A Guide to Callaway & Sant’Anna (2021) and usage in the Matlab DiD Toolbox

(Overview by ChatGPT 5)

**What it estimates.** In panels where units adopt at different times (or never), Callaway–Sant’Anna (C&S) compute treatment effects for each cohort–time cell, written ATT(g,t): “the average effect at calendar time t for the cohort first treated in g.” This avoids the mixing issues of classic TWFE when effects vary over time or across cohorts.

From these building blocks, you can aggregate transparently:

* Event-time (θ\_es): average by time-since-adoption e = t − g (an event-study curve).
* Calendar-time (θ\_c): average across cohorts within each calendar period t.
* Cohort averages (θ\_sel): average over post periods within a cohort g.
* Overall ATET (θ\_overall): a single, weighted average across all post cells.

1) Choosing the control group

**Comparison = "never" (never-treated controls).** Only units never treated serve as controls.  
Pros: clean interpretation; closest to classic DiD.  
Cons: wasteful if few never-treated; impossible if everyone is treated eventually.

**Comparison = "notyet" (not-yet-treated controls).** At time t, controls are units treated later than t.  
Pros: uses more data; works when everyone is eventually treated.  
Cons: relies on a stronger “no anticipation” condition.

2) Allowing for anticipation with Δ

Outcomes can move before formal adoption. Set Delta = δ ≥ 0 to exclude the δ periods before adoption from the pre-treatment baseline. The baseline becomes r = g − δ − 1 (instead of g − 1). Small δ (e.g., 1) protects against mild anticipation; larger δ trades precision for robustness.

3) Four estimation approaches (how we build the counterfactual)

**Approach = "unconditional" (simple DiD on changes).** Compare treated vs controls without covariates on long differences.  
Use when unconditional parallel trends is plausible.

**Approach = "or" (Outcome Regression).** On controls, regress the change in outcomes on X; predict treated counterfactuals; subtract.  
Assumption: the outcome model for controls is well specified (conditional parallel trends).

**Approach = "ipw" (Inverse Probability Weighting).** Estimate propensity Pr(G=g | X); reweight controls to match treated covariates; compare mean changes.  
Assumption: correct propensity model and overlap (treated-like controls exist).

**Approach = "dr" (Doubly Robust).** Combine OR and IPW: adjust treated by an outcome model and weight control residuals by propensities.  
Benefit: consistent if either the outcome or the propensity model is correct.

**Rule of thumb:** If you have credible covariates, use Approach="dr"; otherwise start with "unconditional" and consider a small Delta.

4) Aggregations you’ll read

* Event-time (θ\_es): averages by e = t − g. Pre-period (leads) check parallel trends; post (lags) trace dynamics.
* Calendar-time (θ\_c): averages across cohorts within each calendar period t—useful for macro “when” questions.
* By cohort (θ\_sel): average effect for cohort g across its post periods.
* Overall (θ\_overall): one ATET; typically cohort-share weights are used.

Make sure you compare like with like (event-time vs calendar-time can differ substantially).

5) Inference (SEs & bands)

**Multiplier (wild) bootstrap — SEMethod="multiplier".** Perturb per-unit influence contributions with random weights to get draws for ATT(g,t) and aggregates.  
Pros: fast; delivers simultaneous max‑t bands; robust to heteroskedasticity.  
Use when: you want bands for event-study plots or lack a natural cluster beyond units.  
Caveat: needs enough independent units; dominance by a few units can distort.

**Clustered SEs (one-way) — SEMethod="clustered".** Treat each unit as a cluster and form cluster-robust variance (CRV1).  
Pros: matches Stata’s cluster(id); robust to serial correlation within unit.  
Use when: dependence is mainly within units; you need pointwise SE/t/p for tables.  
Caveat: with few clusters (e.g., < 50) SEs can be biased; be cautious with p‑values.

**Two-way clustered SEs — SEMethod="clustered2".** Cameron–Gelbach–Miller: Var(unit) + Var(second cluster) − Var(intersection).  
Pros: helpful when there are shocks along a unit‑constant dimension (industry/region).  
Note: with per‑unit influence functions, the second cluster must be unit‑constant; true unit×time clustering needs observation‑level IFs.

**Bands vs SEs:** Multiplier → simultaneous bands for plots; Clustered → pointwise SEs (for bands under clustering, bootstrap clusters).

6) Assumptions checklist (informal)

* Parallel trends (unconditional or conditional on X, matching your approach).
* No anticipation (with Delta = δ, no effects before g − δ).
* Overlap for IPW/DR (treated-like controls exist in X‑space).
* Sufficient independent units/clusters for inference.
* Reasonable cell sizes (avoid tiny or empty ATT(g,t) cells driving results).

7) Practical defaults

* Baseline: Comparison="notyet", Delta ∈ {0,1}, Approach="unconditional", with SEMethod="clustered" (tables) or "multiplier" (bands).
* With covariates: prefer Approach="dr".
* Everyone eventually treated: must use Comparison="notyet".
* Need bands: SEMethod="multiplier", B≈999, Studentize=true.
* Common shocks by industry/region: SEMethod="clustered2" with a unit‑constant ClusterVar2.

8) Frequent pitfalls

* Comparing event‑time to calendar‑time aggregates.
* Letting tiny cells dominate; report which cells are used/dropped.
* Few clusters → beware small p‑values; show CIs and the number of clusters.
* Ignoring anticipation; set Delta>0 if pre‑trends move.