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(#)**]

Options	Description
delta_partition(#)	Number of grid points for the treatment-effect support 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 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 CIs on distribution-function bounds; default **95**.
nograph	Suppress the graph.
seed(#)	Initialization for **cluster kmedians, start(krandom(#))**. If **seed(1)** (the default), a random seed is drawn internally; set a fixed integer to reproduce partitions.
qbounds	Compute bounds for the quantile function of treatment effects instead of the CDF. (No CIs are reported in this branch.)
num_quantiles(#)	Number of equally spaced quantiles in [0,1] for **qbounds**; default **100**.

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 program estimates pointwise sharp bounds $F^L(\delta) \le F_\Delta(\delta) \le F^U(\delta)$ over a grid of δ values. With the option **qbounds**, it instead computes bounds on the quantile function $F^{-1}_\Delta(q)$: $F^{-1,U}(q) \le F^{-1}_\Delta(q) \le F^{-1,L}(q)$.

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 final reported bounds intersect the unconditional and the averaged-conditional bounds, which can tighten the region under covariate heterogeneity and overlapping-support restrictions.

Arguments

depvar is the outcome. Continuous outcomes are typical, but binary outcomes are allowed.

treatvar is a numeric indicator **0/1**. The command verifies this within the estimation sample.

indepvars is an optional list of covariates (numeric). Factor-variable notation is not expanded here; pass numeric covariates or pre-built dummies.

Options

delta_partition(#) sets the number of δ grid points used to approximate the support of $\Delta = Y1 - Y0$. The grid is the equally spaced partition of $[a-d, b-c]$, where $[a,b]$ is the support of $Y1$ and $[c,d]$ of $Y0$, estimated from the sample.

delta_values(numlist) specifies explicit δ values. Values outside [*a*−*d*, *b*−*c*] are dropped. This option overrides **delta_partition()**.

cov_partition(#) chooses the number of clusters for **cluster kmedians** over *indepvars*. Cells with no treated or no controls are skipped (positivity). Weights are the empirical cell shares among usable cells and are re-normalized.

level(#) sets the confidence level for pointwise CIs on the distribution-function bounds. The graph shows shaded 95% and 90% bands (by default).

nograph suppresses the graph.

seed(#) passes through to **start(krandom(#))** for **cluster kmedians**. With the default **seed(1)**, a random integer seed is drawn internally; specify a fixed integer (e.g., **seed(12345)**) for replication.

qbounds requests bounds for the quantile function of Δ. No standard errors/CIs are computed in this branch (you may bootstrap if the bound is interior).

num_quantiles(#) sets the number of quantile grid points in [0,1] for **qbounds**.

Examples

Baseline usage

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

With covariates (tighter bounds when supports differ by X)

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

Specify 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()**.

Matrices	Description
r(bounds)	For distribution-function bounds (default branch): an *n* × *2* matrix with columns *LB* and *UB* evaluated at **r(delta_val)**. For **qbounds**: an *n* × *2* matrix with quantile-function bounds *UB* and *LB* evaluated at **r(q_val)**.
r(sigma_2)	(Only when **qbounds** is *not* specified.) *n* × *2* matrix of large-sample variances for the active lower and upper bound at each δ (columns *Var_L*, *Var_U*). These underlie the pointwise CIs shown in the graph.
r(M_delta)	(Distribution-function branch.) *n* × *2* matrix; column 1 stores *M(δ) = sup_y [F1(y) − F0(y − δ)]*, column 2 stores *m(δ) = inf_y [F1(y) − F0(y − δ)]*.
r(M_active)	(Distribution-function branch.) *n* × *2* matrix; 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.) *n* × *1* grid of δ values used for **r(bounds)**.
r(q_val)	(Quantile-function branch.) *n* × *1* grid of quantiles in

[0,1] used for **r(bounds)**.

Remarks

Computation. The program builds empirical CDFs of *Y1* and *Y0+δ* via **cumul**, interpolates with **ipolate**, and searches the intersection support *Y_δ = [a,b] ∩ [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]* before averaging across cells by 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 CI one-sidedness switch).

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

Inference. CIs are pointwise (not uniform) and based on large-sample normal approximations. Shaded bands on the graph correspond to 95% and 90% intervals. For **qbounds**, CIs are not provided; 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.

Authors

Isaac Meza López, Department of Economics, Harvard University.