Who's afraid of selection bias? Robust inference in the presence of competitive selection, by Thomas Noe, Oxford (UK)

Discussant: Alex Boulatov, HSE

HSE, November 9, 2013

Overview

- Motivation
- Contributions
- Model Overview
- Conclusion

Motivation: What is (Stochastic) Selection Dominance?

- In many economic situations, agents have to make inferences about selection-conditioned distributions from unconditional distributions, and vice versa
  - **conditional→unconditional**: Can we infer the relation between quality distributions of Oxford and Cambridge graduates based on salary data for those hired by Goldman Sachs?
  - **Assumption:** Goldman hires the best (criteria – competitive or fixed?)
  - Fixed criteria (exogenous threshold) – monotone likelyhood ratio property (MLRP)

- (*Competitive Stochastic*) **Selection Dominance**: Expected value **(for ANY reasonable utility)** defined w.r.t. the "winning" distribution is higher than that w.r.t. the "rival" distribution (conditional on **competitive selection**)
  - **Unconditional** – Stochastic Dominance (SD)

Contributions

- **Analytic charactreization** of conditions for (CS)**Selection Dominance**, analogous to SD and MLRP
- **Necessary and sufficient** condition – **supermultiplicativity on average**
  - **However**: no ordering for distributions (no transitivity $A \succ B, B \succ C$, but $A \succ\prec C$)
- **Sufficient** condition – **geometric convexity** (geometric dominance, GD)
- Generates **ordering** for distributions
  - Interestingly, there could be cases when GD distributions are dominated in SD sense
  - Makes some sense economically: ex.: good students from bad schools may be extremely good
  - Rare "catastrophe" events
- **However**: SD is itself a sufficient critterion and imposes some limits (**arbitrary** utility function)

Let: $\widetilde{X}$ and $\widetilde{Y}$ be independent r.v., with cdf F and G (resp.), $v(\cdot)$ **valuation function**, strictly increasing.
Value max., Select: $\max\left\{\widetilde{X}, \widetilde{Y}\right\}$

Definition

$\widetilde{X}$ (or F) (comp. stochastically) selection dominates $\widetilde{Y}$ (or G) if, for **any** increasing $v(\cdot)$,

$$E\left[v\left(\widetilde{X}\right)|\widetilde{X} > \widetilde{Y}\right] \geq E\left[v\left(\widetilde{Y}\right)|\widetilde{Y} > \widetilde{X}\right].$$

- What are the conditions on F and G that lead to CSSD?

Definition

$\widetilde{X}$ (or F) stochastically dominates $\widetilde{Y}$ (or G) if, for **any** increasing $v(\cdot)$,

$$E\left[v\left(\widetilde{X}\right)\right] \geq E\left[v\left(\widetilde{Y}\right)\right].$$

Result: equivalent to $F(x) \leq G(x)$

- Standard result; introduces ordering
- Derivation: partial integration, then for **any** $v'(x) \geq 0$,

$$E\left[v\left(\widetilde{X}\right)\right] - E\left[v\left(\widetilde{Y}\right)\right] = -\int_{\underline{x}}^{\overline{x}} dx v'(x)\left(F(x) - G(x)\right).$$

Technical Assumptions

Admissible pair of distributions

- F and G have common support $[\underline{x}, \overline{x}]$, $0 \leq \underline{x} < \overline{x} \leq +\infty$
- F and G are continuous and mutually abs. continuous
- $\int_0^{+\infty} dF(x)\, x < \infty$, $\int_0^{+\infty} dG(x)\, x < \infty$ – finite "expectations"
- Make things less cumbersome

Main result

Characterization of CSSD

- CS (conditional on "winning") cdfs (note: typos in Eqs.(7) and (9)):

$$H(x) = \frac{\int_{\underline{x}}^{x} dF(s)\, G(s)}{\int_{\underline{x}}^{\overline{x}} dF(s)\, G(s)}, \quad J(x) = \frac{\int_{\underline{x}}^{x} dG(s)\, F(s)}{\int_{\underline{x}}^{\overline{x}} dG(s)\, F(s)}$$

- Applying SD, we immediately obtain

$$H(x) \leq J(x).$$

- This is it! Very simple intuition. Can also prove that: $H(x) \leq F(x)\, G(x) \leq J(x)$

- Define a mapping $u : [0;1] \to [0;1]$, $F = u \circ G$, $u = FG^{-1}$ is increasing and continuous
- Also, let $v = u^{-1}, v : [0;1] \to [0;1]$, $G = v \circ F$, $v = GF^{-1}$ – inverse transformation
- Supermultiplicativity on average (for u):

$$\int_0^1 ds \left(u\left(ts\right) - u\left(t\right) u\left(s\right)\right) \geq 0, \quad \forall t \in \left(0;1\right).$$

- Sufficient condition (geometric convexity): $u\left(ts\right) \geq u\left(t\right) u\left(s\right), \quad \forall t, s \in \left(0;1\right)$
- Geometric equivalence: $F\left(x\right) = \left(G\left(x\right)\right)^p, \quad p \in \left(0;1\right)$

Characterizations of Geometrically Dominant pdfs

Suppose:

u is differentiable

- Then u is geometrically convex, iff
- $R(t) = \frac{d(\ln u)}{d(\ln t)}$ is increasing t, when $t \in [0;1]$
- Make relations with SD and MLRP, etc.

Conclusion

- Close form characterization of CSSD
- Inferences on unconditional distributions are limited:
- The notion of SD; **sufficient, but not necessary** by construction