Statistics/Data analysis

Title

fan_park —— Sharp bounds on the distribution (and quantile function) of treatment effects for a binary treatment, with optional covariate tightening and pointwise inference.

Syntax

fan_park *depvar treatvar* [*indepvars*] [*if*] [*in*] [, **delta_partition(#) delta_values(***numlist***) cov_partition(#) level(#) nograph seed(#) qbounds num_quantiles(#)**]

Description

fan_park implements the sharp, nonparametric bounds on the distribution of treatment effects for a binary treatment developed by Fan and Park (2010, *Econometric Theory*). Let $\delta = Y1 - Y0$. Given the marginal distributions *F1* and *F0*, the command estimates pointwise sharp bounds $F\_L(\delta) <= F\_\delta(\delta) <= F\_U(\delta)$ over a grid of δ values (the distribution-function branch).

With option **qbounds**, the command instead computes bounds on the quantile function $F^{-1}\_\delta(q)$: $F^{-1,U}(q) <= F^{-1}\_\delta(q) <= F^{-1,L}(q)$ (the quantile-function branch). No standard errors are reported in this branch.

When *indepvars* are provided, the covariate space is partitioned by **cluster kmedians**. Conditional bounds are computed within each cell using cell-specific supports and then averaged with empirical cell weights. The reported bounds intersect the unconditional and the averaged-conditional bounds, which can tighten the region under heterogeneity and overlapping-support restrictions.

Options

Options	Description
delta_partition(#)	Number of grid points for the support of $\delta = Y1 - Y0$ when computing distribution-function bounds; default **100**.
delta_values(*numlist***)**	Explicit grid of δ values at which to compute bounds. Values outside the feasible range [$a - d$, $b - c$] are ignored. Supersedes **delta_partition()**.
cov_partition(#)	If *indepvars* are supplied, number of k-medians clusters used to partition covariate space (tightening the bounds); default **6**.
level(#)	Confidence level for pointwise confidence intervals on distribution-function bounds; default **95**. The graph shows 95% and 90% bands by default.
nograph	Suppress the graph.
seed(#)	Initialization passed to **cluster ... start(krandom(#))**. With the default **seed(1)**, a random integer is drawn internally; set a fixed integer for replication.
qbounds	Compute bounds for the quantile function of δ instead of the CDF (no SEs/CIs reported).
num_quantiles(#)	Number of equally spaced quantiles in [0,1] for **qbounds**; default **100**.

Examples

Baseline usage

```
use limits_commitment.dta, clear
fan_park apr treat
```

With covariates (potential tightening)

```
fan_park apr treat age educ, cov_partition(6)
```

Specify the δ grid explicitly

```
fan_park apr treat, delta_values(-10(0.5)10)
```

Quantile-function bounds
fan_park apr treat, qbounds num_quantiles(200)

Simulation
do simulation_fan_park.do

Stored results

fan_park stores the following in **r()**.

r(bounds)	Distribution-function branch: *n x 2* matrix with columns **LB** and **UB**, evaluated at **r(delta_val)**. Quantile-function branch: *n x 2* matrix with quantile-function bounds **UB** and **LB**, evaluated at **r(q_val)**.
r(sigma_2)	Distribution-function branch only. *n x 2* matrix of large-sample variances for the active lower and upper bound at each δ (columns **Var_L**, **Var_U**).
r(M_delta)	Distribution-function branch only. *n x 2* matrix; column 1 stores *M(δ) = sup over y of [F1(y) − F0(y − δ)]*, column 2 stores *m(δ) = inf over y of [F1(y) − F0(y − δ)]*.
r(M_active)	Distribution-function branch only. *n x 2* matrix with the sup/inf values corresponding to the bound that is actually active at each δ after intersecting unconditional and conditional bounds (columns **M_L_active**, **M_U_active**).
r(delta_val)	Distribution-function branch only. *n x 1* grid of δ values used for **r(bounds)**.
r(q_val)	Quantile-function branch only. *n x 1* grid of quantiles in [0,1] used for **r(bounds)**.

Remarks

Computation. The command builds empirical CDFs of *Y1* and *Y0 + δ* via **cumul**, interpolates with **ipolate**, and searches the intersection support *Y_δ = [a,b] intersect [c + δ, d + δ]*. With covariates, the same is done cell-by-cell using cell-specific supports {[a_c, b_c]} and {[c_c, d_c]}, and then averaged across cells using empirical shares.

Interpretation. **r(bounds)** contains the final (possibly tightened) bounds after intersecting unconditional and conditional averages. **r(M_delta)** reports the raw sup/inf from the unconditional step; **r(M_active)** reports the sup/inf corresponding to whichever bound is active at each δ (used for the one-sidedness adjustment in CIs).

Positivity and empty cells. A cluster with no treated or no controls is skipped; weights are renormalized over remaining cells and a note is displayed. If all cells are dropped at a given δ (rare), the conditional piece is missing and the unconditional bound is reported.

Inference. CIs are pointwise (not uniform) and based on large-sample normal approximations. The graph displays 95% and 90% bands by default. For **qbounds**, CIs are not reported; if needed, consider bootstrap inference away from boundary cases.

References

Fan, Yanqin, and Sang Soo Park (2010). “Sharp bounds on the distribution of treatment effects and their statistical inference.” *Econometric Theory* 26(3): 931–951. DOI: 10.1017/S0266466609990168

Authors

Isaac Meza López, Department of Economics, Harvard University.