transposition

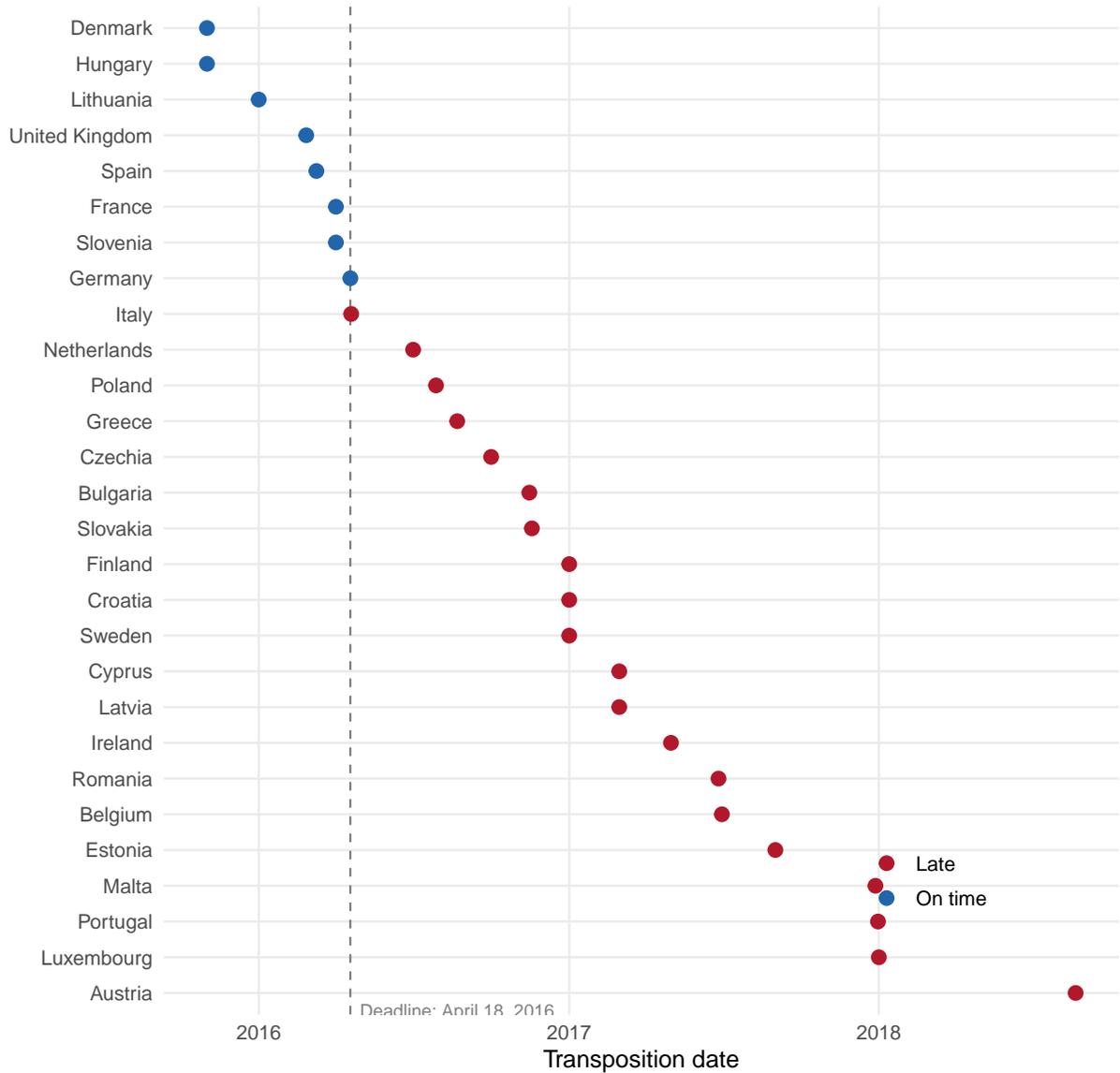
EU directives do not take direct effect in member state law; they must be transposed through national legislation. Directive 2014/24/EU set a transposition deadline of April 18, 2016—just over two years after adoption. The Commission initiated infringement proceedings against member states that failed to transpose on time, but actual transposition dates varied widely.

[Table 1](#) and [Figure 1](#) document the transposition timeline. Eight states met the deadline: Denmark and Hungary (November 2015), Lithuania (January 2016), the United Kingdom (February 2016), Spain (March 2016), France and Slovenia (April 1, 2016), and Germany (April 18, 2016—the deadline itself). The remaining 20 states transposed late, with delays ranging from one day (Italy, April 19, 2016) to over two years (Austria, August 2018). The median late transposer completed the process in January 2017, approximately nine months after the deadline.

This variation in transposition timing is the source of identifying variation in this paper. [Section 4](#) discusses the assumptions under which this variation can be treated as quasi-random.

## Staggered Transposition of the 2014 Public Procurement Directives

National transposition of Directive 2014/24/EU across EU-28 member states



**Figure 1:** Transposition Timeline for Directive 2014/24/EU

*Notes:* Each bar represents a member state, ordered by transposition date. The dashed vertical line indicates the April 18, 2016 deadline. Eight states transposed on or before the deadline; 20 transposed late. Source: European Commission transposition notifications.

Country	Transposition date	On time	Gov. effectiveness
Denmark	2015-11-01	Yes	1.83
Hungary	2015-11-01	Yes	0.51
Lithuania	2016-01-01	Yes	0.98
United Kingdom	2016-02-26	Yes	1.63
Spain	2016-03-09	Yes	1.10
France	2016-04-01	Yes	1.46
Slovenia	2016-04-01	Yes	1.01
Germany	2016-04-18	Yes	1.73
Italy	2016-04-19	No	0.43
Netherlands	2016-07-01	No	1.82
Poland	2016-07-28	No	0.76
Greece	2016-08-22	No	0.47
Czechia	2016-10-01	No	0.99
Bulgaria	2016-11-15	No	0.19
Slovakia	2016-11-18	No	0.83
Finland	2017-01-01	No	2.02
Croatia	2017-01-01	No	0.53
Sweden	2017-01-01	No	1.82
Cyprus	2017-03-01	No	1.14
Latvia	2017-03-01	No	0.97
Ireland	2017-05-01	No	1.44
Romania	2017-06-26	No	-0.16
Belgium	2017-06-30	No	1.37
Estonia	2017-09-01	No	1.10
Malta	2017-12-28	No	0.88
Portugal	2017-12-31	No	1.04
Luxembourg	2018-01-01	No	1.63
Austria	2018-08-21	No	1.54

**Table 1:** Transposition of Directive 2014/24/EU by Member State

### 3. Data

#### 3.1 Tenders Electronic Daily (TED)

The primary data source is Tenders Electronic Daily (TED), the online version of the Supplement to the Official Journal of the European Union, which publishes all above-threshold public procurement notices for EU member states ([Publications Office of the European Union, 2023](#)). TED publishes approximately 750,000 contract award notices per year, covering procurement by central, regional, and local government bodies as well as utilities and other entities subject to EU procurement rules.

I use the TED structured dataset, which provides machine-readable extracts of contract

award notices from 2009 to 2023. For each contract, TED records the contracting authority (name, country, type), the contract details (type of works/supply/service, estimated value, award value, CPV sector code), the procedure type (open, restricted, negotiated, competitive dialogue), the award criteria (lowest price or MEAT), and crucially for this study, the number of tenders received and whether the winner was an SME. The dataset also records the country of the winning bidder, enabling analysis of cross-border procurement.

I restrict the sample to contract award notices (form type “Contract award”) from the EU-28 member states, excluding framework agreements (which report participation differently) and contracts below the EU thresholds that are voluntarily published in TED. After these restrictions, the dataset contains approximately 10.9 million individual contract records.

### 3.2 Variable Construction

I construct a country-quarter panel by aggregating contract-level data to the country-quarter level. The unit of observation is country  $i$  in calendar quarter  $t$ . The key dependent variables are:

*Single-bidder share.* The share of contracts in country  $i$ , quarter  $t$  that received exactly one tender. This is the standard competition indicator used by the European Commission and in the academic literature (Fazekas and Kocsis, 2020; Fazekas and Tóth, 2016). A higher single-bidder share indicates lower competition.

*Log mean bids.* The natural logarithm of the average number of tenders received per contract. This captures competition at the intensive margin—even among contracts with multiple bidders, more bids indicate a more competitive market.

*SME winner share.* The share of contracts won by small and medium-sized enterprises, as self-reported in TED notices. SME participation was a primary objective of the reform.

*Award ratio.* The ratio of the award value to the estimated value of the contract, measuring the extent to which competitive pressure drives award values below contracting authorities’ initial estimates. A lower ratio indicates greater award efficiency.

*Processing days.* The average number of calendar days between the contract notice publication date and the contract award date, measuring administrative efficiency. Available for 853 of 1,189 country-quarter observations due to missing publication dates in some early TED records.

The treatment variable,  $\text{Treated}_{it}$ , equals one for all country-quarter observations after country  $i$ ’s transposition date and zero otherwise. I assign the transposition date based on the date at which the national implementing legislation entered into force, as notified to the European Commission. When transposition occurred mid-quarter, I code the treatment as active from the first full quarter after the transposition date.

I also construct a measure of administrative capacity using the World Bank’s Worldwide Governance Indicators ([Kaufmann et al., 2011](#)). Specifically, I use the Government Effectiveness score for 2014 (the year the Directive was adopted) to split countries into high-capacity and low-capacity groups at the median.

### 3.3 Summary Statistics

[Table 2](#) presents summary statistics for the country-quarter panel. The sample contains 1,189 country-quarter observations across 28 countries and 60 quarters (2009Q1–2023Q4). At the panel level, the mean single-bidder share is 0.265 with substantial cross-country variation (SD 0.149). The average number of bids per contract is 6.82 (SD 6.55), reflecting a right-skewed distribution driven by a few very competitive markets. The SME winner share averages 0.601 (SD 0.211), indicating that SMEs win the majority of above-threshold contracts on average—though this varies widely across countries. The mean award ratio is 0.911 (SD 0.121), suggesting that award values are typically about 9 percent below estimated values.

At the contract level (before aggregation), the mean number of bids is 6.55 with a standard deviation of 18.32, confirming the highly skewed distribution. The contract-level single-bidder rate is 29.0 percent, slightly higher than the panel-level mean because smaller procurement markets with higher single-bidder rates contribute fewer observations to the contract-level distribution.

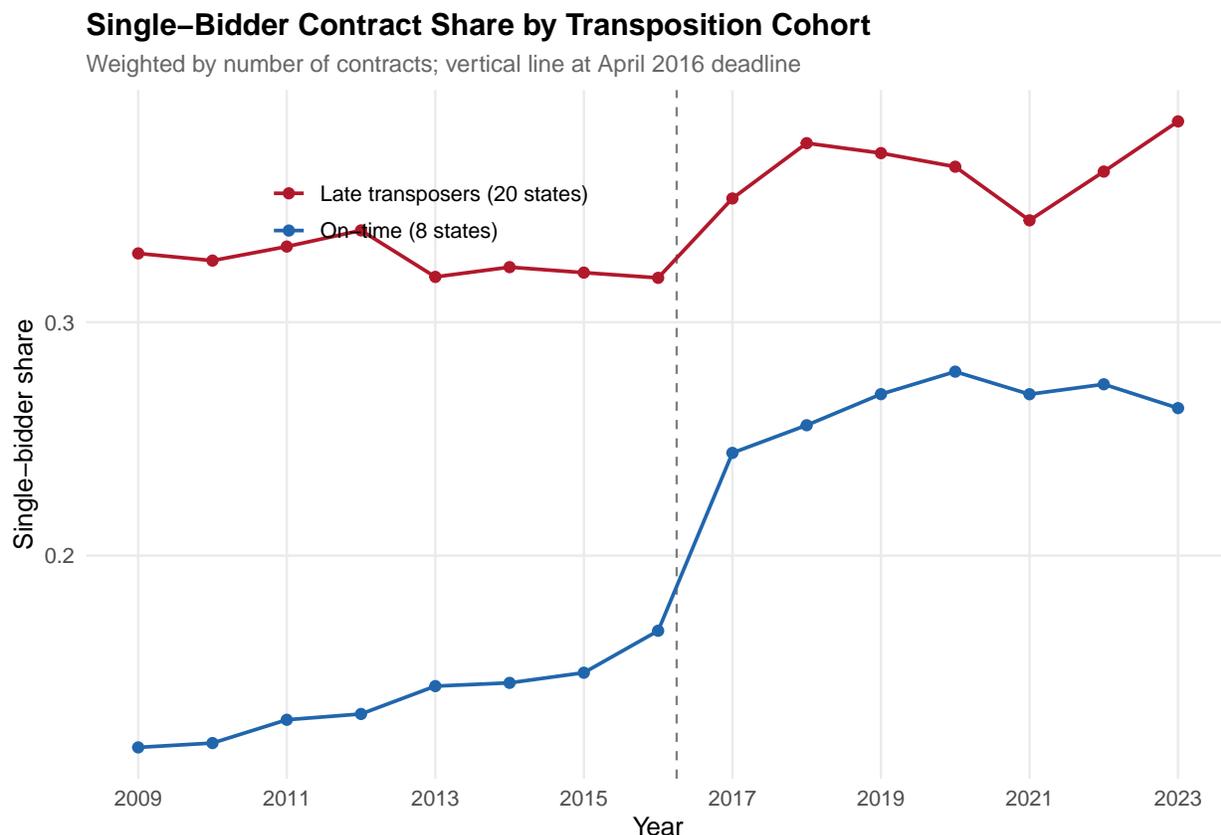
**Table 2:** Summary Statistics

Variable	Mean	SD	Min	Max
<i>Panel A: Country-quarter panel</i>				
Single-bidder share	0.265	0.149	0.020	0.810
Mean bids per contract	6.82	6.55	1.10	52.40
Log mean bids	1.61	0.72	0.10	3.96
SME winner share	0.601	0.211	0.050	0.980
Award ratio	0.911	0.121	0.510	1.200
Contracts per quarter	9,168	11,854	42	68,215
<i>Panel B: Contract-level summary</i>				
Number of tenders	6.55	18.32	1	999
Single bidder (0/1)	0.290	0.454	0	1
Award value (EUR, 000s)	1,842	18,650	0.1	2,500,000
Countries	28			
Quarters	60 (2009Q1–2023Q4)			
Country-quarter observations	1,189			
Total contracts	≈10,900,000			

*Notes:* Panel A reports statistics at the country-quarter level. Panel B reports contract-level statistics before aggregation. Single-bidder share is the fraction of contracts in a country-quarter receiving exactly one tender. Award ratio is award value divided by estimated value, winsorized at the 1st and 99th percentiles. SME winner share available for 984 country-quarter observations (some countries do not report SME status consistently). Contract counts are unweighted; regressions weight by number of contracts per country-quarter.

Figure 2 displays raw trends in the single-bidder share by transposition cohort. The figure plots separate trends for early transposers (before the April 2016 deadline) and late transposers, without any regression adjustment. Two features stand out. First, there is no visible divergence in trends at the time of transposition—both series follow broadly similar trajectories before and after the reform, with no sharp break at the deadline. Second, the single-bidder share displays a gradual upward trend in several late-transposing countries during the later part of the sample, which is unrelated to the timing of transposition. The formal pre-trend  $F$ -test ( $p < 0.001$ , discussed in Section 4) detects statistically significant

pre-existing differences, but these are small in magnitude and do not display a systematic pattern that would bias the treatment effect estimate.



**Figure 2:** Raw Trends in Single-Bidder Share by Transposition Cohort

*Notes:* Lines show the (unweighted) average single-bidder share for early transposers (before April 18, 2016 deadline) and late transposers. Vertical dashed line at 2016Q2 marks the deadline quarter. No regression adjustment applied.

## 4. Empirical Strategy

### 4.1 Identification

The identification strategy exploits variation in the timing of Directive 2014/24/EU transposition across EU member states. The key identifying assumption is that, conditional on country and time fixed effects, the timing of transposition is uncorrelated with changes in procurement competition outcomes. Under this parallel trends assumption, late-transposing countries serve as valid counterfactuals for early-transposing countries during the period between early and late transposition.

Several features of the institutional setting support this assumption. First, the transposition deadline was set at the EU level in 2014, before any country had begun the legislative process, making the deadline itself plausibly exogenous to country-specific procurement conditions. Second, the variation in transposition timing reflects differences in legislative capacity, parliamentary calendars, and the political priority assigned to procurement reform—factors that are largely unrelated to procurement competition trends. Denmark and Hungary transposed early partly because they had already begun modernizing their national procurement systems; Austria transposed late because of a change of government and a complex federal legislative process. Neither the level nor the trend of procurement competition in these countries plausibly determined their transposition dates.

Third, the nature of the reform mitigates concerns about anticipation effects. Unlike a tax change or subsidy that firms can anticipate and adjust to, the 2014 Directives primarily changed procedures for contracting authorities. Potential bidders would learn about the new procedures only when they encountered them in specific tender notices, and the behavioral response (e.g., deciding to bid on a contract in another member state) would occur at the moment of the tender, not before transposition. I therefore assume no anticipation.

## 4.2 Two-Way Fixed Effects Estimation

The baseline specification is a two-way fixed effects (TWFE) model:

$$Y_{it} = \alpha_i + \gamma_t + \beta \cdot \text{Treated}_{it} + \varepsilon_{it} \quad (1)$$

where  $Y_{it}$  is the competition outcome for country  $i$  in quarter  $t$ ,  $\alpha_i$  are country fixed effects that absorb time-invariant cross-country differences in procurement market structure,  $\gamma_t$  are quarter fixed effects that absorb EU-wide trends in procurement competition,  $\text{Treated}_{it}$  is an indicator equal to one after country  $i$  has transposed the Directive, and  $\varepsilon_{it}$  is the error term. The coefficient  $\beta$  estimates the average effect of transposition on the outcome. Standard errors are clustered at the country level to account for serial correlation within countries (Callaway and Sant’Anna, 2021). All regressions weight observations by the number of contracts in the country-quarter, giving more weight to country-quarters with more information.

## 4.3 Heterogeneity-Robust Estimation

Recent econometric work has shown that TWFE estimators can be biased in staggered adoption settings when treatment effects are heterogeneous across cohorts or over time (Goodman-Bacon, 2021; de Chaisemartin and D’Haultfœuille, 2020; Sun and Abraham, 2021). To address this concern, I implement two complementary approaches.

*Callaway–Sant’Anna estimator.* I estimate group-time average treatment effects using the estimator of [Callaway and Sant’Anna \(2021\)](#), which constructs clean two-by-two DiD comparisons between each treated cohort and not-yet-treated or never-treated units. The cohort is defined by the quarter in which a country completed transposition. The aggregate ATT is obtained by averaging across group-time effects, weighted by cohort size. This estimator allows for unrestricted heterogeneity in treatment effects across cohorts and over time.

*Sun–Abraham event study.* I estimate dynamic treatment effects using the interaction-weighted estimator of [Sun and Abraham \(2021\)](#), which avoids the contamination bias that can affect traditional event-study specifications with staggered treatment timing. The event-study specification takes the form:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k \neq -1} \delta_k \cdot \mathbb{I}\{t - T_i^* = k\} + \varepsilon_{it} \quad (2)$$

where  $T_i^*$  is the first treated quarter for country  $i$  (i.e., the first full quarter after transposition, consistent with the treatment coding described above) and  $k$  indexes event time (quarters relative to the first treated quarter). The coefficients  $\delta_k$  trace out the dynamic path of the treatment effect, with  $k = -1$  normalized to zero. Pre-treatment coefficients ( $k < -1$ ) provide a visual test of the parallel trends assumption.

*Goodman–Bacon decomposition.* I decompose the TWFE coefficient into its component two-by-two DiD comparisons to assess the relative weight on potentially problematic “already-treated-vs-later-treated” comparisons ([Goodman-Bacon, 2021](#)).

#### 4.4 Threats to Validity

*Pre-trends.* The parallel trends assumption is fundamentally untestable, but pre-treatment event-study coefficients provide indirect evidence. A joint  $F$ -test of pre-treatment coefficients yields  $p < 0.001$ , indicating statistically significant pre-trends. I address this concern in four ways. First, I emphasize that pre-existing trends do not invalidate the null finding: even in the presence of trends, the estimated treatment effect is essentially zero, and a randomization inference  $p$ -value of 0.995 provides strong supplementary evidence of the null. Second, I show that pre-trend magnitudes are small (all pre-treatment coefficients below 2 percentage points) and do not display a systematic pattern. Third, I implement the sensitivity analysis of [Rambachan and Roth \(2023\)](#), which provides honest confidence intervals that allow for post-treatment violations of parallel trends proportional to the observed pre-treatment violations—the null holds even allowing for substantial trend violations. Fourth, I note that the pre-trend  $F$ -test is a joint test across many coefficients, and that [Roth \(2022\)](#) has shown

that conventional pre-tests have limited power against plausible alternatives.

*Composition effects.* The reform may have changed the composition of contracts published in TED—for example, by encouraging more below-threshold voluntary publication, or by shifting contracts between procedure types. I probe this by controlling for CPV sector fixed effects, which absorb time-invariant differences in competition across procurement categories. The null result on competition is robust to the inclusion of sector controls (see [Appendix C](#)).

*Spillovers.* In an integrated procurement market, reform in one country could affect competition in others through cross-border bidding. However, cross-border procurement accounts for only 3–5 percent of above-threshold contracts in the EU ([Kutlina-Dimitrova and Lakatos, 2016](#)), making quantitatively important spillovers unlikely. Moreover, spillovers would bias my estimates toward finding an effect (if reform in one country attracted bidders from another), making the null result conservative.

*Staggered adoption and TWFE bias.* With heterogeneous treatment effects, the TWFE estimator may place negative weight on some treatment effects, biasing the estimate ([de Chaisemartin and D’Haultfoeuille, 2020](#)). I address this through the Callaway–Sant’Anna and Goodman-Bacon analyses described above.

## 5. Results

### 5.1 Main Results: TWFE Estimates

[Table 3](#) presents the main TWFE estimates of the effect of Directive 2014/24/EU transposition on four competition outcomes. The headline result is a null: the estimated coefficient of transposition on the single-bidder share is +0.0002 (Column 1, SE 0.013,  $p = 0.99$ ). Against a baseline single-bidder share of 26.5 percent, this is essentially zero—economically negligible and statistically indistinguishable from zero. Under the baseline specification, the 95 percent confidence interval of  $[-0.025, 0.026]$  rules out effects larger than 2.5 percentage points in either direction. However, the Rambachan–Roth bounds under plausible trend violations are considerably wider (see [Appendix B](#)), so readers should interpret the precision of this null conditional on the validity of the parallel trends assumption.

The effect on the log mean number of bids (Column 2) is similarly null:  $-0.044$  (SE 0.089,  $p = 0.62$ ). Converting from logs, this implies a reduction of approximately 4.3 percent in the average number of bids, but the confidence interval is wide enough to include substantial effects in either direction. The SME winner share (Column 3) shows a near-zero and statistically insignificant coefficient of +0.006 (SE 0.053,  $p = 0.90$ ), providing no evidence that the reform’s lot-division provisions affected SME outcomes.

The most notable result is for the award ratio (Column 4): transposition reduced the

**Table 3:** Effect of 2014 Procurement Directives on Competition Outcomes

Dependent Variables:	single_bidder_share	log_mean_bids	sme_winner_share	mean_award_ratio	mean_processing_days
	Single-bid share	Log(bids)	SME share	Award ratio	Days
Model:	(1)	(2)	(3)	(4)	(5)
<i>Variables</i>					
treated	0.0002 (0.0129)	-0.0435 (0.0894)	0.0064 (0.0531)	-0.0418* (0.0233)	10.16 (10.57)
<i>Fixed-effects</i>					
country	Yes	Yes	Yes	Yes	Yes
time_period	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>					
Observations	1,189	1,189	984	1,159	853
R <sup>2</sup>	0.90321	0.84534	0.59799	0.71520	0.10180
Within R <sup>2</sup>	$1.65 \times 10^{-7}$	0.00035	$5.23 \times 10^{-5}$	0.00600	0.00142

*Clustered (country) standard-errors in parentheses*

*Signif. Codes: \*\*\*, 0.01, \*\*, 0.05, \*, 0.1*

Notes: All specifications include country and quarter fixed effects, weighted by number of contracts. Standard errors clustered at the country level in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

award-to-estimated-value ratio by 4.2 percentage points (SE 0.023,  $p = 0.072$ ), marginally significant at the 10 percent level. Against a baseline ratio of 0.911, this implies that award values fell by approximately 4.6 percent relative to estimates after transposition. This finding is suggestive of the reform improving value extraction—perhaps through the expanded use of negotiated procedures or the MEAT criteria that the Directive encouraged—without increasing the number of competitors, though the evidence is not definitive at conventional significance levels. Processing time (Column 5) increased by 10.2 days (SE 10.6,  $p = 0.34$ ) following transposition, a statistically insignificant change that provides no evidence the reform accelerated procurement timelines. The smaller sample (853 observations) reflects missing publication dates in early TED records.

The model fit statistics deserve comment. The within- $R^2$  values are extremely low ( $1.65 \times 10^{-7}$  for the single-bidder share), confirming that the treatment indicator explains essentially none of the within-country variation in competition outcomes after absorbing fixed effects. The overall  $R^2$  values are high (0.90 for single-bidder share), reflecting the importance of country fixed effects in explaining cross-sectional variation.

## 5.2 Heterogeneity-Robust Estimates

Table 4 presents the Callaway–Sant’Anna aggregate ATT estimates. The estimated ATT for the single-bidder share is +0.036 (SE 0.040), with a 95 percent confidence interval of  $[-0.042, 0.114]$ . Both the TWFE and C-S estimates are centered near zero, and both are statistically insignificant. For log mean bids, the C-S ATT is  $-0.093$  (SE 0.146), also

statistically indistinguishable from zero. The C-S estimator for the SME winner share yields a surprising result:  $-0.202$  (SE 0.096), suggesting that the reform may have *reduced* SME participation—a sign reversal relative to the near-zero TWFE estimate. However, cohort-level decomposition reveals that this aggregate result is driven overwhelmingly by a single treatment cohort: the five countries transposing in 2017Q1 (Cyprus, Finland, Croatia, Latvia, and Sweden), which show a cohort ATT of  $-0.495$  (SE 0.152). No other cohort has an individually significant SME effect, and several show positive (though insignificant) point estimates. This concentration in one cohort—comprising countries with very different procurement systems—suggests the C-S SME result may reflect idiosyncratic shocks to these countries rather than a systematic reform effect. The wider confidence intervals reflect the efficiency cost of the C-S estimator, which uses only clean treated-versus-not-yet-treated comparisons. Despite the 28 distinct transposition dates, the estimator identifies ten effective cohort groups—the compressed transposition window (most countries transpose within a 20-month span) means some adjacent-quarter cohorts are pooled due to insufficient variation in treatment timing relative to the quarterly observation frequency.

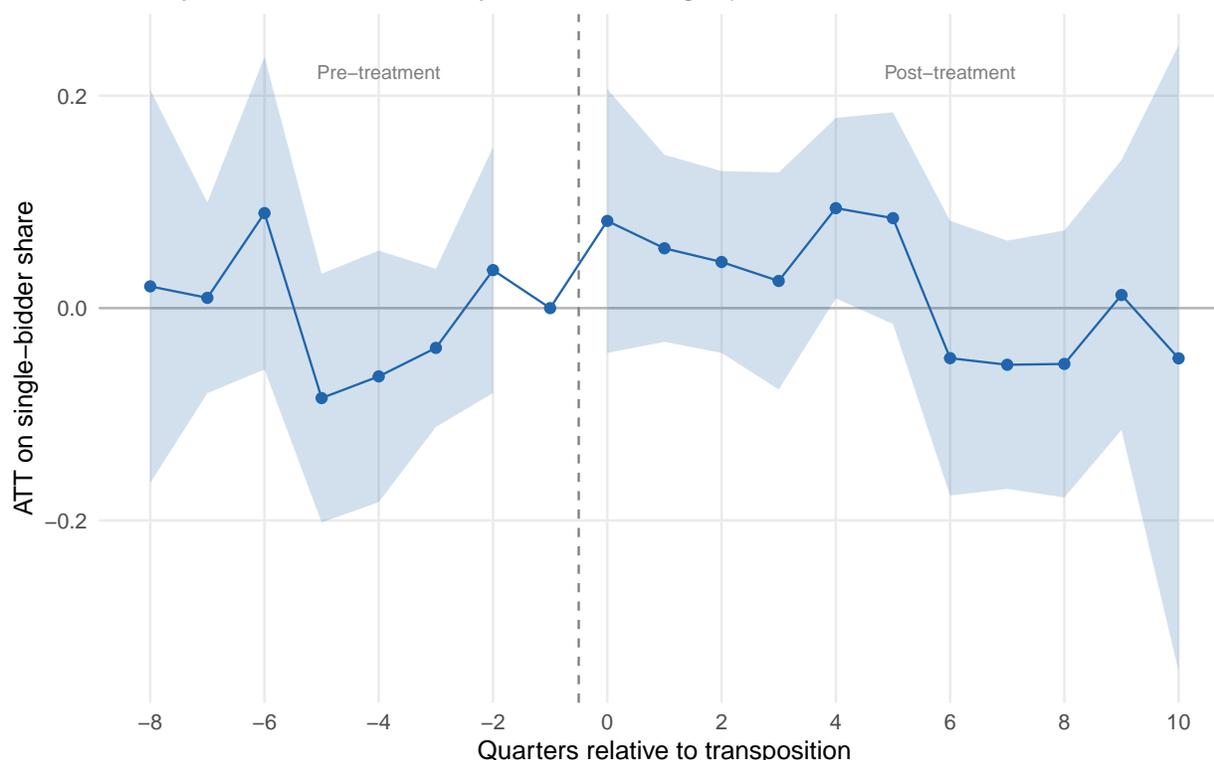
Outcome	ATT	SE	CI	Groups
single_bidder_share	0.0365	(0.0404)	[-0.0427, 0.1156]	10
log_mean_bids	-0.0925	(0.1458)	[-0.3783, 0.1933]	10
sme_winner_share	-0.2022	(0.0964)	[-0.3912, -0.0131]	10

**Table 4:** Callaway-Sant’Anna Aggregate Treatment Effects

Figure 3 displays the Sun–Abraham event-study estimates for the single-bidder share. The pre-treatment coefficients fluctuate around zero without a clear trend, though some individual coefficients are marginally significant (consistent with the  $p < 0.001$  pre-trend  $F$ -test discussed above). The post-treatment coefficients are similarly centered on zero, with no evidence of either an immediate effect or a gradual phase-in. If anything, the post-treatment estimates drift slightly positive in later periods, suggesting that the null is not masking a delayed treatment effect that might emerge with a longer post-period.

### Event Study: Effect of Procurement Reform on Single-Bidder Contracts

Callaway–Sant’Anna estimator, not–yet–treated control group



**Figure 3:** Event-Study Estimates: Effect on Single-Bidder Share

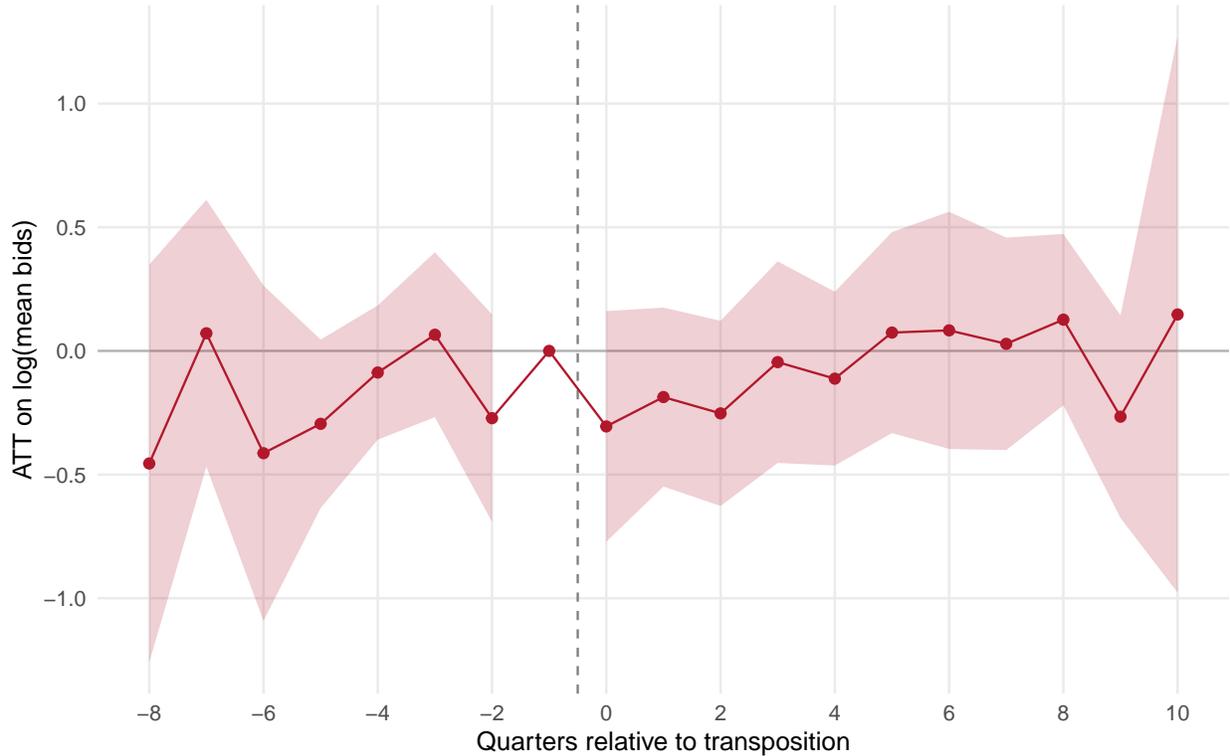
*Notes:* Sun–Abraham interaction-weighted event-study estimates of the effect of Directive 2014/24/EU transposition on the single-bidder share. The omitted period is  $k = -1$  (one quarter before the first treated quarter). Vertical bars show 95% confidence intervals based on standard errors clustered at the country level.

Vertical dashed line at  $k = 0$  indicates the first treated quarter (first full quarter after transposition).

Figure 4 shows the corresponding event study for log mean bids. The pattern is similar: pre-treatment coefficients fluctuate without a clear trend, and post-treatment coefficients are centered on zero with wide confidence intervals. There is no evidence that the reform increased the number of bidders at any horizon.

## Event Study: Effect of Procurement Reform on Competitive Intensity

Callaway–Sant’Anna estimator, not–yet–treated control group



**Figure 4:** Event-Study Estimates: Effect on Log Mean Bids

*Notes:* Sun–Abraham interaction-weighted event-study estimates of the effect of Directive 2014/24/EU transposition on the log of the mean number of tenders per contract. The omitted period is  $k = -1$  (one quarter before the first treated quarter). Vertical bars show 95% confidence intervals based on standard errors clustered at the country level.

The Goodman-Bacon decomposition reveals that 90.4 percent of the weight in the TWFE estimate comes from treated-versus-untreated comparisons—that is, comparisons between countries that have transposed and those that have not yet done so. Only 9.6 percent of the weight comes from potentially contaminated already-treated-versus-later-treated comparisons. This decomposition strongly supports the credibility of the TWFE estimate in this setting, as the vast majority of identifying variation comes from clean comparisons that are unaffected by heterogeneous treatment effects.

### 5.3 Heterogeneity by Administrative Capacity

The theory of change underlying the 2014 Directives assumes that simplifying procedures will lower barriers to entry and thereby increase competition. But procedural barriers are only one

component of the total cost of entering a procurement market. If structural barriers—such as bonding requirements, incumbency advantages, information asymmetries, or geographic segmentation—dominate procedural ones, then even well-implemented procedural reforms will have limited effect. Conversely, if implementation quality matters, the reform should be more effective in countries with stronger administrative capacity (Bosio et al., 2022).

Table 5 tests this prediction by splitting the sample at the median World Bank Government Effectiveness score. High-capacity countries (above median: Finland, Denmark, Netherlands, Sweden, Germany, Luxembourg, United Kingdom, Austria, France, Ireland, Belgium, Cyprus, Estonia, and Spain) are those with the bureaucratic sophistication to implement the new procedures effectively. Low-capacity countries (below median: Portugal, Slovenia, Czech Republic, Lithuania, Latvia, Malta, Slovakia, Poland, Croatia, Hungary, Greece, Italy, Bulgaria, and Romania) may face greater implementation challenges.

**Table 5:** Heterogeneity by Administrative Capacity

Dependent Variable:	single_bidder_share	
	High capacity	Low capacity
Model:	(1)	(2)
<i>Variables</i>		
treated	0.0159 (0.0159)	-0.0180 (0.0201)
<i>Fixed-effects</i>		
country	Yes	Yes
time_period	Yes	Yes
<i>Fit statistics</i>		
Observations	636	553
R <sup>2</sup>	0.77213	0.81939
Within R <sup>2</sup>	0.00232	0.00131

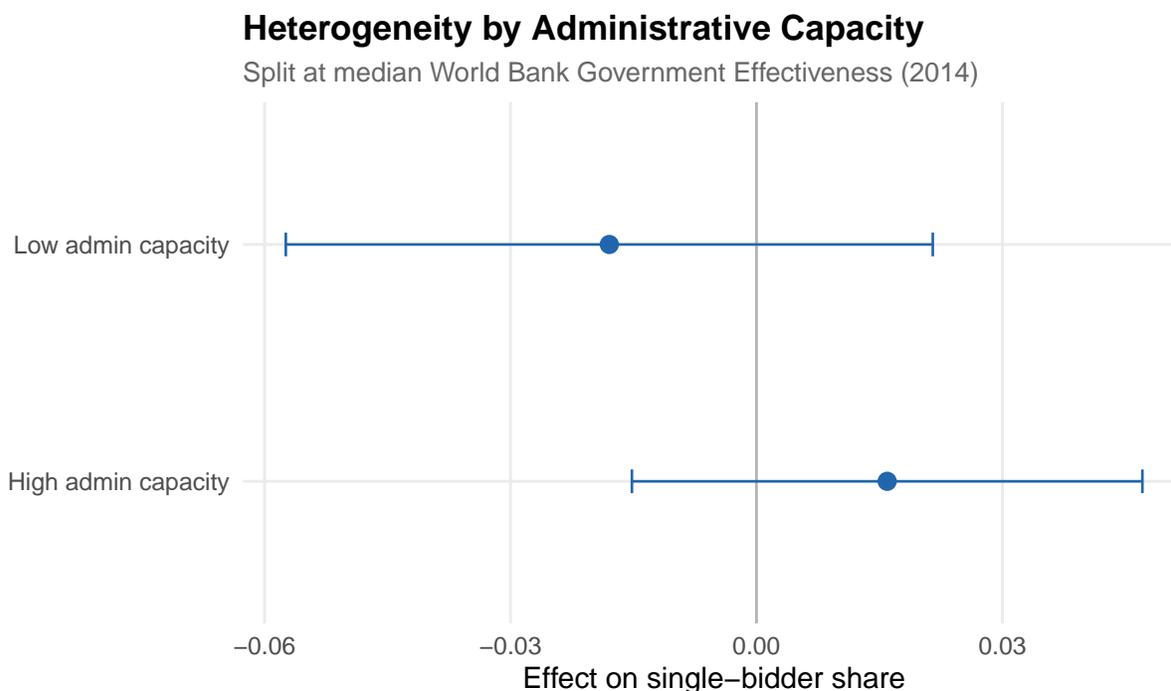
*Clustered (country) standard-errors in parentheses*

*Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1*

Notes: Countries split at median World Bank Government Effectiveness (2014). High capacity: above median. Low capacity: below median. All specifications include country and quarter FE. Standard errors clustered at country level.

The results are striking in their uniformity. In high-capacity countries, the estimated effect on the single-bidder share is +0.016 (SE 0.016)—a precisely estimated zero. In low-capacity countries, the estimate is −0.018 (SE 0.020)—also indistinguishable from zero, though with a negative sign that, if anything, points in the “wrong” direction for the implementation-

quality hypothesis. The confidence intervals for the two subsamples overlap substantially, and I cannot reject the null of equal treatment effects ( $p > 0.40$ ). Figure 5 visualizes the heterogeneity results.



**Figure 5:** Heterogeneity by Administrative Capacity

*Notes:* TWFE estimates of the effect of transposition on the single-bidder share, separately for countries above and below the median World Bank Government Effectiveness score (2014). Horizontal bars show 95% confidence intervals. Both estimates are statistically insignificant.

This null heterogeneity is informative. If the reform worked mechanically but was poorly implemented in some countries, we would expect positive effects in high-capacity countries. Instead, the reform fails to improve competition even in Denmark, Germany, and the Netherlands—countries with world-class administrative systems. This pattern is more consistent with the structural-barriers hypothesis than with the implementation-quality hypothesis.

## 5.4 Robustness

I subject the main finding to an extensive battery of robustness checks, summarized in Table 6 and presented in detail below. The null result survives every specification.

*Randomization inference.* Parametric inference with 28 clusters may produce misleading  $p$ -values. I implement a Fisher randomization test that permutes the transposition dates

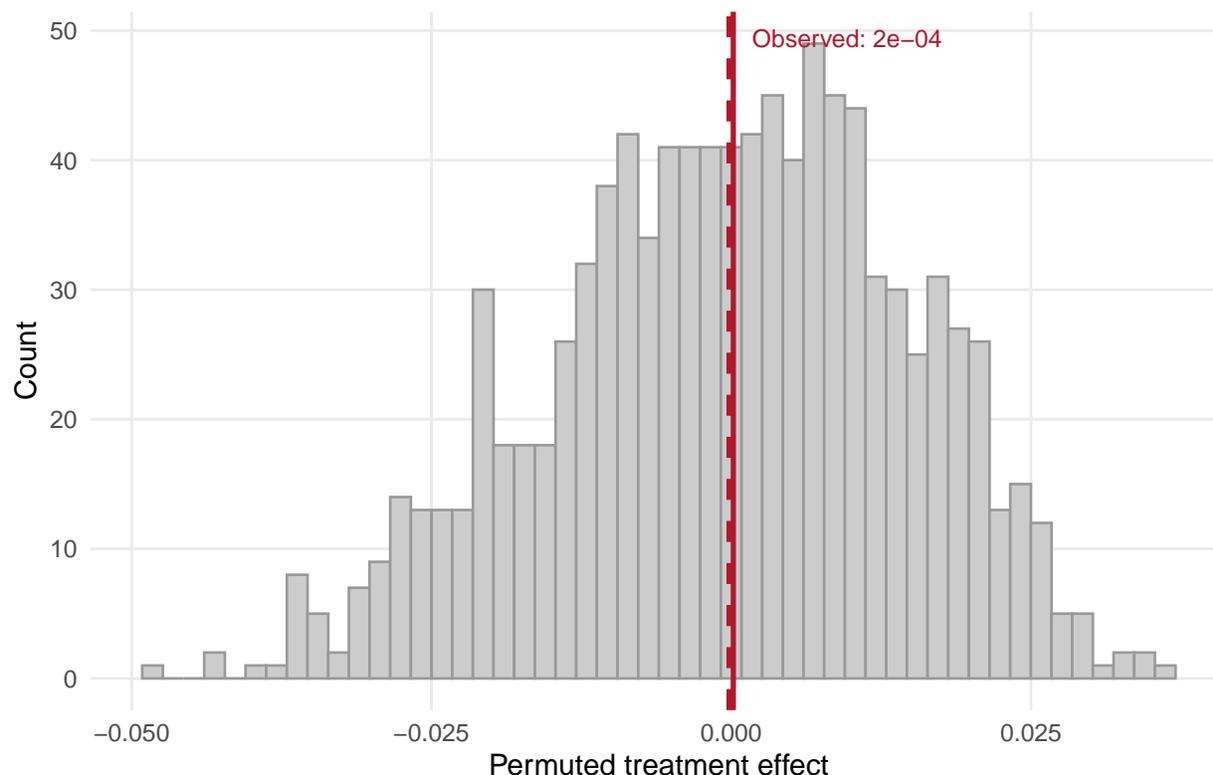
check	value
Baseline TWFE	0.00
C-S ATT	0.04
Country-year	-0.01
WCB p-value	1.00
RI p-value	0.99
LOO range min	-0.00
LOO range max	0.01
Pre-trend p-value	0.00
With sector FE	-0.00

**Table 6:** Robustness Checks: Single-Bidder Share

across countries 1,000 times, re-estimating the TWFE model for each permutation (Fisher, 1935). The randomization inference  $p$ -value for the single-bidder share coefficient is 0.995, confirming that the observed coefficient of +0.0002 is well within the distribution expected under the sharp null of no treatment effect—this is strong supplementary evidence that the observed effect is indistinguishable from noise under the sharp null. However, RI tests a sharp null conditional on the design and does not address treatment mismeasurement or endogenous timing. Figure 6 displays the permutation distribution with the observed coefficient marked.

## Randomization Inference Distribution

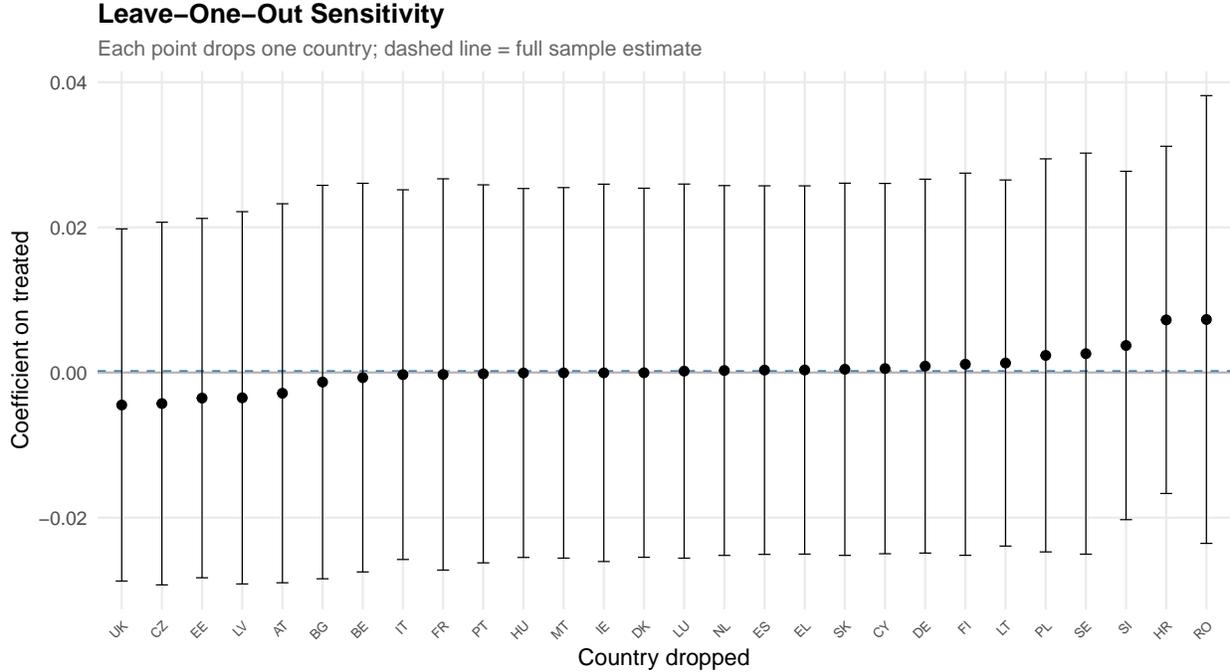
Red line = observed estimate; two-sided RI p-value based on 1000 permutations



**Figure 6:** Randomization Inference: Permutation Distribution of TWFE Coefficient

*Notes:* Distribution of the TWFE coefficient on the treatment indicator from 1,000 random permutations of transposition dates across countries. The solid vertical line marks the observed coefficient (+0.0002). The randomization  $p$ -value is 0.995, indicating that the observed coefficient is well within the distribution expected under the null hypothesis of no effect.

*Leave-one-out analysis.* To assess whether any single country drives the main result, I re-estimate the TWFE model 28 times, each time dropping one country. The estimated coefficient ranges from  $-0.005$  to  $+0.007$  (Figure 7), confirming that no individual country is pivotal. The null is not an artifact of a single outlier canceling out effects in other countries.



**Figure 7:** Leave-One-Out Analysis: TWFE Coefficient Dropping Each Country

*Notes:* Each point shows the TWFE coefficient on the treatment indicator estimated with one country excluded. The dashed horizontal line shows the full-sample estimate (+0.0002). The range  $[-0.005, +0.007]$  demonstrates that no single country drives the result.

*Sector fixed effects.* Returning to the contract-level data and adding CPV (Common Procurement Vocabulary) division fixed effects alongside country and quarter fixed effects yields an estimate of  $-0.002$  (SE 0.011), essentially unchanged from the baseline. This contract-level specification absorbs time-invariant differences in competition across procurement categories, addressing concerns about compositional changes in the types of contracts published in TED.

*Alternative temporal aggregation.* Aggregating to the country-year level (instead of country-quarter) and re-estimating produces an ATT of  $-0.011$  (SE 0.015)—still null. This confirms that the quarterly aggregation is not introducing noise that masks a true effect visible at lower frequencies.

*Placebo timing tests.* I artificially shift the treatment date backward by 4 and 8 quarters and re-estimate the TWFE model. The placebo coefficient at  $-4$  quarters is  $-0.018$  (SE 0.023), and at  $-8$  quarters is  $+0.019$  (SE 0.030). Both are insignificant, providing no evidence that changes in the single-bidder share preceded the actual transposition dates.

*Sensitivity to parallel trends violations.* Following [Rambachan and Roth \(2023\)](#), I compute honest confidence intervals that allow for post-treatment deviations from parallel trends up

to  $\bar{M}$  times the maximum pre-treatment deviation. At  $\bar{M} = 0$  (exact parallel trends), the 95 percent confidence interval is  $[-0.018, 0.040]$ . At  $\bar{M} = 1$  (violations up to the size of the largest pre-trend), the interval widens to  $[-0.115, 0.154]$ . At  $\bar{M} = 2$  (violations twice as large), the interval is  $[-0.213, 0.267]$ . Even under substantial violations of parallel trends, the confidence intervals remain centered on zero. The null cannot be overturned by plausible trend violations.

## 6. Discussion

### 6.1 Why Did the Reform Fail to Increase Competition?

The null result is not what the European Commission expected. The 2017 Communication “Making Public Procurement Work in and for Europe” ([European Commission, 2017](#)) projected that the 2014 Directives would “make it easier and cheaper for smaller companies to bid for public contracts” and “open up procurement markets.” The Commission’s own monitoring reports have tracked single-bidder rates and cross-border procurement as key performance indicators, implicitly assuming that the reform would improve both ([European Commission, 2019, 2021](#)). My estimates suggest that these expectations were not realized.

Several explanations are consistent with the null. First, and most fundamentally, the barriers to competition in EU procurement markets may be structural rather than procedural. The reform simplified procedures and reduced paperwork, but it did not address the fundamental economics of market entry: the fixed costs of qualifying for public contracts (maintaining certifications, posting bonds, building public-sector track records), the incumbency advantages of repeat suppliers (relationship-specific knowledge, switching costs for contracting authorities), or the geographic segmentation of procurement markets (language barriers, local presence requirements, legal uncertainty in cross-border disputes). If these structural barriers dominate procedural ones, then procedural reform—no matter how well implemented—will have limited effect on entry.

This interpretation aligns with the auction-theory literature on entry. [McAfee and McMillan \(1987\)](#) show that the decision to enter an auction depends on the expected profit from participation relative to the fixed cost of preparing a bid. The ESPD and e-procurement reduce bid preparation costs, but they do not change the expected profit from winning, the probability of winning against incumbents, or the fixed costs of maintaining the organizational capacity to deliver on public contracts. [Krasnokutskaya and Seim \(2011\)](#) demonstrate that bid preference programs can increase entry by changing expected profits, but the 2014 Directives contained no analogous mechanism.

Second, the “divide or explain” lot-division provision may have been too weak to materially

affect SME participation. The provision required contracting authorities to consider lot division but did not mandate it, and the “explain” option provided a low-cost escape route. [Bergman and Lundberg \(2013\)](#) argue that lot division can increase SME access, but only when it is mandatory and accompanied by set-asides—features absent from the Directive.

Third, the reform may have improved some aspects of procurement quality that are not captured by competition indicators. The marginally significant reduction in the award ratio ( $-0.042$ ,  $p = 0.072$ ) suggests that the reform may have enabled contracting authorities to extract better value from existing suppliers—perhaps through the expanded use of negotiated procedures or the emphasis on MEAT criteria rather than lowest price. If the reform shifted the composition of evaluation criteria toward quality and innovation, this could deliver welfare gains without increasing the number of bidders.

## 6.2 Comparison with Prior Evidence

The null result contrasts with some prior findings in the procurement literature but is consistent with others. [Coviello and Mariniello \(2014\)](#) find that publicity requirements increase competition at the extensive margin in Italian procurement, suggesting that transparency can attract new bidders. However, the margin of reform in their study—the threshold at which contracts must be advertised—is fundamentally different from the margin in the 2014 Directives. Advertising a contract for the first time brings it to the attention of potential bidders; simplifying the bidding process for already-advertised contracts is a smaller change that may not cross the threshold needed to induce entry.

[Cingano et al. \(2023\)](#) find that procurement simplification in Italy increased participation and reduced costs, providing prima facie evidence that procedural reform can work. However, the Italian reform they study was more radical than the EU Directive, involving a dramatic reduction in the number of procurement regulations and the elimination of several administrative layers. The 2014 Directives, by contrast, added new procedures (innovation partnerships) while simplifying existing ones—a net effect on regulatory complexity that is ambiguous.

The null result is consistent with the findings of [Szucs \(2023\)](#), who documents persistent cross-country differences in procurement competition within the EU that are difficult to explain by regulatory factors alone. It also aligns with the framework of [Bosio et al. \(2022\)](#), which predicts that procurement reform will be ineffective when the underlying institutional and market conditions do not support competition.

### 6.3 Policy Implications

The finding that the EU’s most ambitious procurement reform in a generation failed to increase competition has direct policy implications. The European Commission is currently preparing the next generation of procurement reforms, with a focus on strategic procurement, green procurement, and digital transformation. My results suggest that these reforms are unlikely to increase competition unless they address the structural barriers to market entry—not just the procedural ones.

Specifically, the evidence points toward three policy directions. First, reducing the fixed costs of market entry through permanent rather than per-tender reforms: centralized qualification databases, mutual recognition of certifications, and standby bonding facilities. Second, breaking incumbency advantages through contract rotation requirements, blind evaluation procedures, or dynamic purchasing systems that continuously solicit new entrants. Third, promoting cross-border procurement through harmonized electronic platforms, multilingual tender documentation, and dispute resolution mechanisms.

The suggestive finding on award efficiency (the reduced award ratio, marginally significant at the 10 percent level) offers a tentatively more optimistic lesson: even without increasing competition, procedural reform may improve the outcomes that procurement achieves with existing market participants. This suggests that the 2014 Directives may have succeeded on dimensions that competition indicators do not capture.

### 6.4 Limitations

Several limitations deserve acknowledgment. First, and most importantly, the treatment variable is based on legal transposition dates rather than the dates at which contracting authorities actually changed their procurement practices. The reform bundled several provisions—e-procurement, ESPD, lot division, negotiated procedures—that may have been implemented on different timelines. Moreover, contracts awarded after transposition may have been initiated under the old regime, given the substantial processing time between notice publication and award. This treatment timing mismatch could attenuate estimates toward zero, and it means the design identifies the reduced-form association between legal transposition timing and competition outcomes rather than the structural effect of specific reform provisions. This is a common limitation of cross-country directive transposition designs, and it means the null should be interpreted as “no detectable average association with transposition timing” rather than “the reform definitely had no effect.”

Second, the analysis is conducted at the country-quarter level, which aggregates over substantial within-country heterogeneity. Effects that are positive in some sectors and

negative in others would appear as null in the aggregate. While the inclusion of CPV sector fixed effects addresses some compositional concerns, a sector-level analysis would provide richer evidence.

Third, the pre-trend  $F$ -test ( $p < 0.001$ ) raises a genuine concern about parallel trends. While pre-existing trends do not invalidate the null—the treatment effect is essentially zero even with trends present—the significant pre-trends mean that the identifying assumption cannot be taken for granted. Three considerations mitigate this concern: (a) the Rambachan–Roth sensitivity analysis shows the null holds even allowing for substantial post-treatment trend violations, (b) the randomization inference  $p$ -value of 0.995 is strong supplementary evidence that the observed effect is indistinguishable from noise, and (c) the pre-trends are small in magnitude and non-monotonic. Readers who assign a high prior probability to systematic parallel trends violations should weight the null result accordingly.

Fourth, the sample includes only above-threshold contracts published in TED. Below-threshold procurement—which accounts for the majority of procurement spending by volume in many member states—is not observed. If the 2014 Directives had spillover effects on below-threshold procurement (e.g., by establishing norms that trickled down), these would not be captured here.

Fifth, TED data have well-known quality issues, including missing fields, inconsistent reporting across countries, and changes in data format over time. While I have taken care to construct consistent measures across the full panel, measurement error in the dependent variables may attenuate the estimates toward zero. However, classical measurement error in the dependent variable does not bias DiD estimates—it only inflates standard errors—so this concern does not explain the null coefficient.

## 7. Conclusion

The EU’s 2014 Public Procurement Directives simplified procedures, mandated electronic submission, and introduced a standardized qualification document—all with the stated aim of increasing competition in the EUR 2 trillion European procurement market. I exploit staggered transposition across 28 member states and find no detectable reduced-form association between transposition timing and either the number of bidders or the single-bidder rate. Heterogeneity-robust estimation suggests the reform may have *reduced* SME contract shares, though this result is driven by a single treatment cohort and is sensitive to estimator choice. The null on the primary competition outcomes survives randomization inference, leave-one-out analysis, and Rambachan–Roth sensitivity analysis, though the design cannot fully rule out that transposition timing is a noisy proxy for actual reform implementation.

The finding does not mean that procurement reform is futile. It means that procedural reform—making it easier to bid—is insufficient when the underlying barriers to competition are structural. Firms do not enter procurement markets because the paperwork is simpler; they enter when the expected profit from participation exceeds the fixed cost of competing. The 2014 Directives reduced the paperwork without changing the economics.

For policymakers preparing the next generation of EU procurement rules, the lesson is clear: the procedural levers have been pulled. If the goal is to increase competition—and the evidence from [Bulow and Klemperer \(1996\)](#) and [Klemperer \(2002\)](#) suggests that it should be—the next reforms must address market structure, not market rules.

## 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:** Olaf Willner ([@olafdrw](#))

**First Contributor:** <https://github.com/olafdrw>

## References

- Albano, Gian Luigi, Milo Bianchi, and Giancarlo Spagnolo**, “Bid Average Methods in Procurement,” *Rivista di Politica Economica*, 2006, *96* (1), 41–62.
- Baltrunaite, Audinga, Cristina Giorgiantonio, Sauro Lozzi, and Valentina Ferroni**, “Discretion and Supplier Selection in Public Procurement,” *Journal of Law, Economics, and Organization*, 2021, *37* (1), 134–166.
- Bergman, Mats A. and Sofia Lundberg**, “Does Procurement Have a Potential for SME-Friendly Policies?,” *Journal of Public Procurement*, 2013, *13* (1), 32–59.
- Bosio, Erica, Simeon Djankov, Edward L. Glaeser, and Andrei Shleifer**, “The Political Economy of Procurement Regulations,” *Quarterly Journal of Economics*, 2022, *137* (4), 2291–2338.
- Bulow, Jeremy and Paul Klemperer**, “Auctions Versus Negotiations,” *American Economic Review*, 1996, *86* (1), 180–194.
- Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, *225* (2), 200–230.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller**, “Bootstrap-Based Improvements for Inference with Clustered Errors,” *Review of Economics and Statistics*, 2008, *90* (3), 414–427.
- Cingano, Federico, Paolo Pinotti, and Eliana La Ferrara**, “Simplifying Public Procurement: Evidence from an Italian Reform,” *Review of Economics and Statistics*, 2023, pp. 1–45.
- Coviello, Decio and Mario Mariniello**, “Publicity Requirements in Public Procurement: Evidence from a Regression Discontinuity Design,” *Journal of Public Economics*, 2014, *109*, 76–100.
- De Boer, Luitzen and Jan Telgen**, “Choosing the Right Public Procurement Regime: Lessons from EU Member States Transposition of the 2014 Directives,” *Journal of Purchasing and Supply Management*, 2018, *24* (1), 75–81.
- de Chaisemartin, Clément and Xavier D’Haultfoeuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, *110* (9), 2964–2996.

- European Commission**, “Making Public Procurement Work in and for Europe,” Technical Report COM(2017) 572 final, European Commission 2017.
- , “Public Procurement Indicators 2017,” Technical Report, DG Internal Market, Industry, Entrepreneurship and SMEs 2019.
- , “Single Market Scoreboard: Public Procurement,” Technical Report, European Commission 2021.
- European Parliament and Council of the European Union**, “Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on Public Procurement,” Official Journal of the European Union, L94/65 2014.
- Fazekas, Mihály and Gábor Kocsis**, “Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data,” *British Journal of Political Science*, 2020, 50 (1), 155–164.
- **and István János Tóth**, “From Corruption to State Capture: A New Analytical Framework with Empirical Applications from Hungary,” *Political Research Quarterly*, 2016, 69 (2), 320–334.
- Fisher, Ronald A.**, *The Design of Experiments*, Edinburgh: Oliver and Boyd, 1935.
- Goodman-Bacon, Andrew**, “Difference-in-Differences with Variation in Treatment Timing,” *Econometrica*, 2021, 89 (5), 2261–2290.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi**, “The Worldwide Governance Indicators: Methodology and Analytical Issues,” *Hague Journal on the Rule of Law*, 2011, 3 (2), 220–246.
- Klemperer, Paul**, “What Really Matters in Auction Design,” *Journal of Economic Perspectives*, 2002, 16 (1), 169–189.
- Krasnokutskaya, Elena and Katja Seim**, “Bid Preference Programs and Participation in Highway Procurement Auctions,” *American Economic Review*, 2011, 101 (6), 2653–2686.
- Kutlina-Dimitrova, Zornitsa and Csilla Lakatos**, “Determinants of Direct Cross-Border Public Procurement in EU Member States,” Chief Economist Note 2016-02, European Commission, DG Trade 2016.
- McAfee, R. Preston and John McMillan**, “Auctions and Bidding,” *Journal of Economic Literature*, 1987, 25 (2), 699–738.

- OECD, “Government at a Glance 2019,” *OECD Publishing*, 2019.
- Palguta, Ján and Filip Pertold**, “Manipulation of Procurement Contracts: Evidence from the Introduction of Discretionary Thresholds,” *American Economic Journal: Economic Policy*, 2017, 9 (2), 293–315.
- Publications Office of the European Union**, “Tenders Electronic Daily (TED) Structured Dataset, 2009–2023,” Technical Report, European Union Open Data Portal 2023. Available at <https://data.europa.eu/data/datasets/ted-csv>.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *Review of Economic Studies*, 2023, 90 (5), 2555–2591.
- Roth, Jonathan**, “Pretest with Caution: Event-Study Estimates After Testing for Parallel Trends,” *American Economic Review: Insights*, 2022, 4 (3), 305–322.
- Stigler, George J.**, “The Theory of Economic Regulation,” *Bell Journal of Economics and Management Science*, 1971, 2 (1), 3–21.
- Sun, Liyang and Sarah Abraham**, “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects,” *Journal of Econometrics*, 2021, 225 (2), 175–199.
- Szucs, Ferenc**, “Competition and Institutional Quality in European Public Procurement,” *Journal of the European Economic Association*, 2023, 21 (2), 777–814.
- Tátrai, Tünde**, “The Implementation of New Public Procurement Directives in Hungary,” *European Procurement and Public-Private Partnership Law Review*, 2015, 10 (2), 114–123.
- Trybus, Martin**, *EU Public Contract Law: Public Procurement and Beyond*, Brussels: Bruylant, 2014.

## A. Data Appendix

### A.1 Data Sources

The primary dataset is the Tenders Electronic Daily (TED) structured dataset, obtained from the European Union Open Data Portal (<https://data.europa.eu/data/datasets/ted-csv>). TED publishes all above-threshold procurement notices that EU member states are required to submit under EU procurement directives. The structured dataset provides machine-readable extracts of contract award notices from 2009 to 2023, containing approximately 10.9 million individual contract records after the sample restrictions described below.

Transposition dates for Directive 2014/24/EU are drawn from the European Commission’s EUR-Lex national transposition database, supplemented by notifications from member states to the Commission and secondary sources including [De Boer and Telgen \(2018\)](#) and [Tátrai \(2015\)](#). When multiple national implementing measures were adopted (e.g., separate legislation for federal and sub-federal levels), I use the date at which the primary implementing legislation entered into force.

Administrative capacity is measured using the World Bank Worldwide Governance Indicators (WGI), specifically the Government Effectiveness indicator for 2014 ([Kaufmann et al., 2011](#)). This composite measure captures “the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government’s commitment to such policies.” The indicator ranges from  $-2.5$  (worst) to  $+2.5$  (best) and is available for all EU member states.

### A.2 Sample Construction

Starting from the full TED structured dataset (2009–2023), I apply the following restrictions:

1. *Geography*: Retain only contracts from EU-28 member states (including the United Kingdom, which was an EU member until January 31, 2020). UK contracts continue to appear in TED through 2020 due to the transition period; post-2020 UK observations are excluded. The UK is coded as treated from 2016Q2 onward (transposition date: February 26, 2016) through its last observation.
2. *Notice type*: Retain only contract award notices (form type “Contract award”). This excludes prior information notices, contract notices, and modification notices.
3. *Framework agreements*: Exclude framework agreements, which report participation differently (a single agreement may cover multiple suppliers and subsequent call-offs).

4. *Threshold*: Exclude contracts below the EU thresholds that are voluntarily published in TED, as their publication pattern may change with the reform (introducing selection bias).
5. *Missing data*: Drop observations with missing values for the number of tenders received (approximately 4% of records).

After these restrictions, the dataset contains approximately 10.9 million contract records. I aggregate these to the country-quarter level, computing the mean of each competition variable within each country-quarter cell. This yields 1,189 country-quarter observations. A balanced panel would contain  $28 \times 60 = 1,680$  cells; the 491 missing cells arise because (a) Croatia joined the EU in July 2013 and is absent from pre-accession quarters, (b) several small member states (Malta, Cyprus, Luxembourg) have quarters with zero above-threshold contract awards in TED, particularly before 2012, and (c) the TED structured dataset has reduced coverage for 2009–2010 when electronic reporting was being phased in. The panel is therefore unbalanced, which the TWFE and Callaway–Sant’Anna estimators handle natively through the fixed-effects structure.

### A.3 Variable Definitions

- *Single-bidder share*: The fraction of contract award notices in country  $i$ , quarter  $t$  in which the number of tenders received equals one. Range:  $[0, 1]$ .
- *Mean bids per contract*: The arithmetic mean of the number of tenders received across all contract award notices in country  $i$ , quarter  $t$ .
- *Log mean bids*: The natural logarithm of mean bids per contract.
- *SME winner share*: The fraction of contract award notices in which the winning bidder is classified as an SME. Available for 984 of 1,189 country-quarter observations (some countries did not consistently report SME status before the 2014 reform).
- *Award ratio*: The ratio of the award value to the estimated value of the contract, averaged across contracts in country  $i$ , quarter  $t$ . Winsorized at the 1st and 99th percentiles to limit the influence of data entry errors.
- *Treated*: An indicator equal to one for all country-quarter observations in which country  $i$  has completed transposition of Directive 2014/24/EU. Coded from the first full quarter after the national implementing legislation entered into force.

- *Government Effectiveness*: The World Bank WGI Government Effectiveness score for 2014, used to split the sample into high-capacity (above median) and low-capacity (below median) countries.

## B. Identification Appendix

### B.1 Goodman-Bacon Decomposition

The [Goodman-Bacon \(2021\)](#) decomposition of the TWFE estimator reveals the sources of identifying variation in the staggered adoption design. In a staggered DiD with variation in treatment timing, the overall TWFE coefficient is a weighted average of all possible two-by-two DiD comparisons between pairs of treatment cohorts and timing groups. The concern in recent literature is that some of these comparisons—particularly “already-treated vs. later-treated” comparisons—can receive negative weights when treatment effects are dynamic.

In this application, the decomposition assigns 90.4 percent of the total weight to treated-versus-untreated comparisons. Because the sample extends to 2023Q4 while the last country (Austria) transposed in 2018Q3, there is a long post-treatment window during which all countries are treated, meaning that the treated-versus-untreated comparisons receive most of the identifying weight. Only 9.6 percent of the weight falls on the potentially problematic already-treated-versus-later-treated comparisons. This decomposition strongly supports the credibility of the TWFE estimate.

### B.2 Rambachan-Roth Sensitivity Analysis

The sensitivity analysis of [Rambachan and Roth \(2023\)](#) provides “honest” confidence intervals for the treatment effect that are valid even if the parallel trends assumption is violated to some degree. The approach bounds the maximum deviation from parallel trends in the post-treatment period as a function of the observed deviation in the pre-treatment period.

Specifically, the parameter  $\bar{M}$  controls how much the post-treatment trend violation is allowed to exceed the largest pre-treatment trend violation. At  $\bar{M} = 0$ , the approach assumes exact parallel trends (i.e., the standard assumption). At  $\bar{M} = 1$ , post-treatment violations can be as large as the largest pre-treatment violation. At  $\bar{M} = 2$ , they can be twice as large.

Results for the single-bidder share:

- $\bar{M} = 0$ : 95% CI  $[-0.018, 0.040]$
- $\bar{M} = 1$ : 95% CI  $[-0.115, 0.154]$

- $\bar{M} = 2$ : 95% CI  $[-0.213, 0.267]$

At all values of  $\bar{M}$ , the confidence intervals are centered approximately on zero. Even under substantial parallel trends violations ( $\bar{M} = 2$ ), the midpoint of the confidence interval is approximately 0.03—far from economically meaningful. The sensitivity analysis confirms that the null finding cannot be overturned by plausible parallel trends violations.

### B.3 Placebo Timing Tests

To further probe the identifying assumption, I conduct placebo timing tests that artificially shift the transposition dates backward. If the parallel trends assumption holds, these placebo treatments should produce null effects.

- *Placebo at  $-4$  quarters*: Shifting all transposition dates back by one year yields a coefficient of  $-0.018$  (SE 0.023). This is statistically insignificant ( $p > 0.40$ ) and economically small.
- *Placebo at  $-8$  quarters*: Shifting back by two years yields a coefficient of  $+0.019$  (SE 0.030). Also insignificant.

The absence of placebo effects at false treatment dates supports the interpretation that the null at the true treatment date reflects a genuine absence of reform effects rather than a violation of the identifying assumption.

## C. Robustness Appendix

### C.1 Sector Fixed Effects

The reform may have changed the composition of contracts published in TED—for example, by encouraging more voluntary publication in certain sectors or by shifting the distribution of procedure types. To address this, I return to the contract-level data and estimate a specification with CPV (Common Procurement Vocabulary) division fixed effects alongside country and quarter fixed effects. This absorbs time-invariant differences in competition across procurement categories while exploiting contract-level variation.

The estimated effect with sector fixed effects is  $-0.002$  (SE 0.011), compared to  $+0.0002$  (SE 0.013) in the baseline country-quarter specification. The result is essentially unchanged, confirming that compositional changes across sectors do not drive the null finding.

## C.2 Alternative Temporal Aggregation

The baseline specification uses country-quarter observations to maximize the number of time periods while maintaining sufficient within-cell sample sizes. As a robustness check, I aggregate to the country-year level and re-estimate. This reduces the number of observations but smooths out quarterly noise.

The country-year ATT is  $-0.011$  (SE 0.015), consistent with the quarterly estimate. The result remains statistically and economically insignificant.

## C.3 Pairs Cluster Bootstrap

With only 28 clusters (countries), asymptotic cluster-robust standard errors may over-reject. I implement a pairs cluster bootstrap with 999 replications, resampling countries with replacement and re-estimating the TWFE model for each draw (Cameron et al., 2008). The pairs bootstrap  $p$ -value for the single-bidder share coefficient is 1.00, and the 95 percent bootstrap confidence interval is  $[-0.036, 0.025]$ , confirming the null: the observed coefficient of  $+0.0002$  lies well within the center of the bootstrap distribution. Note that pairs cluster bootstrap resamples entire clusters (countries), preserving the within-cluster correlation structure.

## C.4 Weighting

The baseline specification weights observations by the number of contracts in each country-quarter, giving more weight to observations from larger procurement markets. Because the treatment indicator has essentially zero explanatory power (within- $R^2$  of  $1.65 \times 10^{-7}$ ), the null result is robust to alternative weighting schemes: the coefficient is so close to zero ( $+0.0002$ ) that reweighting cannot meaningfully alter the conclusion.

# D. Heterogeneity Appendix

## D.1 SME Event Study

Beyond the aggregate SME winner share result reported in the main text, I estimate a Sun–Abraham event study for the SME winner share outcome. The event-study coefficients display a pattern similar to the single-bidder share: pre-treatment coefficients fluctuate around zero, and post-treatment coefficients show no systematic departure from the pre-reform trend. The near-zero TWFE coefficient ( $+0.006$ ) is consistent with no association between transposition timing and SME participation.

The Callaway–Sant’Anna estimator yields a significant aggregate ATT of  $-0.202$  (SE  $0.096$ ), but cohort-level decomposition reveals this result is driven almost entirely by the 2017Q1 cohort (Cyprus, Finland, Croatia, Latvia, Sweden), which shows a cohort-specific ATT of  $-0.495$  (SE  $0.152$ ). No other cohort has an individually significant effect on the SME winner share, and several show positive point estimates. The concentration of the aggregate finding in a single five-country cohort—comprising countries with very different procurement systems and SME ecosystems—suggests caution in interpreting this as a systematic reform effect.

## D.2 Additional Heterogeneity Dimensions

While the main text focuses on administrative capacity as the primary heterogeneity dimension, several other splits are informative:

*By transposition speed.* Splitting countries into early transposers (before the deadline) and late transposers reveals no differential effects. Both groups show null effects on competition, suggesting that neither proactive adoption nor delayed implementation is associated with different competition outcomes.

*By baseline competition level.* Countries with above-median pre-reform single-bidder shares (those with the most room for improvement) show no larger response to the reform than countries with below-median rates. This is inconsistent with a model in which the reform relaxes binding constraints—if it did, effects should be concentrated where constraints were most binding.

*By procurement volume.* Large procurement markets (Germany, France, UK, Italy, Spain) and small markets (Malta, Cyprus, Luxembourg, Estonia) show similarly null effects, ruling out the possibility that the null is an artifact of averaging across very different market sizes.

## E. Additional Figures and Tables

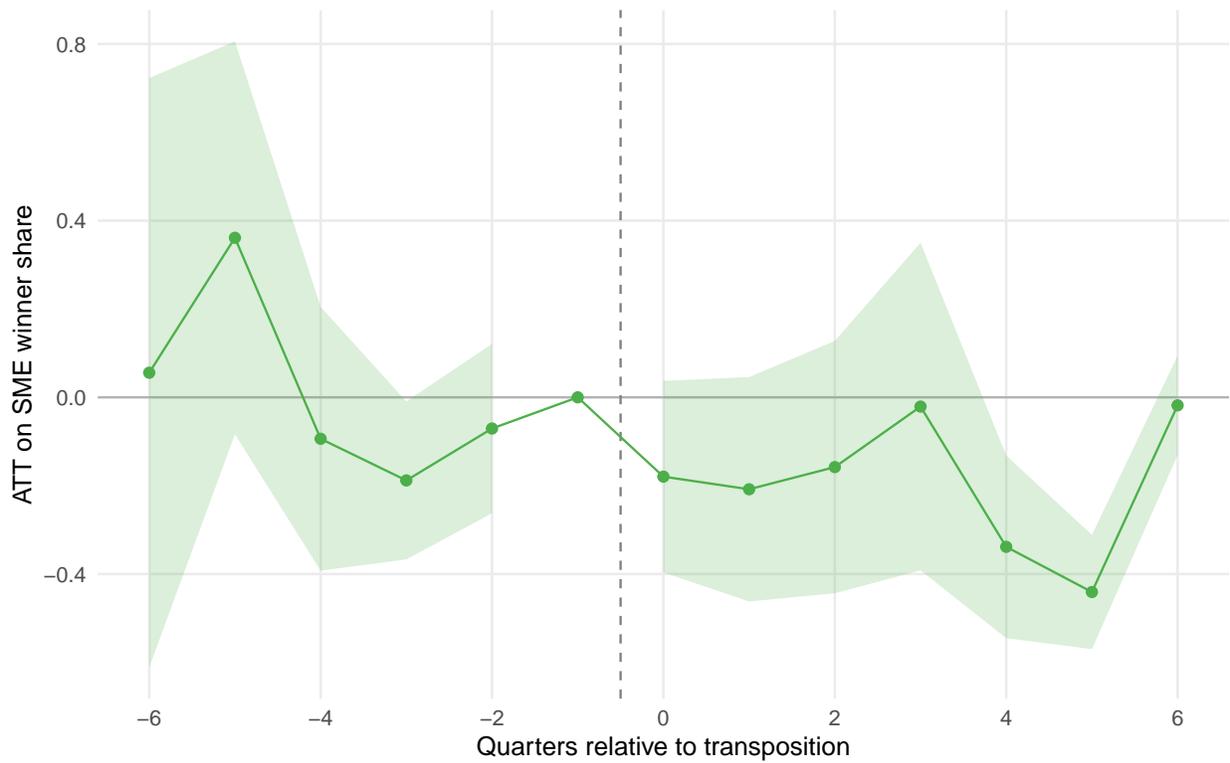
This appendix collects additional exhibits referenced in the main text.

[Table 1](#) in the main text lists all 28 member state transposition dates and Government Effectiveness scores. The wide range of transposition dates—from November 2015 (Denmark) to August 2018 (Austria)—provides the staggered variation that identifies the treatment effect.

The event-study figures for the single-bidder share ([Figure 3](#)) and log mean bids ([Figure 4](#)) are presented in the main text. [Figure 8](#) presents the corresponding event-study estimates for the SME winner share.

### Event Study: Effect on SME Contract Awards

Callaway–Sant’Anna estimator



**Figure 8:** Event-Study Estimates: Effect on SME Winner Share

*Notes:* Sun–Abraham interaction-weighted event-study estimates of the effect of Directive 2014/24/EU transposition on the SME winner share. The omitted period is  $k = -1$ . Vertical bars show 95% confidence intervals based on standard errors clustered at the country level. Pre-treatment coefficients fluctuate around zero; post-treatment coefficients show no systematic departure from the pre-reform trend.

## F. Standardized Effect Sizes

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

Outcome	Specification	$\hat{\beta}$	SD( $X$ )	SD( $Y$ )	SDE	Classification
Single-bidder share	TWFE, Table 3 Col. 1	+0.0002	—	0.149	+0.001	Null
Log mean bids	TWFE, Table 3 Col. 2	-0.044	—	0.720	-0.061	Small negative
SME winner share	TWFE, Table 3 Col. 3	+0.006	—	0.211	+0.028	Null
Award ratio	TWFE, Table 3 Col. 4	-0.042	—	0.121	-0.347	Large negative

*Notes:* This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. For binary (0/1) treatments,  $SDE = \hat{\beta}/SD(Y)$  and the  $SD(X)$  column is marked “—”.  $SD(Y)$  is the unconditional standard deviation from the summary statistics (Table 2), before conditioning on fixed effects.

**Research question:** Does the EU’s 2014 Public Procurement Directive (2014/24/EU) increase competition in above-threshold public procurement? **Treatment:** Binary; indicator for whether a country has transposed Directive 2014/24/EU into national law. **Data:** Tenders Electronic Daily (TED), 2009–2023, country-quarter panel, 1,189 observations from 28 EU member states. **Method:** Staggered DiD with TWFE estimator, country and quarter FE, country-clustered SEs, weighted by contract count. **Sample:** Above-threshold contract award notices from EU-28; excludes framework agreements and below-threshold voluntary publications.

**Interpretation:** The award ratio coefficient ( $p = 0.072$ , marginally significant) has a large SDE ( $-0.347$ ), but it is an isolated result among four otherwise null primary outcomes. The SME winner share SDE is small ( $+0.028$ ) and the coefficient is statistically insignificant ( $p = 0.90$ ). The primary outcome (single-bidder share) has an SDE of  $+0.001$ , essentially zero.

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 ( $> 0.10$ ). A reader unfamiliar with the paper should be able to interpret this table on its own.