Comparing Latent Inequality with Ordinal Data

David M. Kaplan
University of Missouri

Longhao Zhuo
University of Missouri

Yale Econometrics Seminar
4 Sept 2019

Outline

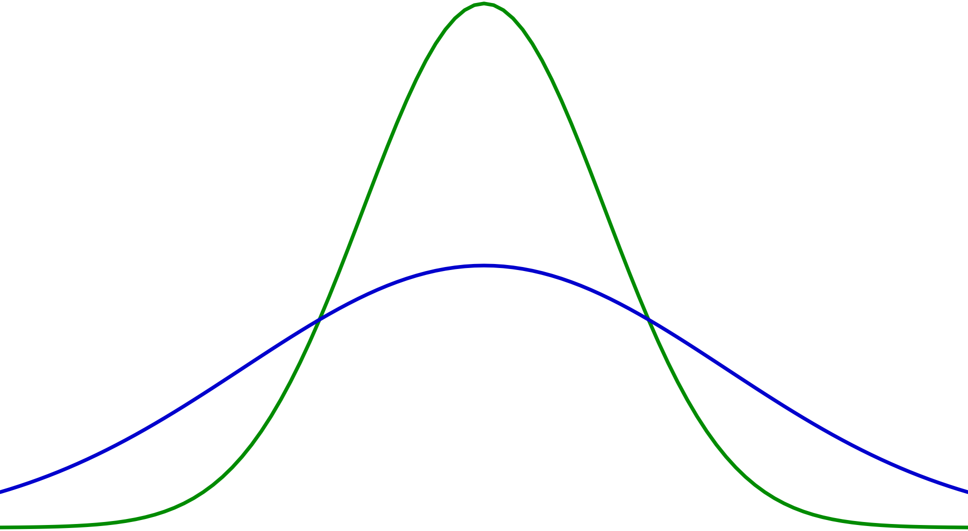
Outline

Motivation: health inequality

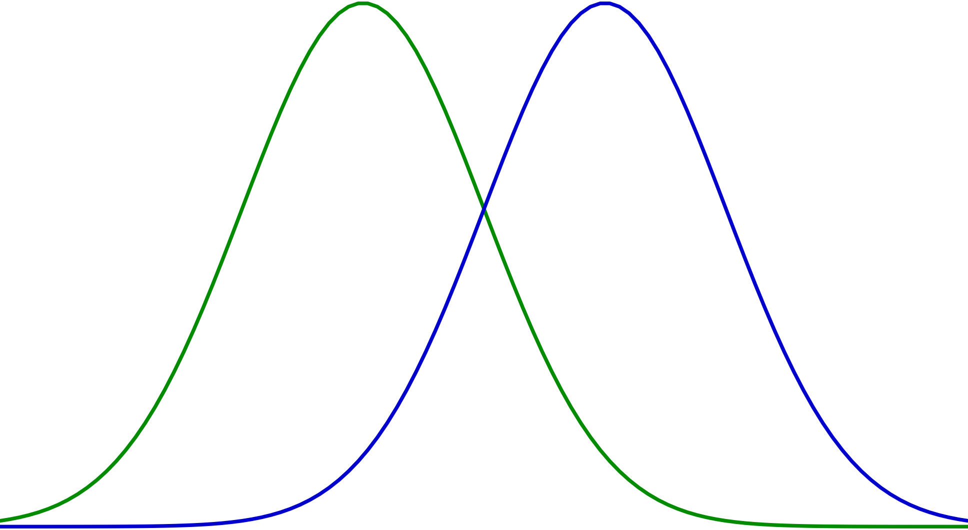
From Deaton and Paxson (1998a, pp. 248–9) and Deaton and Paxson (1998b, pp. 431–2):

- "Our interest in health inequality stems from a more general interest in the distribution of welfare"
- Interested in "whether inequality in health. . . increases with age" or "within-cohort dispersion" (within-group)
- and "[inequality] across socioeconomic groups" (between-group)

Inequality: within-group (dispersion)



Inequality: between-group (better/worse)

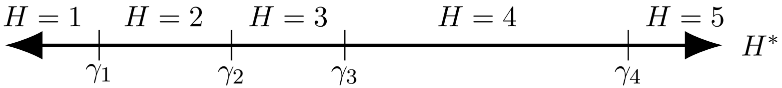


Motivation: SRHS

- SRHS: self-reported health status
- Ordinal scale: excellent, very good, good, fair, poor
- Benefits:
  1. "Useful over the complete adult life cycle"
  2. Strongly correlated with objective measures
  3. Widely available (PSID, NHIS, etc.)
  4. Synthesizes all dimensions of health

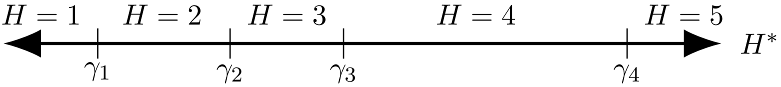
SRHS: latent model

- Latent health H^*: cardinal, continuous, of interest, but censored
- SRHS H, fixed thresholds γ_j



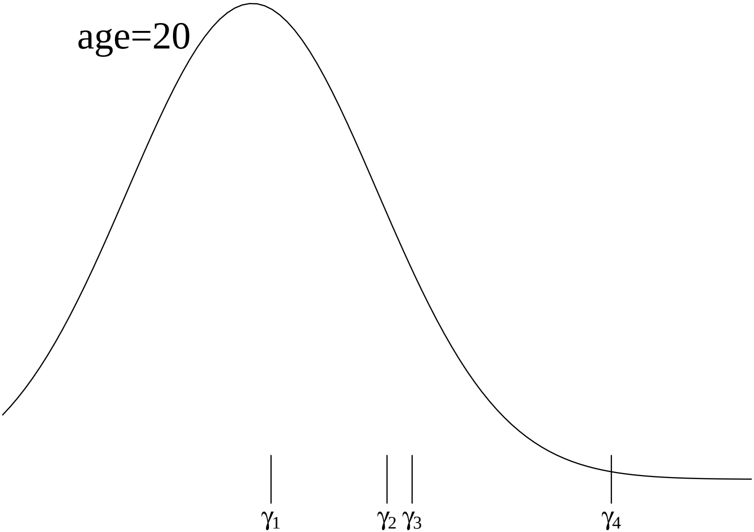
SRHS: latent model

- Latent health H^*: cardinal, continuous, of interest, but censored
- SRHS H, fixed thresholds γ_j

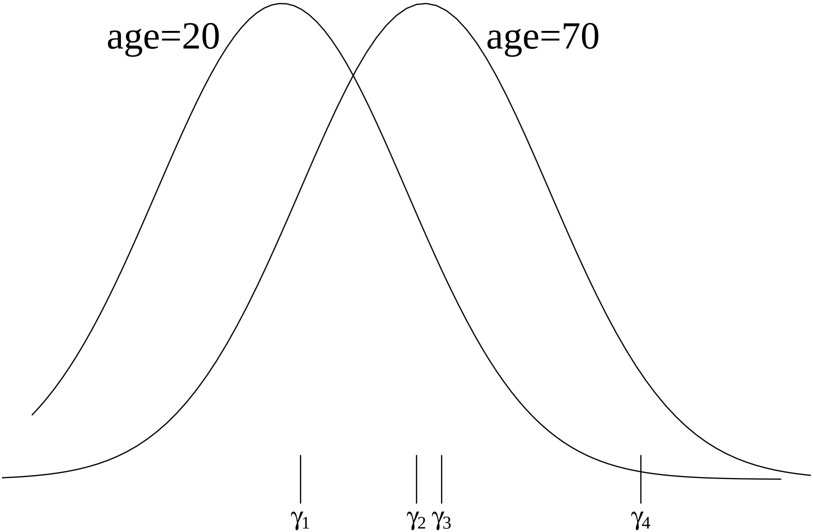


- Why isn't ordinal SRHS distribution of ultimate interest?
  1. No cardinal meaning
  2. Ignores within-category variation

SRHS: simulated pure latent location shift



SRHS: simulated pure latent location shift



SRHS: simulated pure latent location shift

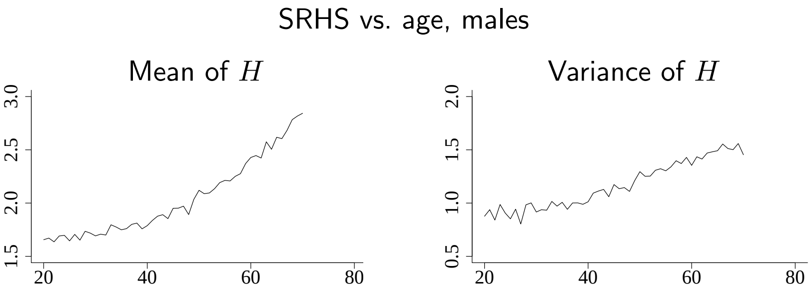
SRHS vs. age, males

Mean of H

Variance of H

DGP: for ages $a = 20, \ldots, 70$, sample 1000 iid $\mathrm{N}(\mu_a, 1)$ each for increasing μ_a, convert to ordinal using fixed thresholds

SRHS: simulated pure latent location shift



DGP: for ages $a = 20, \ldots, 70$, sample 1000 iid $\mathrm{N}(\mu_a, 1)$ each for increasing μ_a, convert to ordinal using fixed thresholds

SRHS: empirical

SRHS vs. age, males

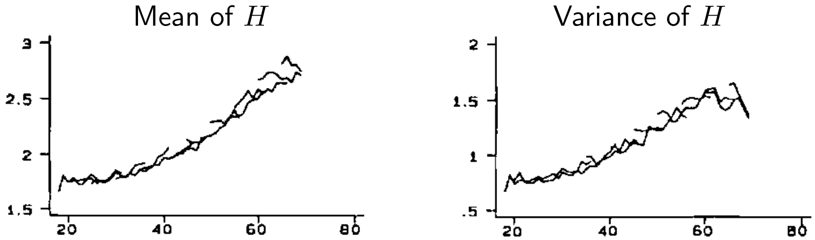


Fig. 10.4 in Deaton and Paxson (1998b), using National Health Interview Survey (NHIS) data, 1983–94

Literature: SRHS inequality methodology

- Goal: compare latent health distributions using ordinal data
- Literature: parametric/MLE or discrete latent distribution
- Literature: statistical inference rare

Literature: SRHS inequality methodology

- Goal: compare latent health distributions using ordinal data
- Literature: parametric/MLE or discrete latent distribution
- Literature: statistical inference rare
- Allison and Foster (2004): "median-preserving spread," called "the breakthrough in analyzing inequality with [SRHS] data" by Madden (2014, p. 206)

Literature: SRHS inequality methodology

- Goal: compare latent health distributions using ordinal data
- Literature: parametric/MLE or discrete latent distribution
- Literature: statistical inference rare
- Allison and Foster (2004): "median-preserving spread," called "the breakthrough in analyzing inequality with [SRHS] data" by Madden (2014, p. 206)
- SRHS-based inequality indexes:
  - Among others: Abul Naga and Yalcin (2008), Lazar and Silber (2013), Lv, Wang, and Xu (2015), and Yalonetzky (2016)
  - Good: "complete ordering" of distributions
  - Bad: many possible indexes/weights

Literature: SRHS inequality methodology

Some quotations from David Madden's chapter in the Encyclopedia of Health Economics (2014):

Literature: SRHS inequality methodology

Some quotations from David Madden's chapter in the Encyclopedia of Health Economics (2014):

- "In the literature there is still only a limited number of indices specifically designed for ordinal data": and, most treat latent distribution as discrete
- "For the case of ordinal health measures, which are arguably more widely employed, dominance results are generally less applicable, there are fewer inequality indices and statistical inference is less well developed"
- "The breakthrough in analyzing inequality with [SRHS] data came from Allison and Foster (2004)"

Our contribution

- Identification: characterize ordinal conditions informative about latent inequality
  - Semi/nonparametric restrictions (not parametric)
  - Continuous latent distribution (not discrete)

Our contribution

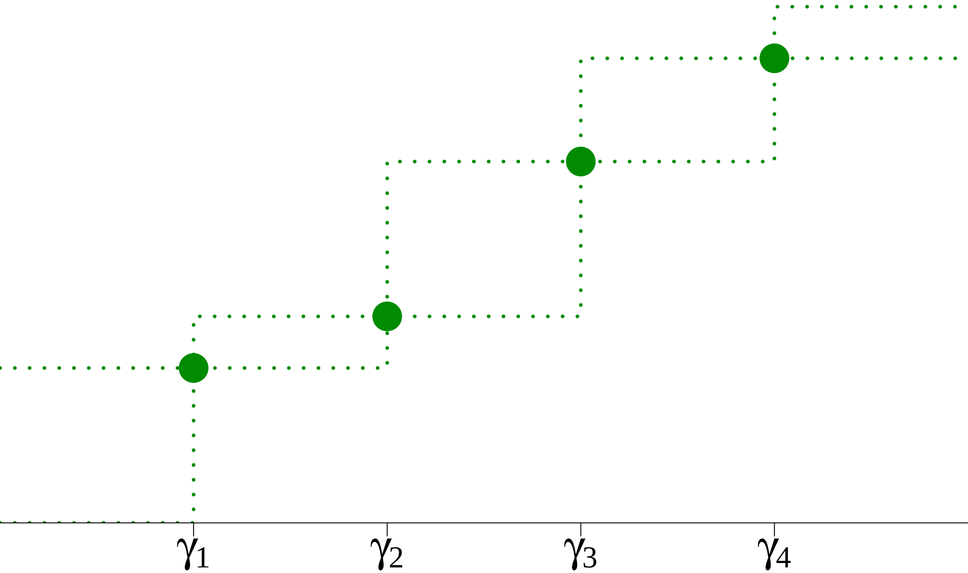
- Identification: characterize ordinal conditions informative about latent inequality
  - Semi/nonparametric restrictions (not parametric)
  - Continuous latent distribution (not discrete)
- Inference: frequentist and Bayesian (example code)
- Inference: separate paper characterizing frequentist properties of Bayesian inequality tests

Outline

Partial identification

- Given H distributions, what can we learn about H^* distributions?
- Latent CDFs partially identified, given γ_j
- Related: Stoye (2010)

Partial identification



Thresholds: assumptions

- Do all populations have same γ_j thresholds?
- Evidence of "yes" or "constant shift": Lindeboom and van Doorslaer (2004), Hernández-Quevedo, Jones, and Rice (2005)

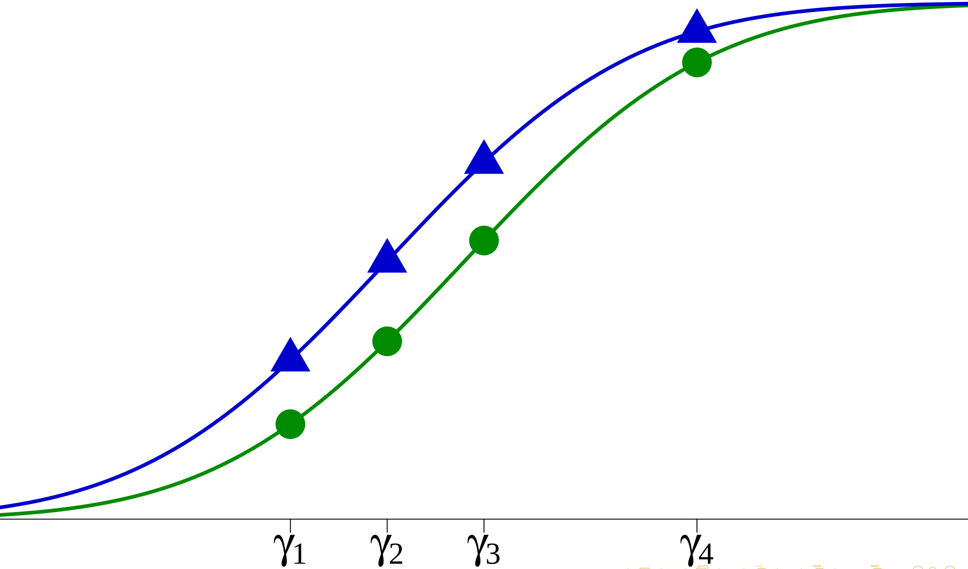
1st-order stochastic dominance (SD1): Theorem 2

- Between-group inequality
- Same γ_j

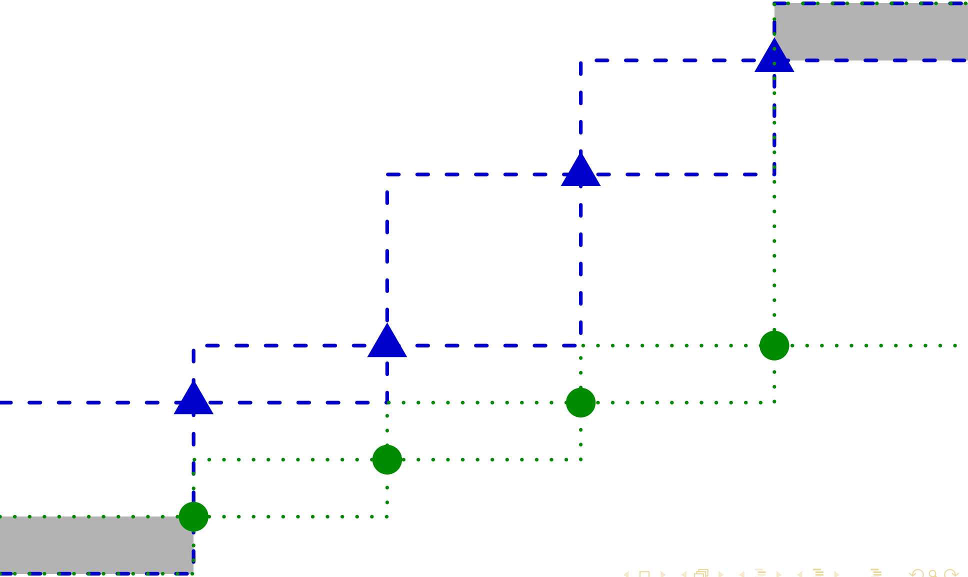
1st-order stochastic dominance (SD1): Theorem 2

- Between-group inequality
- Same γ_j
- "Healthier" if latent 1st-order stochastic dominance (SD1)
  - $X \text{ SD}_1 Y \iff F_X(\cdot) \le F_Y(\cdot)$

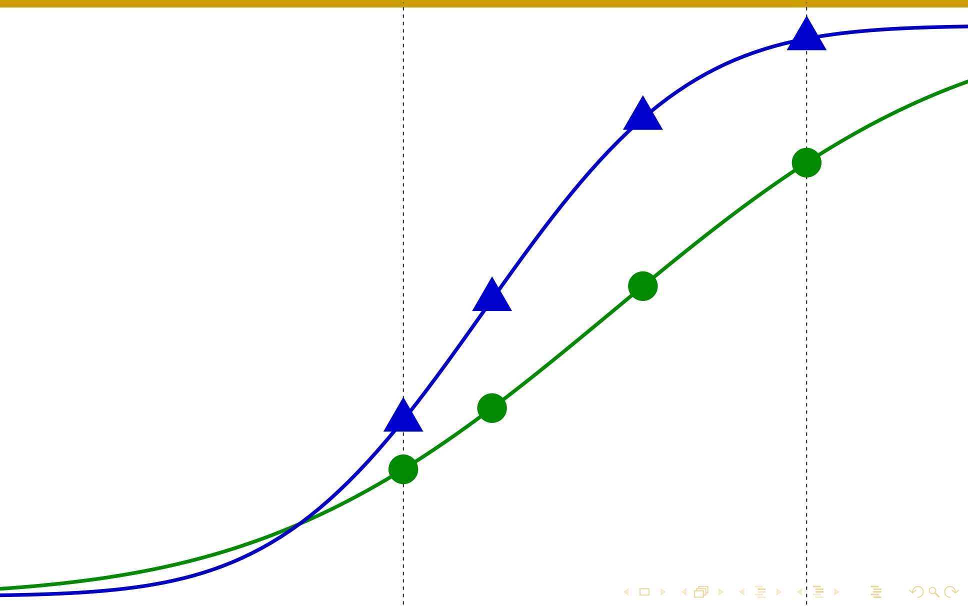
SD1: Thm 2(i)



SD1: Thm 2(v)



SD1: Thm 2(vi)



SD1: Thm 2

- Thm 2(i): latent SD1 $\Longrightarrow$ ordinal SD1, but not $\Longleftarrow$
  - Reject ordinal SD1 $\Longrightarrow$ reject latent SD1

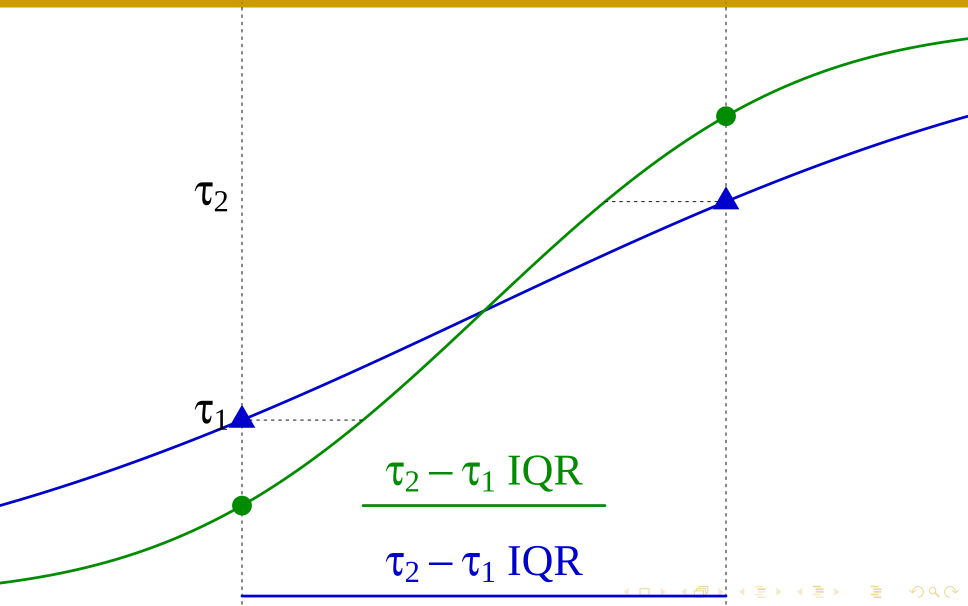
SD1: Thm 2

- Thm 2(i): latent SD1 $\Longrightarrow$ ordinal SD1, but not $\Longleftarrow$
  - Reject ordinal SD1 $\Longrightarrow$ reject latent SD1
- Thm 2(v,vi): ordinal SD1 $\Longrightarrow$ latent "restricted SD1" (Atkinson, 1987).

Dispersion: Theorem 3 (CDF crossing)

- Within-group inequality
- Pure location shift has zero effect on "dispersion"
- Similarly, γ_j can all shift by a constant (unlike for SD1)

Dispersion: Thm 3(i,ii)



Dispersion: Thm 3

- Thm 3(i): can learn about latent interquantile range (IQR) differences
- Thm 3(ii): location–scale $\implies$ extrapolate to all IQR differences
- Thm 3(iii): even stronger assumptions $\implies$ latent SD2

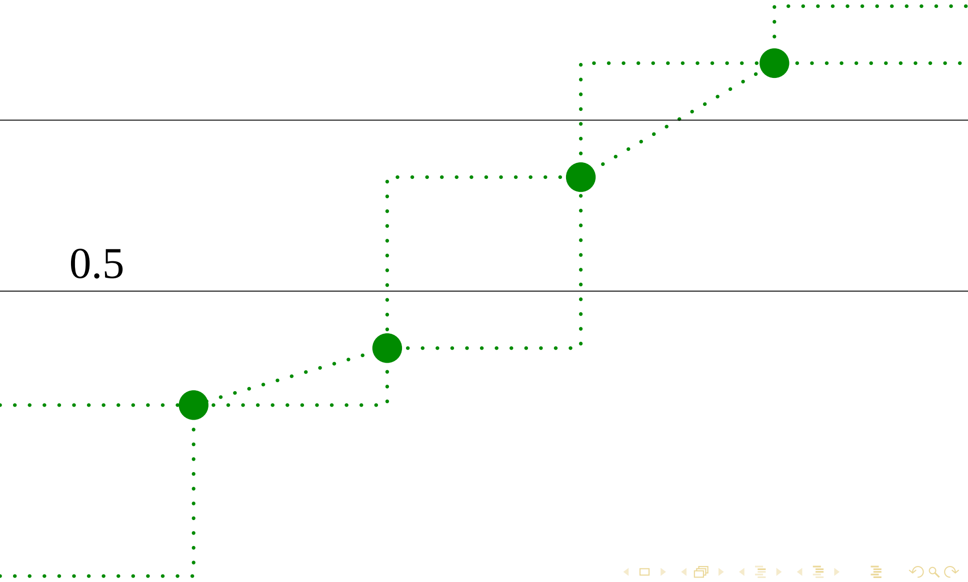
Dispersion: Theorem 4 (fanning out)

- Can we ever infer dispersion changes without a CDF crossing?

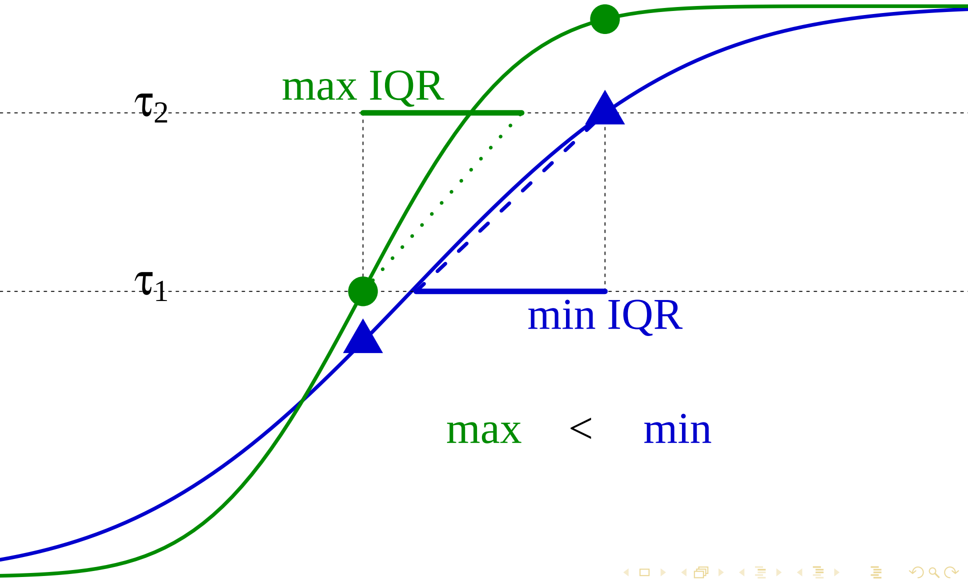
Dispersion: Theorem 4 (fanning out)

- Can we ever infer dispersion changes without a CDF crossing?
- Yes, if symmetric, unimodal latent distributions
- Location–scale $\Longrightarrow$ extrapolation

Dispersion: Thm 4(i)



Dispersion: Thm 4(i)



Outline

Ordinal SD1 inference

- Let $\theta_j \equiv F_2(j) - F_1(j) \equiv \mathrm{P}(H_2 \le j) - \mathrm{P}(H_1 \le j)$
- Ordinal SD1: $H_2 \mathrm{SD}_1 H_1 \iff \theta_j \le 0,\ j = 1, 2, 3, 4$

Ordinal SD1 inference

- Let $\theta_j \equiv F_2(j) - F_1(j) \equiv \mathrm{P}(H_2 \leq j) - \mathrm{P}(H_1 \leq j)$
- Ordinal SD1: $H_2 \ \mathrm{SD}_1 \ H_1 \iff \theta_j \leq 0, \ j = 1, 2, 3, 4$
- Frequentist: recent moment inequality tests from Andrews and Barwick (2012), Romano, Shaikh, and Wolf (2014), McCloskey (2015), et al.
- Bayesian: "nonparametric" posterior for category probabilities $\Longrightarrow$ posterior probabilities for all relationships
  - iid: Dirichlet–multinomial model; category probability vectors have Dirichlet posterior

Other relationships

- Unions and/or intersections of inequalities
- Bayes: just compute posteriors
- Frequentist: intersection–union test (sometimes)

Bayes vs. frequentist?

- How do Bayesian and frequentist inference compare for these various relationships? Or more generally?
- Kline (2011, Lemma 1.4): Bayesian posterior probability of ordinal SD1 is (stochastically) more aggressive than a p-value (or generally for $H_0 \colon \boldsymbol{\theta} \leq \mathbf{0}$)
- Efron and Tibshirani (1998, p. 1690): depends if null region "curves away from" or "curves toward" point estimate; but theory based on asymptotic approx of smooth boundary

Bayes vs. frequentist: setting

- $H_0 : \boldsymbol{\theta} \in \Theta_0$ (vs. $\boldsymbol{\theta} \notin \Theta_0$)
- Single draw, $\mathbf{X} - \boldsymbol{\theta} \mid \boldsymbol{\theta} \sim \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}})$, $\boldsymbol{\theta} - \mathbf{X} \mid \mathbf{X} \sim \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}})$ [can weaken somewhat]
- Bayes test: reject H_0 iff posterior of H_0 below α
  - i.e., treat as p-value
  - (generalized) Bayes rule given loss function with $1 - \alpha$ for type I error, α for type II
- Frequentist properties of Bayes test?

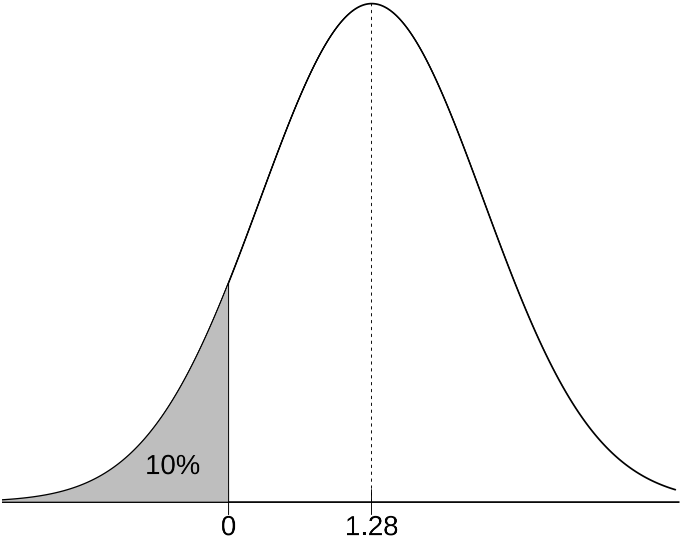
Bayes vs. frequentist: more motivation

- Generalizing the ordinal/latent relationships here: Kline and Tamer (2016, p. 330)
- Does a certain property hold for all possible partially identified parameter values (or, at least one) within the identified set?
  - E.g., H_0: ordinal non-SD1 $\iff$ "There is no latent distribution pair in the identified set that satisfies SD1"
  - E.g., H_0: ordinal SD1 $\iff$ "There is at least one latent distribution pair in the identified set that satisfies SD1"

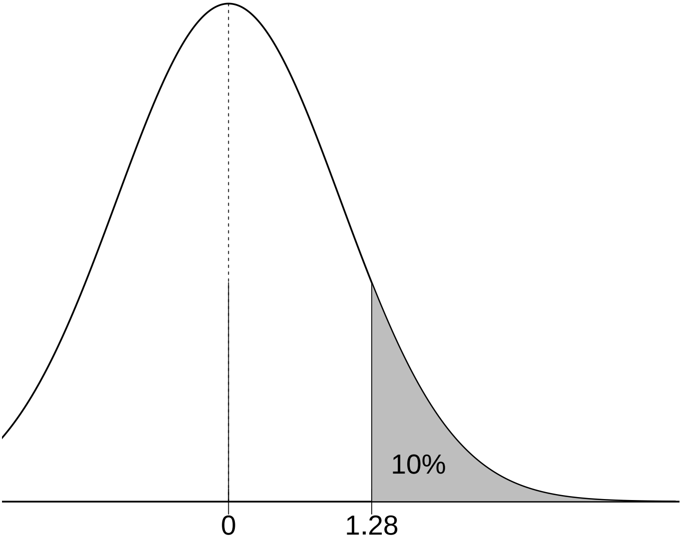
Bayes vs. frequentist: more motivation

- Generalizing the ordinal/latent relationships here: Kline and Tamer (2016, p. 330)
- Does a certain property hold for all possible partially identified parameter values (or, at least one) within the identified set?
  - E.g., H_0: ordinal non-SD1 $\iff$ "There is no latent distribution pair in the identified set that satisfies SD1"
  - E.g., H_0: ordinal SD1 $\iff$ "There is at least one latent distribution pair in the identified set that satisfies SD1"
- Or mis/specification test: is point identified parameter in region corresponding to non/empty identified set?

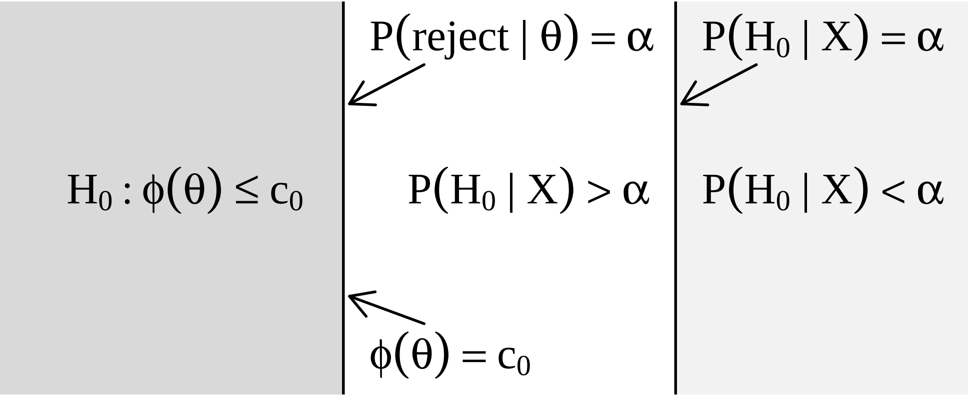
Theorem 1(i) for $H_0\colon \theta \leq 0$



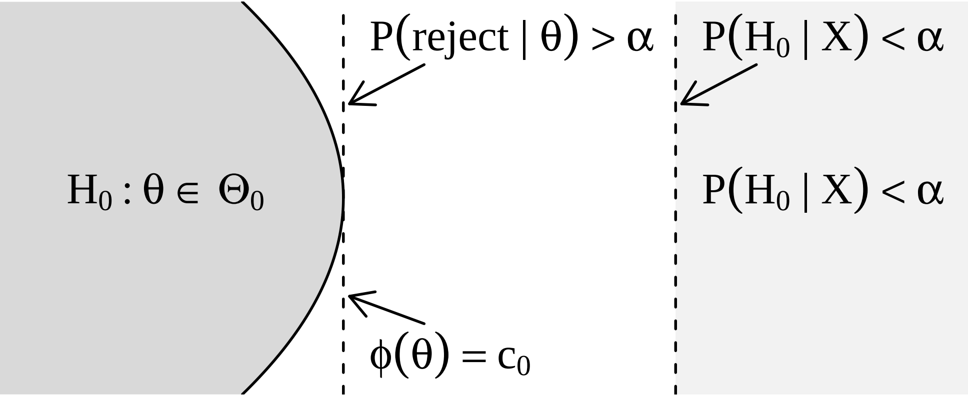
Theorem 1(i) for $H_0\colon \theta \leq 0$



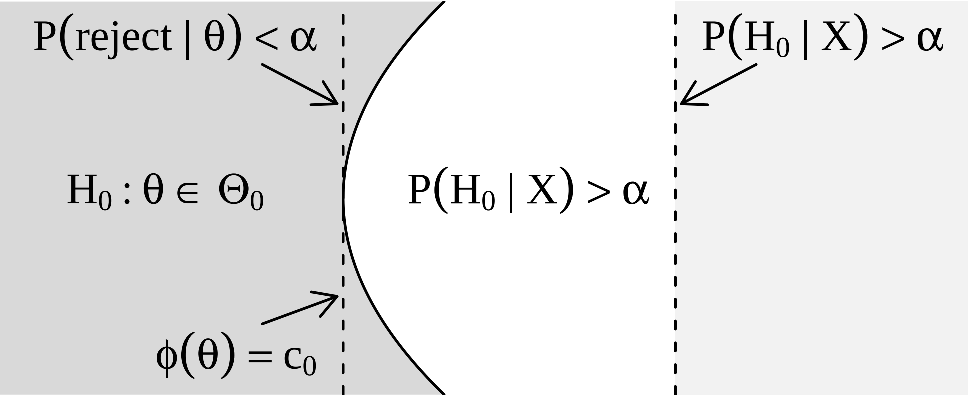
Theorem 1(i) [again]



Theorem 1(iii)



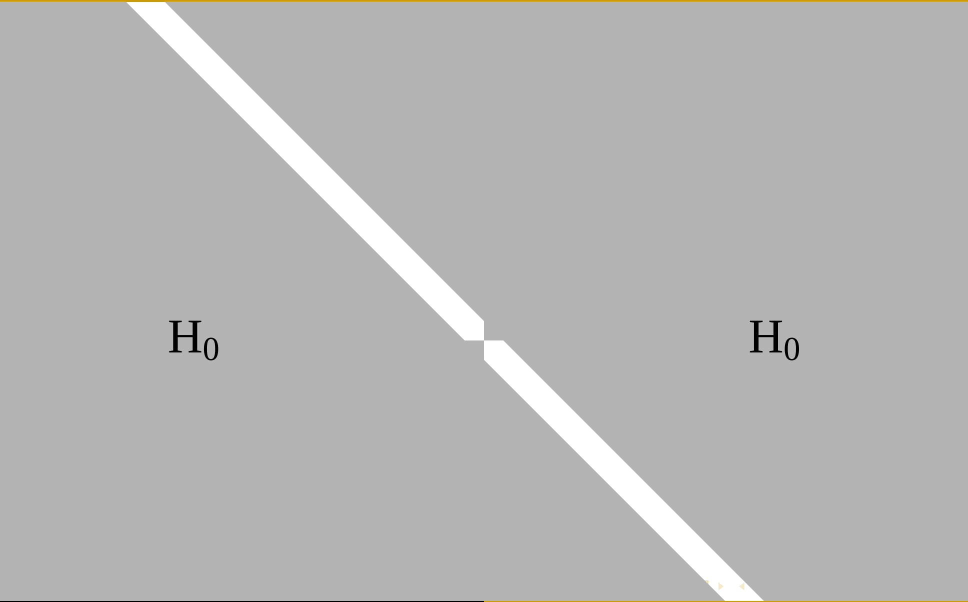
Theorem 1(v)



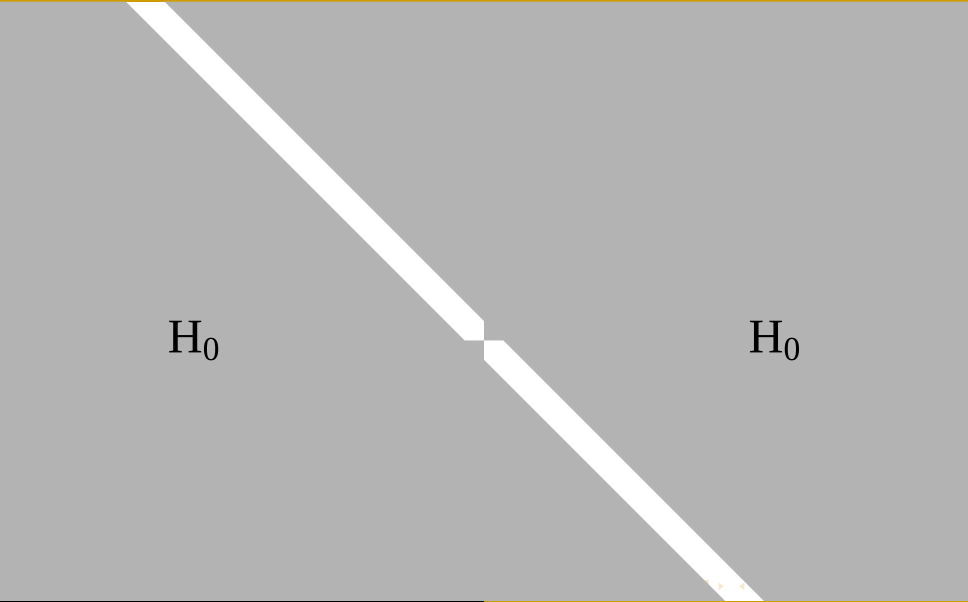
Theorem 1(iv)

- If Θ_0 not contained in half-space, then size$\{>,=,<\}\alpha$
- Further: may depend on distribution, not just shape of Θ_0
- Examples follow: bivariate normal distribution, unit variances, correlation ρ

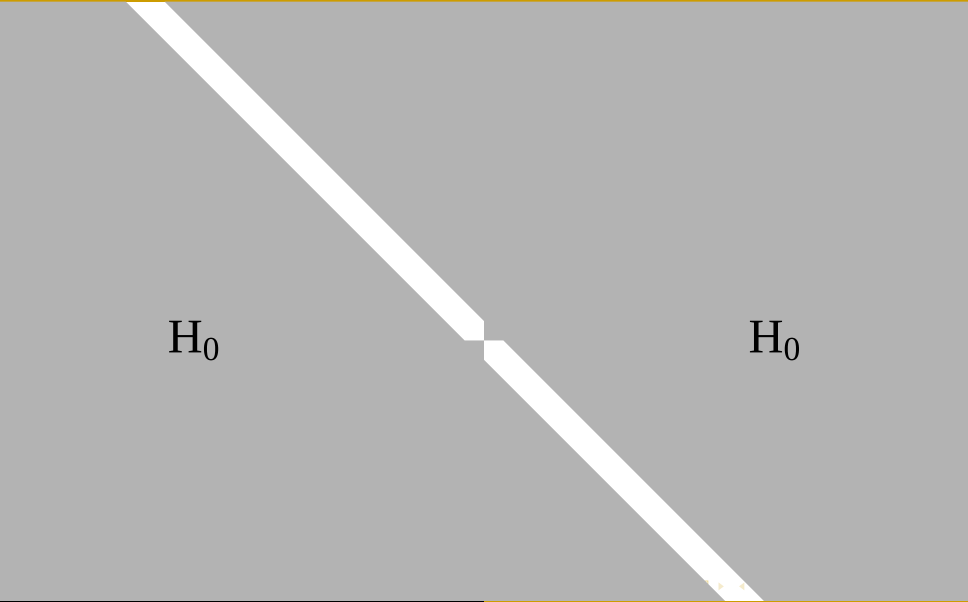
Thm 1(iv): example of size depending on ρ



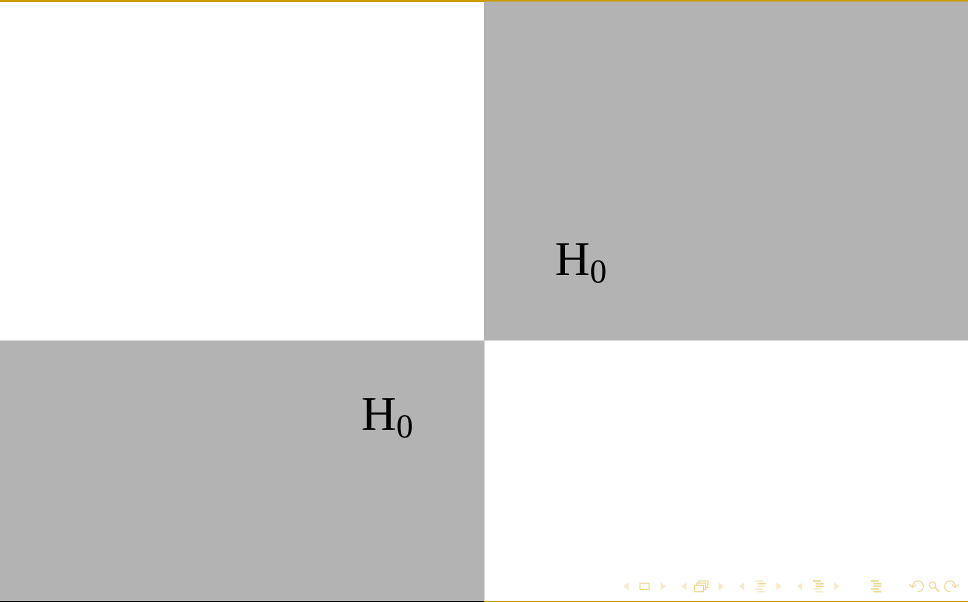
Thm 1(iv): size is 0% if $\rho = 1$ ($\mathrm{P}(H_0 \mid \mathbf{X}) \approx 1$, $\forall \mathbf{X}$)



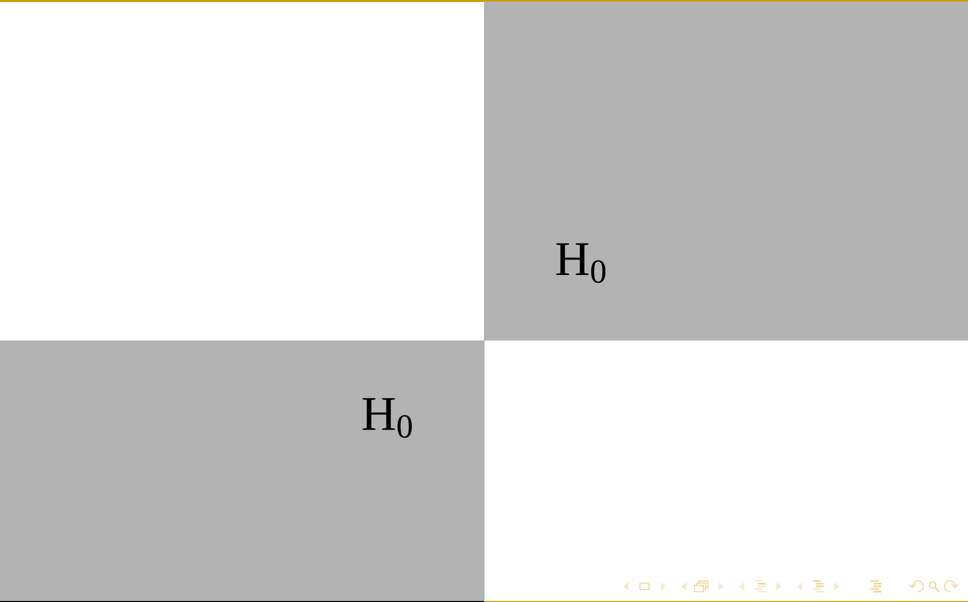
Thm 1(iv): size is 100% if $\rho = -1$ (set $\theta_1 = \theta_2 = 0$)



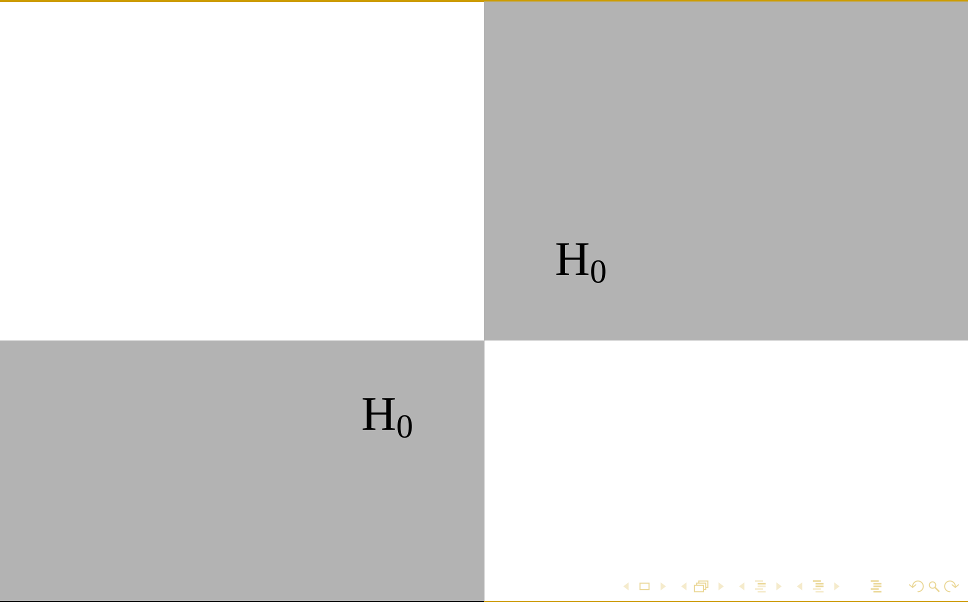
Thm 1(iv): example of $H_0\colon \theta_1\theta_2 \geq 0$



Thm 1(iv): size is 100% if $\rho = -1$ (set $\theta_1 = \theta_2 = 0$)



Thm 1(iv): size is α if $\rho = 1$ (let $\theta_1 \to \infty$, $\theta_2 = 0$)



Outline

Comparisons of U.S. states

- Goal: compare latent health across states
- Data: 2011 PSID, $\sim 50\text{–}300$ observations per state
- Posteriors: Bayesian bootstrap of Dong, Elliott, and Raghunathan (2014) (stratification, clustering, weights)

PSID 2011 posterior probabilities (%)

		SD_1		SC		fans out	
X	Y	$\succ$	$\prec$	$\succ$	$\prec$	$\succ$	$\prec$
AZ	MO	0	90	4	2	11	34
NY	UT	0	3	0	94	1	70
IL	NY	20	0	66	1	92	3
MN	NY	24	0	57	0	96	3
IA	MO	0	10	2	16	42	98

PSID 2011 posterior probabilities (%)

	X SD$_1$ Y; Y is:					X SC Y; Y is:				
X	MO	KS	NE	IA	IL	MO	KS	NE	IA	IL
MO	—	0	10	10	0	—	7	67*	16	30
KS	34*	—	20*	10	3	6	—	24	6	16
NE	3	0	—	6	0	2	0	—	3	0
IA	0	0	6	—	0	2	0	66*	—	6
IL	40*	4	18*	43*	—	4	14	48	15	—

Asterisk (*): satisfied in-sample

Outline

Purpose and setup

- Goal: compare type I error rates of different frequentist and Bayesian tests
- DGP 1: $\mathrm{P}(X = j) = \mathrm{P}(Y = j) = 1/5$, $j = 1, \ldots, 5$; all SD1 inequalities binding
- DGP 2: change to $\mathrm{P}(X = j) = 1/10$ for $j = 1, 2, 3$ and $\mathrm{P}(X = 4) = 1/2$; only one binding inequality

Methods

- KS: Kolmogorov–Smirnov
- RMS: refined moment selection of Andrews and Barwick (2012)
- Bayes: Dirichlet–multinomial, uninformative prior on parameters, reject if posterior below α
- adj: adjust prior to $\mathrm{P}(H_0) = 1/2$ (Goutis, Casella, and Wells, 1996)

Results: $\alpha = 0.1$

		$H_0\colon X \text{ SD}_1\ Y$				$H_0\colon X \text{ SC } Y$		
DGP	n	KS	RMS	Bayes	adj	RMS	Bayes	adj
1	50	0.038	0.089	0.436	0.204	0.032	0.439	0.175
1	100	0.022	0.084	0.430	0.205	0.029	0.359	0.142
1	500	0.027	0.092	0.447	0.199	0.034	0.428	0.171
1	1000	0.032	0.079	0.454	0.228	0.032	0.408	0.155
2	50	0.004	0.057	0.127	0.032	0.031	0.125	0.034
2	100	0.002	0.068	0.105	0.031	0.085	0.133	0.041
2	500	0.006	0.087	0.098	0.029	0.095	0.114	0.032
2	1000	0.003	0.074	0.084	0.025	0.060	0.084	0.018

RMS better than KS

Even with adj, cannot treat posterior as p-value, or vice-versa

More DGPs

- Goal: importance of "local" vs. "global" shape
- Let $\theta_j \equiv F_X(j) - F_Y(j)$ [ordinal]
- SC of Y by X requires $\theta_1 < 0 < \theta_4$ (quadrant II)
- SC of Y by X violated by $\theta_2 > 0 > \theta_3$, but quadrants I, II, III of (θ_2, θ_3) ok
- DGP 1: at "convex corner" of Θ_0 with $F_X(1) = F_Y(1) = 0.2$, $F_X(2) = 0.39 < 0.4 = F_Y(2)$, $F_X(3) = 0.5 < 0.6 = F_Y(3)$, and $F_X(4) = F_Y(4) = 0.8$
- DGP 2: at local non-convexity of Θ_0 with $F_X(1) = 0.18 < 0.22 = F_Y(1)$, $F_X(2) = F_Y(2) = 0.4$, $F_X(3) = F_Y(3) = 0.6$, and $F_X(4) = 0.82 > 0.78 = F_Y(4)$
- Both on boundary of Θ_0

Results: $\alpha = 0.1$

		H_0: SC		H_0: non-SC	
DGP	n	RP	RP (adj)	RP	RP (adj)
1	20	0.422	0.170	0.000	0.014
1	100	0.484	0.186	0.000	0.010
1	500	0.531	0.249	0.000	0.008
2	20	0.246	0.056	0.001	0.050
2	100	0.128	0.022	0.009	0.148
2	500	0.011	0.002	0.130	0.501

Outline

Conclusion

- Summary: link latent and ordinal inequality, without parametric or discrete assumptions
- Summary: relationships defined by inequalities; frequentist and Bayesian inference (code on website)
- Summary: characterize rejection probabilities of a Bayesian test of $H_0 : \boldsymbol{\theta} \in \Theta_0$

Conclusion

- Summary: link latent and ordinal inequality, without parametric or discrete assumptions
- Summary: relationships defined by inequalities; frequentist and Bayesian inference (code on website)
- Summary: characterize rejection probabilities of a Bayesian test of $H_0 : \boldsymbol{\theta} \in \Theta_0$
- Future work: other shape restrictions? multivariate? regression? health, happiness, . . . ?
- Thank you!
- (And further questions or comments are welcome)

References I

Abul Naga, R. H., Yalcin, T., 2008. Inequality measurement for ordered response health data. Journal of Health Economics 27 (6), 1614–1625.
URL https://doi.org/10.1016/j.jhealeco.2008.07.015

Allison, R. A., Foster, J. E., 2004. Measuring health inequality using qualitative data. Journal of Health Economics 23 (3), 505–524.
URL https://doi.org/10.1016/j.jhealeco.2003.10.006

Andrews, D. W. K., Barwick, P. J., 2012. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. Econometrica 80 (6), 2805–2826.
URL https://www.jstor.org/stable/23357242

References II

Atkinson, A. B., 1987. On the measurement of poverty. Econometrica 55 (4), 749–764.
URL https://doi.org/10.2307/1911028

Deaton, A., Paxson, C., 1998a. Aging and inequality in income and health. American Economic Review (Papers and Proceedings) 88 (2), 248–253.
URL https://www.jstor.org/stable/116928

Deaton, A., Paxson, C., 1998b. Health, income, and inequality over the life cycle. In: Frontiers in the Economics of Aging. University of Chicago Press, pp. 431–462.
URL https://www.nber.org/chapters/c7309

References III

Dong, Q., Elliott, M. R., Raghunathan, T. E., 2014. A nonparametric method to generate synthetic populations to adjust for complex sampling design features. Survey Methodology 40 (1), 29.
URL https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400114003

Efron, B., Tibshirani, R., 1998. The problem of regions. Annals of Statistics 26 (5), 1687–1718.
URL https://projecteuclid.org/euclid.aos/1024691353

Goutis, C., Casella, G., Wells, M. T., 1996. Assessing evidence in multiple hypotheses. Journal of the American Statistical Association 91 (435), 1268–1277.
URL https://www.jstor.org/stable/2291745

References IV

Hernández-Quevedo, C., Jones, A. M., Rice, N., 2005. Reporting bias and heterogeneity in self-assessed health. evidence from the British Household Panel Survey. HEDG Working Paper 05/04, Health, Econometrics and Data Group, The University of York. URL https://ideas.repec.org/p/yor/hectdg/05-04.html

Kline, B., 2011. The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean. Journal of Statistical Planning and Inference 141 (9), 3131–3141. URL https://doi.org/10.1016/j.jspi.2011.03.034

Kline, B., Tamer, E., 2016. Bayesian inference in a class of partially identified models. Quantitative Economics 7 (2), 329–366. URL https://doi.org/10.3982/QE399

References V

Lazar, A., Silber, J., 2013. On the cardinal measurement of health inequality when only ordinal information is available on individual health status. Health Economics 22 (1), 106–113.
URL https://doi.org/10.1002/hec.1821

Lindeboom, M., van Doorslaer, E., 2004. Cut-point shift and index shift in self-reported health. Journal of Health Economics 23 (6), 1083–1099.
URL https://doi.org/10.1016/j.jhealeco.2004.01.002

Lv, G., Wang, Y., Xu, Y., 2015. On a new class of measures for health inequality based on ordinal data. Journal of Economic Inequality 13 (3), 465–477.
URL https://doi.org/10.1007/s10888-014-9289-4

References VI

Madden, D., 2014. Dominance and the measurement of inequality. In: Culyer, A. J. (Ed.), Encyclopedia of Health Economics. Vol. 1. Elsevier, pp. 204–208.
URL https://doi.org/10.1016/B978-0-12-375678-7.00725-2

McCloskey, A., 2015. On the computation of size-correct power-directed tests with null hypotheses characterized by inequalities, working paper, available at http://www.brown.edu/Departments/Economics/Faculty/Adam_McCloskey/Research.html.

Romano, J. P., Shaikh, A. M., Wolf, M., 2014. A practical two-step method for testing moment inequalities. Econometrica 82 (5), 1979–2002.
URL https://www.jstor.org/stable/24029299

References VII

Stoye, J., 2010. Partial identification of spread parameters. Quantitative Economics 1 (2), 323–357.
URL https://doi.org/10.3982/QE24

Yalonetzky, G., 2016. Robust ordinal inequality comparisons with Kolm-independent measures. Working Paper 401, ECINEQ, Society for the Study of Economic Inequality.
URL https://ideas.repec.org/p/inq/inqwps/ecineq2016-401.html