MGMT MFE 431-3

Statistical Arbitrage

Lecture 09: Advanced Shrinkage

Professor Olivier Ledoit

University of California Los Angeles
Anderson School of Management
Master of Financial Engineering

Fall 2019

What Did We Learn Last Thursday?

Plan of Lecture 09

1. Introduction
2. General Overview of Cornell-Welch Model
3. Analysis without other-group candidates
4. Analysis with other-group candidates
5. Evidence from a Natural Experiment
6. Statistical Analysis: The Bootstrap
7. Conclusion

Advanced Covariance Matrix Shrinkage

Accepted by the Annals of Statistics

ANALYTICAL NONLINEAR SHRINKAGE OF LARGE-DIMENSIONAL COVARIANCE MATRICES

By Olivier Ledoit and Michael Wolf

University of Zurich

This paper establishes the first analytical formula for nonlinear shrinkage estimation of large-dimensional covariance matrices. We achieve this by identifying and mathematically exploiting a deep connection between nonlinear shrinkage and nonparametric estimation of the Hilbert transform of the sample spectral density. Previous nonlinear shrinkage methods were of numerical nature: QuEST requires numerical inversion of a complex equation from random matrix theory whereas NERCOME is based on a sample-splitting scheme. The new analytical method is more elegant and also has more potential to accommodate future variations or extensions. Immediate benefits are (i) that it is typically 1,000 times faster with basically the same accuracy as QuEST and (ii) that it accommodates covariance matrices of dimension up to $10,000$ and more. The difficult case where the matrix dimension exceeds the sample size is also covered.

Key Talking Points

- Kernel estimation of the density of sample covariance matrix eigenvalues
  - And of its Hilbert transform
  - Using the Epanechnikov kernel
  - And locally adaptive bandwidth
- Combine with a generalization of the Marčenko-Pastur (1967) equation
- Can estimate 10,000-dimensional covariance matrix in 2 minutes!

We are going to do it the easy way..

- Advanced shrinkage of the vector (not the matrix)

- First anchor point: Matlab program shrink.m uploaded to CCLE for Lecture 05

- Second anchor point: spreadsheet ShrinkHedge2011P.xls on hedge-fund-of-funds case study uploaded to CCLE for Lecture 04

Plan of Lecture 09

1. Introduction
2. General Overview of Cornell-Welch Model
3. Analysis without other-group candidates
4. Analysis with other-group candidates
5. Evidence from a Natural Experiment
6. Statistical Analysis: The Bootstrap
7. Conclusion

MENU

THE UNIVERSITY OF CHICAGO PRESS JOURNALS

Journal of Political Economy

Volume 104, Number 3 | Jun., 1996

SUBSCRIBE/RENEW | BROWSE ISSUES | FORTHCOMING | CONTRIBUTORS

Previous Article | Next Article

Culture, Information, and Screening Discrimination

Bradford Cornell, and **Ivo Welch**

Abstract | Cited by | PDF

Abstract

We show that discrimination can occur even when it is common knowledge that underlying group characteristics do not differ and when employers do not prefer same-group candidates. When employers can judge job applicants' unknown qualities better when candidates belong to the same group and hire the best prospect from a large pool of applicants, the top applicant is likely to have the same background as the

Main Qualitative Features

1. Cultural type: interpreted broadly to include groups defined by language, religious belief, ethnic background, race, sex, sexual preference, neighborhood upbringing, schooling, or membership in social organization
2. Employer comes from some group (type)
3. Candidates come either from the same group as the employer or from some other group
4. Employer can distinguish between high- and low-quality **more accurately** when the candidates being sorted come from the same group

What the Model Does not Contain

- No assumption that one country's quality differs from the other
- Employer is <u>not biased</u> towards working with same-country candidates

Figure 1 and Appendix B

- Continuous model: normal prior, normal signals
- *s* same-group candidates + *o* from other group
- Unobservable candidate quality: $Q_i \sim \mathcal{N}(0,1)$
- Mean 0 and variance 1: *without loss of generality*
- $\forall$i = 1,…,*s* signal: $x_i = Q_i + \sigma_s \, \varepsilon_i$
- $\forall$i=*s*+1,…, *s*+*o* signal: $x_i = Q_i + \sigma_o \, \varepsilon_i$
- Errors ε_i are i.i.d. $\sim \mathcal{N}(0,1)$
- Cultural barriers $\Rightarrow \sigma_s \leq \sigma_o$

Plan of Lecture 09

1. Introduction
2. General Overview of Cornell-Welch Model
3. Analysis without other-group candidates
4. Analysis with other-group candidates
5. Evidence from a Natural Experiment
6. Statistical Analysis: The Bootstrap
7. Conclusion

Inferred Quality

- Candidate quality Q_i is unobservable
- Employer observes only a noisy signal x_i
- Signal estimates quality, but can be improved
- Better to use *inferred quality* $\hat{Q}_i$=E[Q_i|x_i]
- Conditional expectation = "best guess"
- Not 100% accurate, but more so than signal x_i
- <u>Bayesian interpretation</u>: posterior mean

How to Infer Quality

<u>Theorem</u> $\hat{Q}_i$ = E[Q_i|x_i] = $\beta_s x_i$ where $\beta_s = \frac{1}{1+\sigma_s^2}$

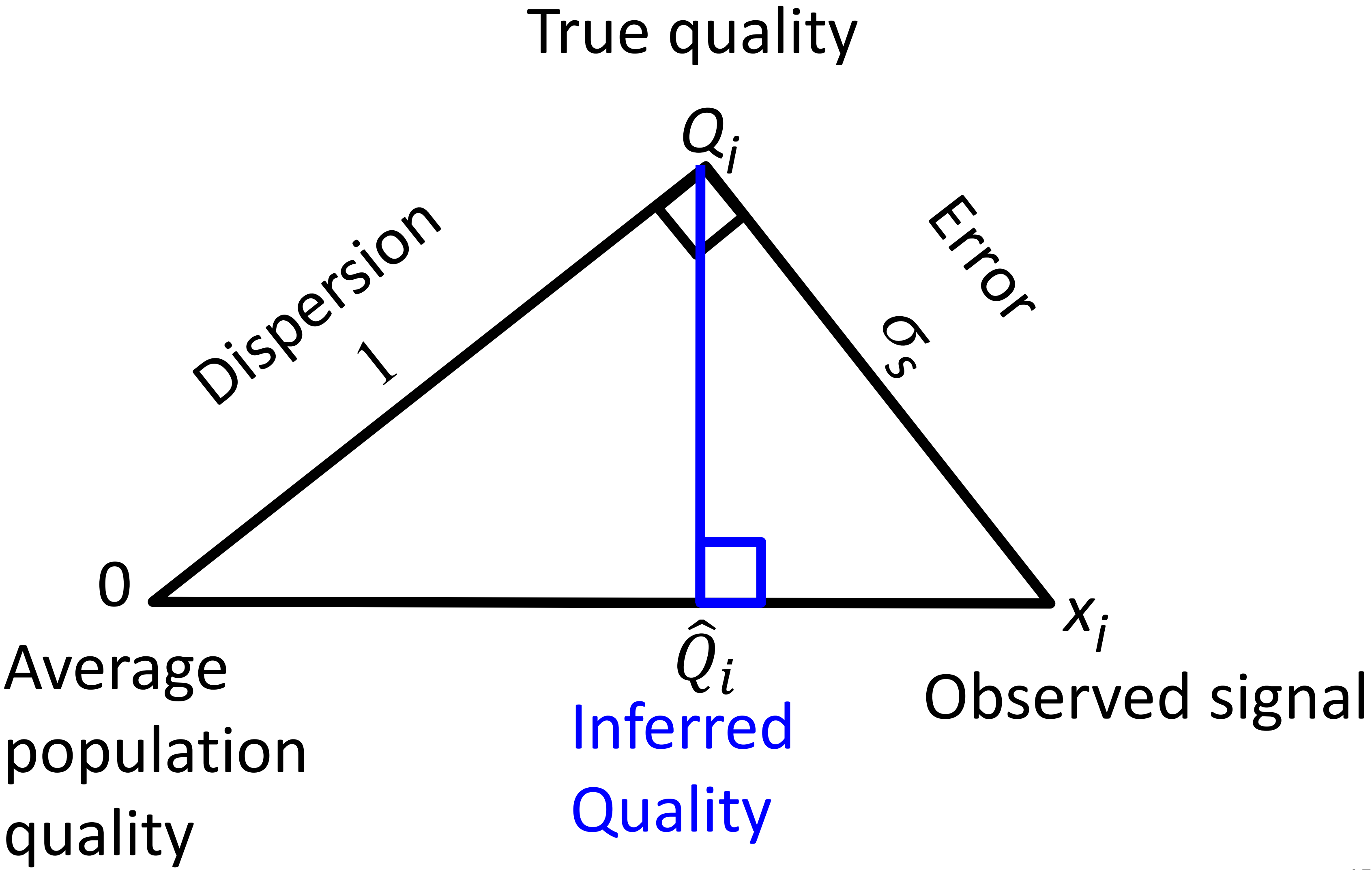
<u>Proof</u> If (y,z) are bivariate normal with mean 0:

$$E[y|z] = \frac{Cov[y,z]}{Var[z]}\, z$$

$Cov[Q_i,x_i]=Cov[Q_i,Q_i+\sigma_s\varepsilon_i]=cov[Q_i,Q_i]=var[Q_i]=1$

$Var[x_i]=Var[Q_i+\sigma_s\varepsilon_i]=Var[Q_i]+\sigma_s^2 Var[\varepsilon_i]=1+ \sigma_s^2$

Geometric Interpretation



Four Interpretations to Shrinkage

1. Bayesian interpretation: mean of the posterior distribution
2. Geometric interpretation: orthogonal projection of (unobservable) quality onto the one-dimensional space spanned by the signal
3. Decision-theoretic: it is the linear correction that reduces mean squared error the most
4. Ordinary Least Squares: Regressing quality onto a mismeasured version of itself causes *attenuation* (also known as *dilution*)

Shrinking Is Good for Accuracy

$$\mathbb{E}[(x_i - Q_i)^2] = E[(\sigma_s \varepsilon_i)^2] = \sigma_s^2$$

$$\begin{aligned}
\mathbb{E}[(\widehat{Q}_i - Q_i)^2] &= \mathbb{E}\left[\left(\frac{1}{1+\sigma_s^2} x_i - Q_i\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{1+\sigma_s^2} Q_i + \frac{\sigma_s}{1+\sigma_s^2}\varepsilon_i - Q_i\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{\sigma_s}{1+\sigma_s^2}\varepsilon_i - \frac{\sigma_s^2}{1+\sigma_s^2} Q_i\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{\sigma_s}{1+\sigma_s^2}\varepsilon_i\right)^2\right] + \mathbb{E}\left[\left(\frac{\sigma_s^2}{1+\sigma_s^2} Q_i\right)^2\right] \\
&= \frac{\sigma_s^2}{(1+\sigma_s^2)^2} + \frac{\sigma_s^4}{(1+\sigma_s^2)^2} = \frac{\sigma_s^2}{1+\sigma_s^2} = \beta\sigma_s^2 \leq \sigma_s^2
\end{aligned}$$

Unconditional Distribution of $\hat{Q}_i$

- It is normally distributed (Gaussian bell curve)
- Mean: $E[\hat{Q}_i] = E[\beta_s x_i] = \beta_s E[Q_i + \sigma_s \varepsilon_i] = 0$
- $\mathrm{Var}[\hat{Q}_i] = \beta_s^2 \mathrm{Var}[Q_i + \sigma_s \varepsilon_i]$

$$= \beta_s^2 (\mathrm{Var}[Q_i] + \sigma_s^2 \mathrm{Var}[\varepsilon_i])$$

$$= \beta_s^2 (1 + \sigma_s^2) = \beta_s = \frac{1}{1+\sigma_s^2}$$

- In summary: $\hat{Q}_i \sim \mathcal{N}(0, \beta_s)$

Cross-Sectional Dispersions

- There are **3 different** quantities *for each* candidate: observed signal x_i, unobservable quality Q_i and inferred quality $\hat{Q}_i$
- Unconditional distributions:
  - $x_i \sim \mathcal{N}(0, 1/\beta_s)$
  - $Q_i \sim \mathcal{N}(0, 1)$
  - $\hat{Q}_i \sim \mathcal{N}(0, \beta_s)$
- Signal is more dispersed than true quality, but inferred quality is <u>less dispersed</u> (by same ratio!)

Goals per Game Season 2009-2010

		Aug-Dec 2009	Jan-May 2010	Change
W. Rooney	Man Utd	0.70	0.67	↘
F. Torres	Liverpool	0.60	0.33	↘
Agbonlahor	Aston Villa	0.40	0.28	↘
R. van Persie	Arsenal	0.37	0.11	↘
F. Lampard	Chelsea	0.30	0.89	↗
N. Anelka	Chelsea	0.25	0.33	↗
S. Gerrard	Liverpool	0.25	0.22	↘
Dirk Kuyt	Liverpool	0.25	0.22	↘
John Carew	Aston Villa	0.20	0.33	↗
Kevin Davies	Bolton	0.17	0.20	↗

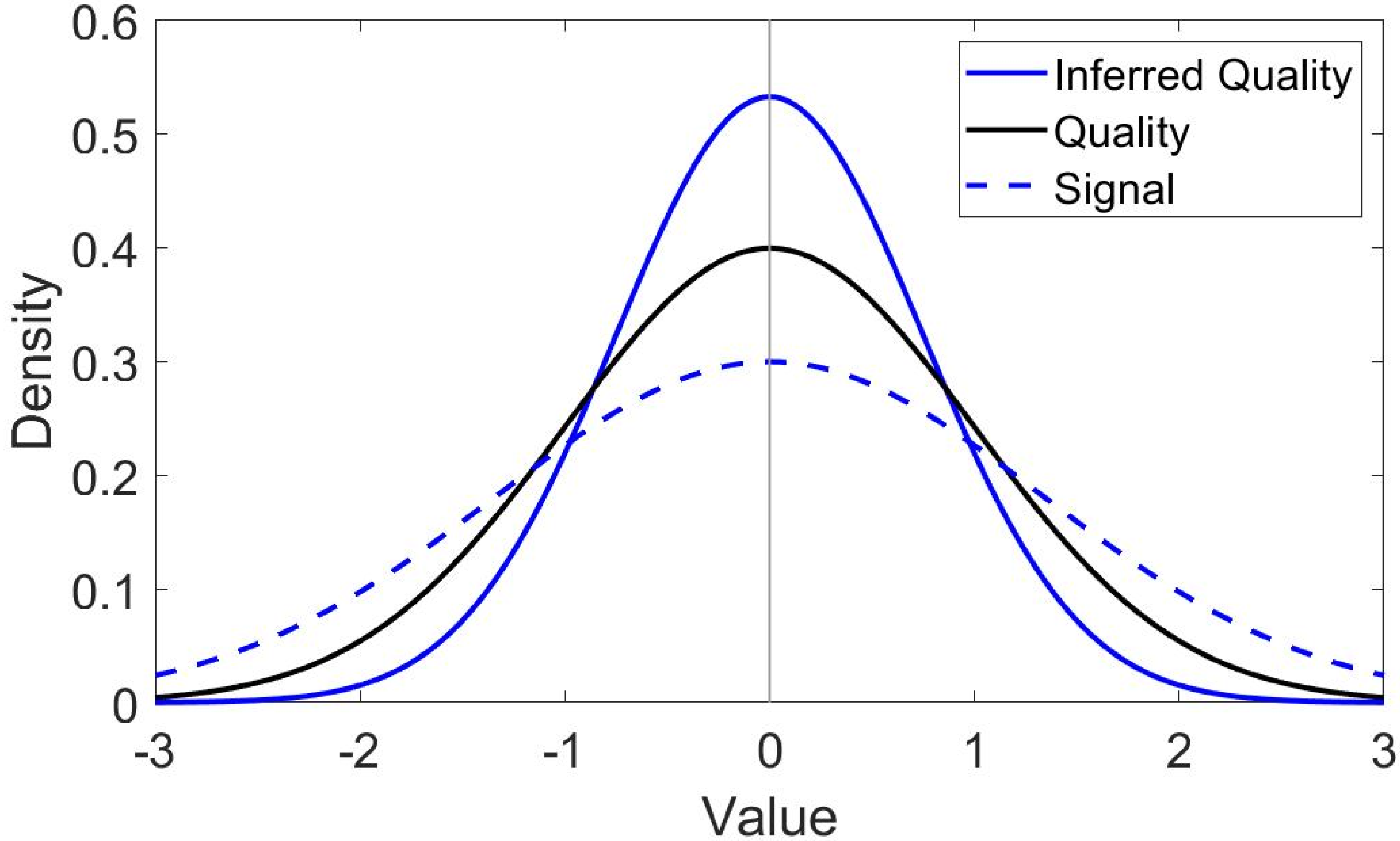
Goals per Game Season 2009-2010

		15-16 Aug 2009	Jan-May 2010
W. Rooney	Man Utd	1.00	0.67
F. Torres	Liverpool	0.00	0.33
Agbonlahor	Aston Villa	0.00	0.28
R. van Persie	Arsenal	0.00	0.11
F. Lampard	Chelsea	0.00	0.89
N. Anelka	Chelsea	0.00	0.33
S. Gerrard	Liverpool	1.00	0.22
Dirk Kuyt	Liverpool	0.00	0.22
John Carew	Aston Villa	0.00	0.33
Kevin Davies	Bolton	0.00	0.20

Intuitive Explanation: Information

- Dispersion of inferred quality is a measure of **information precision**
- Best-case scenario: zero noise $\Rightarrow$ dispersion of inferred quality = 1
- Worst-case scenario: infinite noise $\Rightarrow$ dispersion of inferred quality = 0
- Realistic scenario: some amount of noise $\Rightarrow$ dispersion of inferred quality between 0 and 1
- In same-group case, shrinkage *preserves ordering*

Comparing Densities



Plan of Lecture 09

1. Introduction
2. General Overview of Cornell-Welch Model
3. Analysis without other-group candidates
4. Analysis with other-group candidates
5. Evidence from a Natural Experiment
6. Statistical Analysis: The Bootstrap
7. Conclusion

Introducing Other Groups

- Candidate quality is still: $Q_i \sim \mathcal{N}(0,1)$
- s same-group candidates + o from other group
- $\forall$i = 1,…,s signal: $x_i = Q_i + \sigma_s\, \varepsilon_i$
- $\forall$i=s,…,s+o signal: $x_i = Q_i + \sigma_o\, \varepsilon_i$
- Cultural barriers $\Rightarrow \sigma_s \leq \sigma_o$

Employer Infers Quality from Signal

- Same country: $\forall i = 1,\ldots,s$

  $\hat{Q}_i = E[Q_i | x_i] = \beta_s x_i$ where $\beta_s = \dfrac{1}{1+\sigma_s^2}$

- Other country: $\forall i = s+1,\ldots,s+o$

  $\hat{Q}_i = E[Q_i | x_i] = \beta_o x_i$ where $\beta_o = \dfrac{1}{1+\sigma_o^2}$

- Cultural barriers $\Rightarrow 1 \geq \beta_s \geq \beta_o \geq 0$

Unconditional Distributions

same country	other country
$i = 1, \ldots, s$	$i = s+1, \ldots, s+o$
$x_i \sim \mathcal{N}(0, 1/\beta_s)$	$x_i \sim \mathcal{N}(0, 1/\beta_o)$
$Q_i \sim \mathcal{N}(0, 1)$	$Q_i \sim \mathcal{N}(0, 1)$
$\widehat{Q}_i \sim \mathcal{N}(0, \beta_s)$	$\widehat{Q}_i \sim \mathcal{N}(0, \beta_o)$

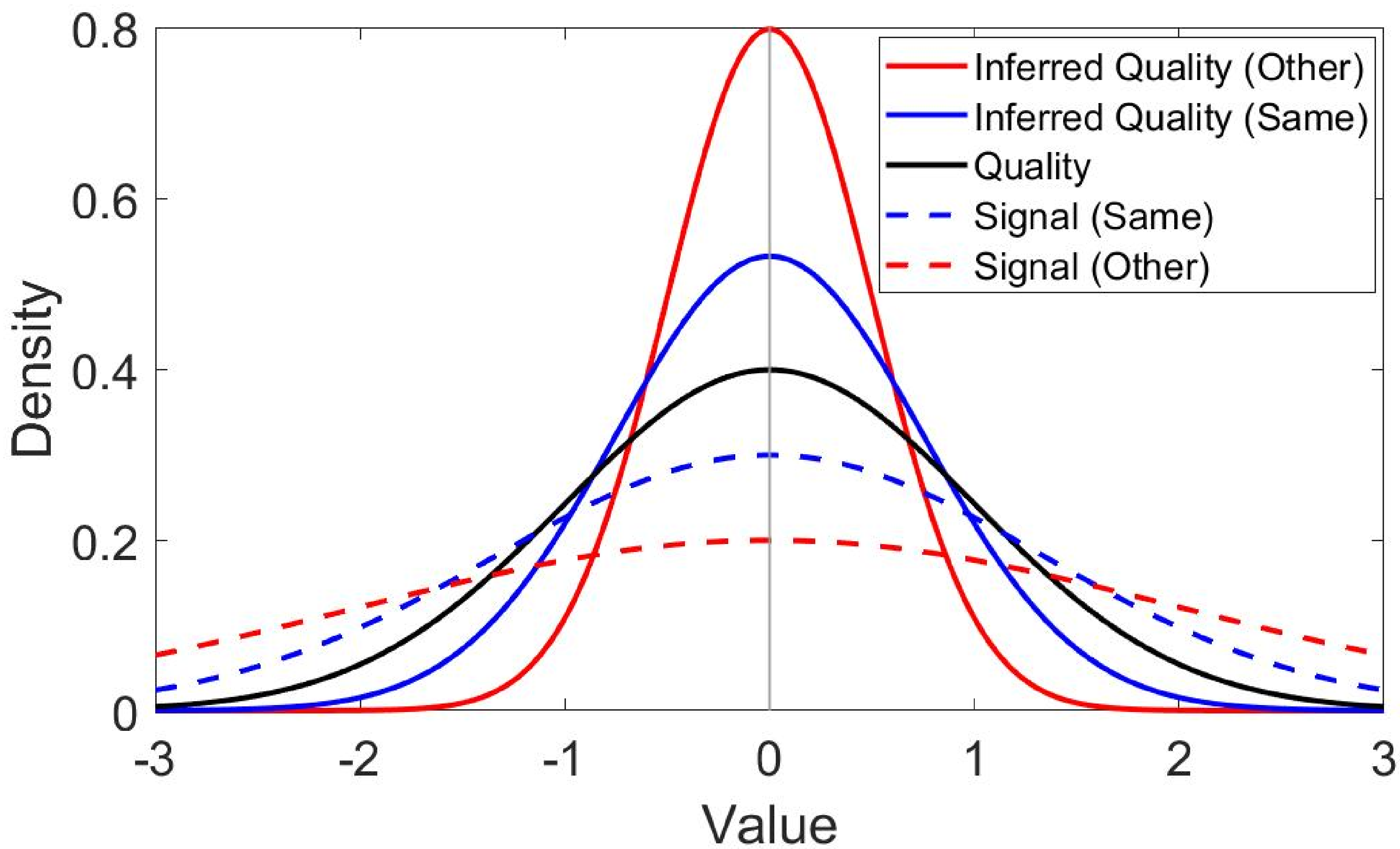
Inferred quality is more dispersed for same-country than for other-country candidates

Key Features of the Model

- No assumption that one country's quality differs from the other
- Employer is <u>not biased</u> towards working with same-country candidates
- When employer ranks candidates from top to bottom, same-country candidates are *also* over-represented at the bottom

$\Rightarrow$ Coming from same country as employer is a **double-edged sword**

Comparing Spread of Distributions



Break (10 minutes)

Plan of Lecture 09

1. Introduction
2. General Overview of Cornell-Welch Model
3. Analysis without other-group candidates
4. Analysis with other-group candidates
5. Evidence from a Natural Experiment
6. Statistical Analysis: The Bootstrap
7. Conclusion

Features of the Experiment

- Different from employer-candidate relations because it is symmetrical
- Students have to choose other students in a class
- Cultural type is defined by **country of origin**
- On average: 44 students from 11 countries
- Form ≈11 groups of ≈4 for high-stakes exercise
- Test is whether students seek out compatriots

Self-Organization

Group Name	A	B	C	D	E	F	G	H
Member 1								
Member 2								
Member 3								
Member 4								

Summary Statistics

	Total	China	India	United States	Singletons	Number of Groups
Class 1	41	19	11	0	6	10
Class 2	33	19	11	0	1	9
Class 3	49	20	11	5	7	13
Class 4	57	11	21	9	9	13
Class 5	52	25	11	4	8	13
Class 6	34	12	6	5	8	10
Average	44.3	17.7	11.8	3.8	6.5	11.3

Any Mistakes?

Group	Last Name	First Name	Country	Gender
ALPHA	CAI	YUE	China	Male
ALPHA	LIN	BRYAN	China	Male
ALPHA	LU	XIAOQING	China	Female
ALPHA	LUO	XICHEN	China	Female
ALPHA	YU	CHARLES	China	Male
BRAVO	FAZLANI	ABDUL MOIZ	Pakistan	Male
BRAVO	KARNANI	BIJOY TARUN	India	Male
BRAVO	SHINKARUK	RUSTEM	Kazakhstan	Male
DELTA	DHAKA	ASHISH	India	Male
DELTA	DONG	SHUYU	China	Female
DELTA	IASHVILI	NIKOLOZ	Georgia	Male
DELTA	SHUKLA	MOHIT	India	Male
DELTA	YU	CHENLU	China	Female

Group	Last Name	First Name	Country	Gender
ECHO	HONG	SEUNGMIN	Korea	Male
ECHO	HUANG	RUNHONG	China	Male
ECHO	PENG	TONGSU	China	Female
ECHO	TAN	JUSTIN	Canada/China	Male
GOLF	DONG	XIYUAN	China	Female
GOLF	HUANG	JIAMING	China	Male
GOLF	KIM	HOGUN	Korea	Male
GOLF	WANG	XIAHAO	China	Male
GOLF	ZHAO	YANXIANG	China	Male
INDIA	BHARATI	SEJAL	India	Female
INDIA	MOHANTY	ELEENA	India	Female
INDIA	ORTIZ CASILLAS	PABLO	Mexico	Male
INDIA	SOLANKI	NUPUR BEN	India	Female

Group	Last Name	First Name	Country	Gender
JULIET	HUANG	YINING	China	Male
JULIET	LI	JIAQI	China	Male
JULIET	MEI	XIANGUI	China	Female
JULIET	ZHU	XINYUE “JULIET”	China	Female
LIMA	JUNG	HYEUK	Korea	Female
LIMA	LUO	JINGWEI	China	Male
LIMA	WASSMANN	MAXIMILIAN	Germany	Male
LIMA	ZHANG	YING	China	Female
MOTHER	DESHPANDE	ARJUN	India	Male
MOTHER	MURDIA	AYUSH	India	Male
MOTHER	SHARMA	ANIKET HARISHBHAI	India	Male
MOTHER	THATISHETTI	RANADHEER	India	Male
MOTHER	VENKATESHWARAN	RITESH	India	Male
QUEBEC	CAO	CHONG	China	Male
QUEBEC	WEI	YUFAN	China	Male

National Affinity Index

- For every class, for every student, how many same-country students are in same group *(not counting self)* ?
- Average over students in class, and 6 classes
- Depends on how many of students from same-country, other-country, and group size
- Empirical result: **1.55**
- Is this high or low??

Plan of Lecture 09

1. Introduction
2. General Overview of Cornell-Welch Model
3. Analysis without other-group candidates
4. Analysis with other-group candidates
5. Evidence from a Natural Experiment
6. Statistical Analysis: The Bootstrap
7. Conclusion

Invented by Brad Efron from Stanford

B Efron

FOLLOW

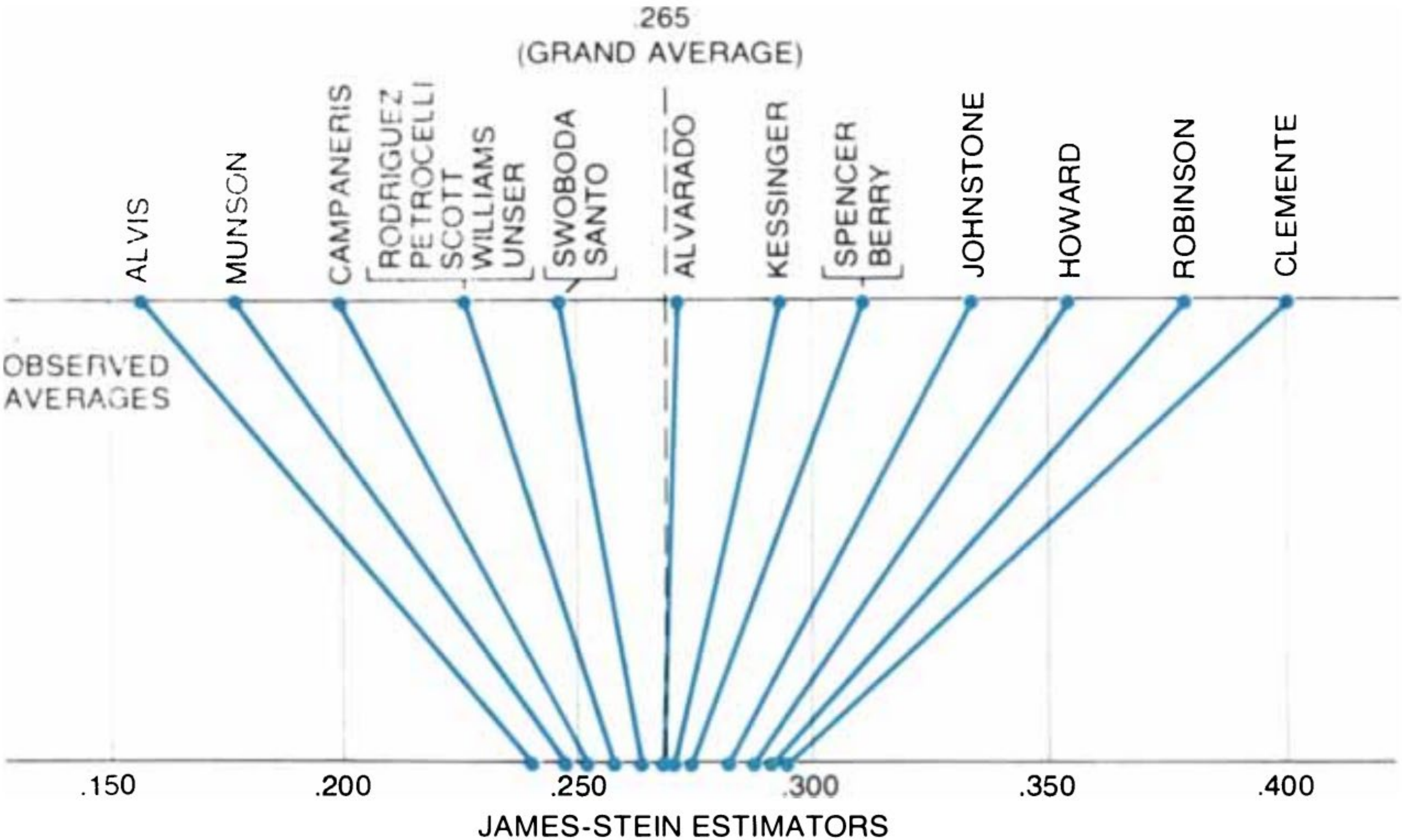
Professor of statistics, Stanford University
Verified email at stat.stanford.edu

statistics biostatistics astrostatistics

TITLE	CITED BY	YEAR
An introduction to the bootstrap B Efron, RJ Tibshirani CRC press	41376	1994
Bootstrap methods: another look at the jackknife B Efron Breakthroughs in statistics, 569-593	18247	1992
The jackknife, the bootstrap, and other resampling plans B Efron Siam	10047	1982

Recipient of the 1990 Wilks Medal

Wrote my Favorite Shrinkage Paper

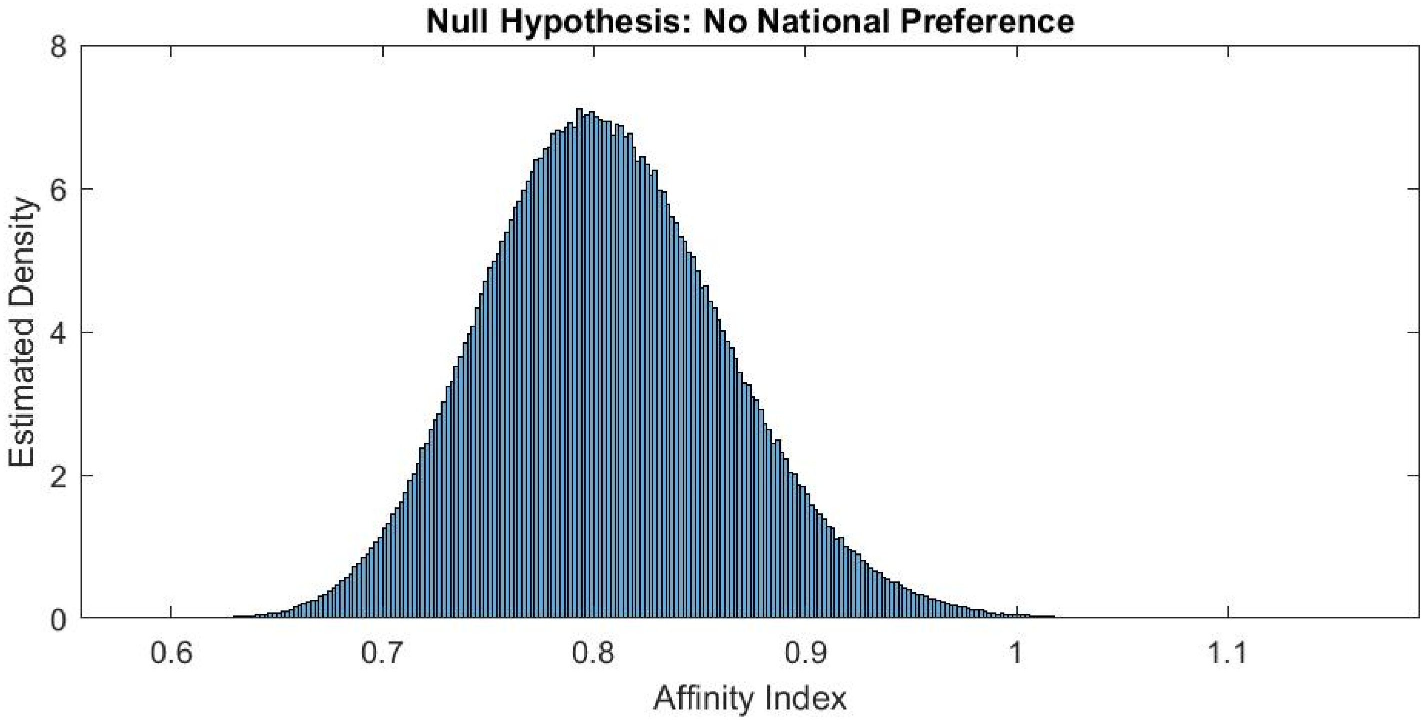


Published in *Scientific American* in 1997

Bootstrap Exercise

- Take all the original data
- Observed affinity index: θ = **1.55**
- B=1,000,000 bootstrap simulations
- Permute group assignments randomly
- Null hypothesis: no country affinity
- Recompute simulated affinity index $\underline{\theta}_1,\dots,\underline{\theta}_B$
- p-value: $$PV = \frac{\#\{\underline{\theta}_b \geq \theta\}+1}{B+1}$$

Bootstrapped Density under the Null



- Mean: 0.81; std dev: 0.06; range: [0.59, 1.16]
- Observed value of 1.55 is 6 std dev > highest
- Reject hypothesis of national indifference: $p<10^{-6}$

Spread Ratio

$$\rho \equiv \sqrt{\frac{\beta_s}{\beta_o}} = \sqrt{\frac{1+\sigma_o^2}{1+\sigma_s^2}} \geq 1$$

- Excess spread of other-country observed signals relative to same-country signals

- Excess spread of same-country <u>inferred quality</u> relative to other-country <u>inferred quality</u>

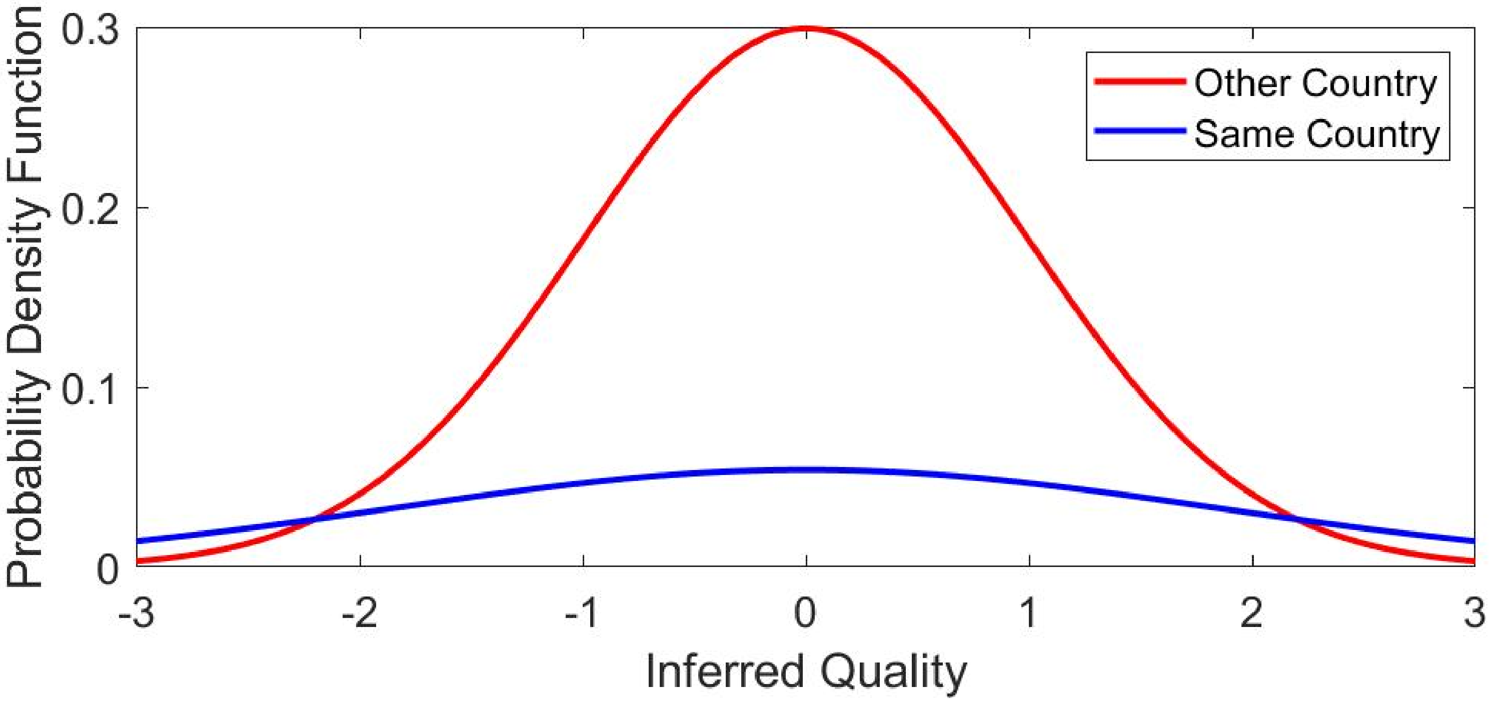
Estimation of the Spread Ratio

- For every class, for every student: s same-country, o other-country, k same group (excluding self: $k \approx 3$)
- Simulate s inferred qualities from $\mathcal{N}(0,\rho^2)$
- Simulate o inferred qualities from $\mathcal{N}(0,1)$
- Find the top k out of $s + o$
- How many out of k come from same-country?
- Average over 10,000 simulations, 6 classes
- Search for ρ that gives country affinity of **1.55**

Market-Clearing Condition

- You are on my short-list, but what if I am not on your short-list?
- Negotiations are hard to model
- Nucleus of 2 or 3: all want a say in choice of 4^{th} groupmate
- Not clear in what direction it affects results
- Sanity check: when $\rho = 1 \Rightarrow$ average affinity index is **0.81** (like before with permutation)

Density of Inferred Quality



- 75% other-country, 25% same-country
- Std Dev: other-country=1, same-country=**1.85**
- Wish-List Choices: 1st=3.0, 2nd=2.3, 3rd=1.9

Robustness Check: Affinity by Class

	Observed	Null:mean	Null:std dev	z-score	p-value
Class 1	1.61	0.93	0.15	4.51	<0.001
Class 2	2.67	1.3	0.19	7.28	<0.001
Class 3	1.47	0.69	0.13	6.09	<0.001
Class 4	1.3	0.67	0.12	5.31	<0.001
Class 5	1.58	0.83	0.12	6.31	<0.001
Class 6	0.71	0.41	0.12	2.36	0.028

2^{nd} Robustness Check: By Country

	Observed	Null:mean	Null:std dev	z-score	p-value
China	2.03	1.22	0.1	8.15	<0.001
India	2.11	0.77	0.11	11.71	<0.001
Other	0.22	0.1	0.05	2.44	0.03

In the Gender Dimension

- Average per class: 33 males, 11 females
- Observed Gender Affinity Index: **1.94**
- On average, your group has **1.94** other students of same gender as yourself
- Average Affinity Index bootstrapped 10,000 times under the null hypothesis of gender indifference: **1.93 (**with std dev 0.06)
- Z-score=0.23, p-value=0.39 $\Rightarrow$ can’t reject null
- Puzzle: students appear willing to spend time with opposite sex

Plan of Lecture 09

1. Introduction
2. General Overview of Cornell-Welch Model
3. Analysis without other-group candidates
4. Analysis with other-group candidates
5. Evidence from a Natural Experiment
6. Statistical Analysis: The Bootstrap
7. Conclusion

Main Lessons

- People are faced with the problem of screening other people in virtually every aspect of life
- This sorting problem involves not only major decisions such as:
  - choosing friends,
  - accepting members in social organizations,
  - and hiring,
- but everyday decisions such as deciding whether the person walking toward you on the sidewalk means well or ill
- And it is an ‘advanced’ or ‘differential’ shrinkage problem

Plan of Lecture 10

- 8:30 to 9:20: class as usual
- 9:20 to 9:30: ‘early’ break
- 9:30 to 10:40: talk by David Don from RCM-X on equity market-neutral strategies via video link with Chicago
- 10:40 to 11:00 (approximately): some personal remarks from Joe Signorelli (RCM-X)
- 11:00 to 11:20: window for evaluating your instructor on the UCLA intranet

RCM-X in Chicago

Assignment: explore their website for more info
https://www.rcmalternatives.com/rcmx/