Advanced Panel Data: Interpreting FE?

Chris Conlon

April 28, 2020

NYU Stern

Causal FE

Recall the FE Assumptions

$$y_{it} = x'_{it}\beta + \eta_i + \varepsilon_{it}$$

- η_i is a fixed effect.
- To estimate everything consistently, we need $E[\varepsilon_{it}|x_{it}, \eta_i] = 0$
- Mostly this is not true. Instead usually treat η_i as a control variable or nuisance parameter.
  - A nuisance parameter is one that we estimate but don't care about interpreting.
  - If we only care about β then η_i is a nuisance parameter.
- With a control or nuisance parameter we only require that $E[\varepsilon_{it}|\eta_i] = E[\varepsilon_{it}|x_{it}, \eta_i]$ conditional mean independence.

Recall the FE Assumptions

$$y_{it} = x'_{it}\beta + \eta_i + \varepsilon_{it}$$

- η_i is a fixed effect.
- To estimate everything consistently, we need $E[\varepsilon_{it}|x_{it}, \eta_i] = 0$
- Mostly this is not true. Instead usually treat η_i as a control variable or nuisance parameter.
  - A nuisance parameter is one that we estimate but don't care about interpreting.
  - If we only care about β then η_i is a nuisance parameter.
- With a control or nuisance parameter we only require that $E[\varepsilon_{it}|\eta_i] = E[\varepsilon_{it}|x_{it}, \eta_i]$ conditional mean independence.
- Once we condition on η_i it is as if ε_{it} and x_{it} are uncorrelated.

Causal FE

- We can get away with conditional mean independence if we don't care about η_i.
- But suppose that we care about $\widehat{\eta}_i$?
  - Teacher FE
  - Physician/Hospital FE
  - Location/County FE
  - Suppose we take someone from the $10th$ percentile and move them to the $90th$ percentile

Causal FE

- Now we have to really believe $E[\varepsilon_{it}|x_{it}, \eta_i] = 0$
- We should worry about the conventional omitted variable bias problem.
- Suppose there exists a variable w_{it} so that:

$$y_{it} = x'_{it}\beta + w'_{it}\gamma + \eta_i + \varepsilon_{it}$$

- Recall the conditions for OVB
  - w_{it} is correlated with x_{it}
  - w_{it} is a determinant of y_{it}
- New one: w_{it} is correlated with η_i
  - This is easy to satisfy!
  - w_{it} needs to be uncorrelated with anything about the individual i.

Example: Test Scores

- Students s, Teachers t
- Want to measure effect of Teachers on Test Scores

$$TestScore_{st} = \beta x_s + \gamma w_t + \eta_i + \varepsilon_{st}$$

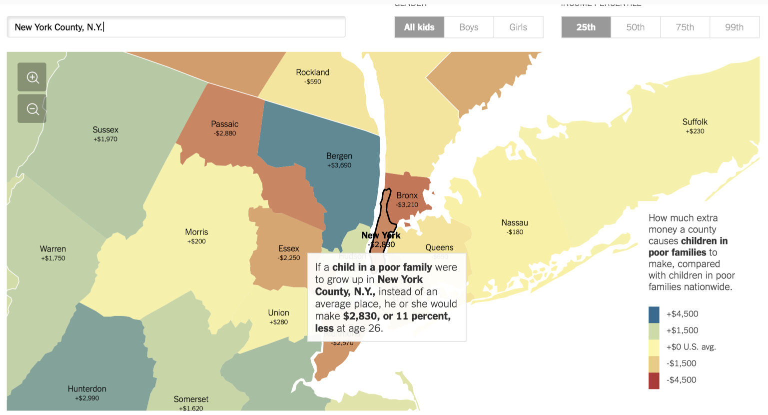
- We observe some features of students but not all of them (parent's education, household income, language spoken at home).
- We also observe some school specific variables w_t but not all of them (district spending per pupil, % free lunch, etc.).
- But we don't observe other things (jackhammering outside the classroom, which students have disruptive home lives,etc.).
    - If the mean of those things varies across teachers $\rightarrow$ we are screwed!
    - Can't get an accurate estimate of η_i.

Example: Test Scores

We need a better design:

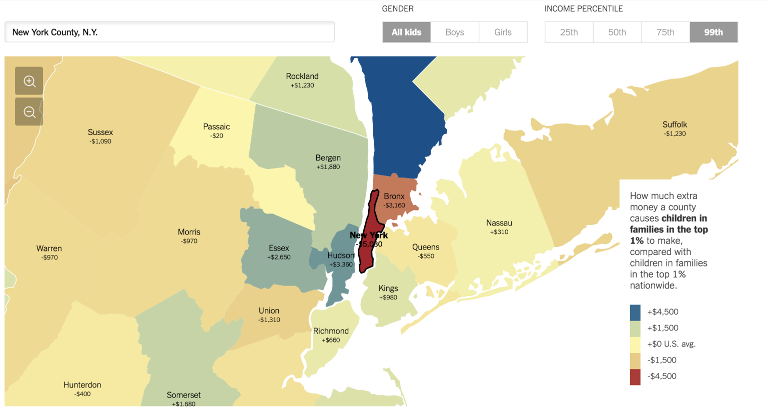
- We probably need random assignment of students to teachers.
- Ideally we would be able to control for student and school unobservables.
- Might want to see many students match with many teachers.

Best Counties? (Chetty Hendren 2016 AER)



Manhattan is **very bad** for income mobility for **children in poor families.** It is better than **only about 7 percent of counties.**

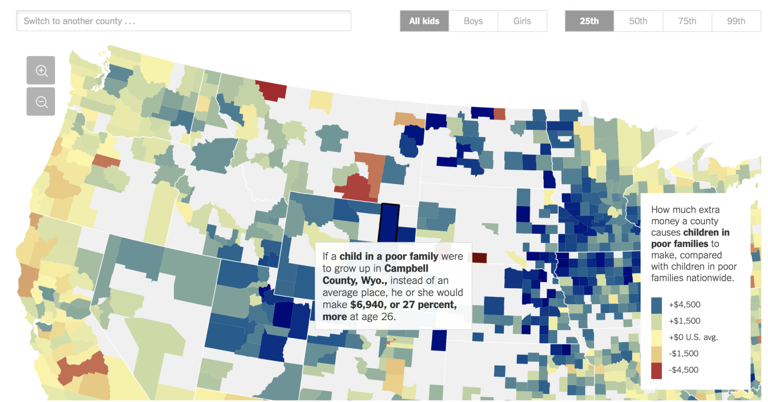
Best Counties?



Manhattan is **extremely bad** for **children in families in the top 1%.**
It is among the worst counties in the U.S.

Best Counties?

50,000 residents (and very few people move here).



Campbell County is **extremely good** for income mobility for **children in poor families.** It is among the best counties in the U.S.

Time to move to Campbell County, WY?

Less than 10% of Campbell County residents graduated college (among worst in US):

Quick Stats (2020)

- Grades: **9-12**
- Students: **1,103 students**
- Student:Teacher Ratio: **16:1**
- Minority Enrollment: **18%**
- Graduation Rate: **86%** (Btm 50% in WY)
- Overall School Rank: **Bottom 50%**
- Math Proficiency: **35%** (Btm 50%)
- Reading Proficiency : **30%** (Btm 50%)
- Diversity Score: **0.31** (Btm 50%)

What's going on?

Is Campbell county the land of opportunity?

- Or is it a small place that nobody moves to with very imprecisely measured FE γ_i?
- Helped that people found coal in their backyards (literally)
- If you moved there today would you still expect to do well?

This problem is endemic:

- Our "best" and "worst" teachers tend to be those who teach the fewest students.
- Presumably FE are estimated with precision that is not equal.

Healthcare Exceptionalism: Static Reallocation

Is quality (or productivity) correlated with market share?

$$\ln(N_h) = \beta_0^s + \beta_1^s q_h + \gamma_M^S + \varepsilon_h^s$$

- N_h measures market size for hospital h
- γ_M^s are market FE
- q_h is measure of hospital quality
- Goal: Is $\beta_1^s > 0$ or not. $\beta_1^s < 0$ is usually only Soviet countries or 1970's steel.
- $\beta_1^s > 0$ means allocation towards productive firms (or just returns to scale?)

Healthcare Exceptionalism: Dynamic Reallocation

$$\Delta_h = \beta_0^d + \beta_1^d q_h + \gamma_M^d + \varepsilon_h^d$$
$$\Delta_h = \frac{N_{h,2010} - N_{h,2008}}{\frac{1}{2}\left(N_{h,2010} + N_{h,2008}\right)}$$

- $\beta_1^d > 0$ means growth towards productive firms (not just returns to scale)
- Same idea but now we capture dynamics.
- Patients may still be attracted to unobservables correlated with quality.

TABLE 1—STATIC AND DYNAMIC ALLOCATION METRICS ACROSS CONDITIONS

Condition:	AMI (1)	Heart failure (2)	Pneumonia (3)	Hip/knee (4)
Panel A. Composition of all Medicare discharges in 2008				
Number of patients in 2008	263,485	545,363	475,756	350,536
Share through emergency dept.	0.71	0.76	0.76	0.02
Share of all Medicare discharges	0.03	0.06	0.05	0.04
Share of Medicare hospital spending	0.04	0.05	0.04	0.05
Number of hospitals in 2008	4,257	4,547	4,607	3,297
Panel B. Static allocation: patients in 2008				
Patients (index events)	190,189	308,122	354,319	267,557
Average number of patients per hospital	65.8	76.6	81.9	101.7
SD of patients per hospital	67.6	78.2	70.8	118.0
Hospitals	2,890	4,023	4,325	2,632
Average number of hospitals per market	9.4	13.1	14.1	8.6
Panel C. Dynamic allocation: growth in patients from 2008 to 2010				
Average growth rate across hospitals	−0.17	−0.10	−0.13	−0.03
SD across hospitals	0.42	0.38	0.36	0.46
Hospitals	2,890	4,023	4,325	2,632

Notes: Panel A is calculated on a 100 percent sample of age 65 and older fee-for-service Medicare patients in 2008 and counts all patients with the condition, not just the index events that are the subject of the remainder of this study and panels B and C. The sample in panels B and C is all hospitals that had at least 1 index admission in 2008 for the condition shown in the column heading and had a valid risk-adjusted survival rate for that condition (risk-adjusted readmission for hip/knee replacement). There are 306 hospital markets, called Hospital Referral Regions (HRRs). Growth is calculated based on the formula in equation (3) that restricts values to between −2 and 2.

Healthcare Exceptionalism

TABLE 2—SUMMARY STATISTICS ON QUALITY METRICS ACROSS CONDITIONS

Condition:	AMI (1)	Heart failure (2)	Pneumonia (3)	Hip/knee (4)
Panel A. Risk-adjusted survival rates (30 days): patients in 2006–2008				
Average 30-day survival rate	0.82	0.89	0.88	
SD of risk-adjusted measure	(0.03)	(0.02)	(0.02)	
Hospitals in risk-adjusted measure	2,890	4,023	4,325	
Panel B. Risk-adjusted readmission rates (30 days): patients in 2006–2008				
Average 30-day readmission rate	0.21	0.21	0.16	0.06
SD of risk-adjusted measure	(0.03)	(0.02)	(0.02)	(0.02)
Hospitals in risk-adjusted measure	2,322	3,904	4,264	2,632
Panel C. Processes of care: shares of patients receiving appropriate treatments in 2006–2008				
Average score	0.93	0.83	0.88	
SD	(0.05)	(0.14)	(0.07)	
Hospitals	2,398	3,666	3,920	
Average number of processes reported	4.40	3.30	6.22	
Panel D. Patient survey: survey covers all patients in 2008 (not limited to particular condition)				
Average overall rating (1–3, higher is better)	2.53	2.53	2.53	2.53
SD	(0.14)	(0.14)	(0.14)	(0.14)
Hospitals	3,498	3,598	3,610	3,061

Notes: Sample restrictions are specific to the condition and quality metric; see text for more details of the metric definitions and sample restrictions. Summary statistics are reported across hospitals. In panels A and B, the standard deviations are of the risk-adjusted measures and are empirical-Bayes-adjusted to account for measurement error (see online Appendix Section C.3.1). In panel D, the number of hospitals differs across conditions even though the patient survey metric is not condition-specific because we calculate the ratings on the subset of hospitals that reported at least one patient with the condition in 2008.

TABLE 3—CORRELATION OF QUALITY METRICS WITHIN CONDITION

	AMI				HF			
Metric	Risk-adj survival (1)	Risk-adj readm (2)	Process of care Z (3)	Patient survey Z (4)	Risk-adj survival (5)	Risk-adj readm (6)	Process of care Z (7)	Patient survey Z (8)
Risk-adjusted survival	1.00 [2,890]				1.00 [4,023]			
Risk-adjusted readmission	0.03 [2,322]	1.00 [2,322]			0.35 [3,904]	1.00 [3,904]		
Process of care *Z*-score	0.24 [2,346]	−0.25 [2,214]	1.00 [2,398]		0.17 [3,607]	−0.15 [3,578]	1.00 [3,666]	
Patient survey *Z*-score	−0.06 [2,799]	−0.26 [2,293]	0.18 [2,370]	1.00 [3,498]	−0.18 [3,447]	−0.36 [3,398]	0.01 [3,392]	1.00 [3,598]
	Pneumonia				**Hip/knee replacement**			
Metric	Risk-adj survival	Risk-adj readm	Process of care Z	Patient survey Z	Risk-adj survival	Risk-adj readm	Process of care Z	Patient survey Z
Risk-adjusted survival	1.00 [4,325]							
Risk-adjusted readmission	0.08 [4,264]	1.00 [4,264]				1.00 [2,632]		
Process of care *Z*-score	0.08 [3,871]	−0.18 [3,847]	1.00 [3,920]					
Patient survey *Z*-score	−0.03 [3,527]	−0.36 [3,503]	0.18 [3,512]	1.00 [3,610]		−0.23 [2,542]		1.00 [3,061]

Notes: Hospitals used to calculate correlation in brackets. All quality metrics are condition-specific except the patient survey, which is only available as an all-patient average. Correlations involving risk-adjusted survival and

Healthcare Exceptionalism

Table 4—Allocation across Conditions

	Static allocation				Dynamic allocation			
Measure/condition	AMI (1)	HF (2)	Pneu (3)	Hip/knee (4)	AMI (5)	HF (6)	Pneu (7)	Hip/knee (8)
Risk-adjusted survival								
Coef. on survival rate	17.496	15.360	5.140		1.533	0.774	1.220	
	(0.995)	(1.320)	(0.777)		(0.379)	(0.501)	(0.354)	
Hospitals	2,890	4,023	4,325		2,890	4,023	4,325	
Risk-adjusted readmission								
Coef. on readmission rate	−9.162	−10.346	0.499	−21.037	−1.428	−2.300	−1.138	−1.112
	(1.621)	(1.782)	(1.575)	(2.027)	(0.611)	(0.651)	(0.679)	(0.836)
Hospitals	2,322	3,904	4,264	2,632	2,322	3,904	4,264	2,632
Process of care Z-score								
Coef. on process Z-score	0.319	0.332	0.211		0.048	0.043	0.026	
	(0.026)	(0.016)	(0.015)		(0.010)	(0.009)	(0.009)	
Hospitals	2,398	3,666	3,920		2,398	3,666	3,920	
Patient survey Z-score								
Coef. on survey Z-score	−0.321	−0.252	−0.210	0.057	−0.065	−0.003	0.007	0.037
	(0.052)	(0.038)	(0.030)	(0.051)	(0.015)	(0.011)	(0.011)	(0.022)
Hospitals	3,498	3,598	3,610	3,061	3,498	3,598	3,610	3,061

Notes: The static allocation results are estimated using equation (1), a hospital-level regression of log-patients in 2008 on market fixed effects and the quality measure named in the row. The dynamic allocation results are estimated using equation (2), which is an identical regression except for the dependent variable, which is now growth in patients from 2008 to 2010. Growth is defined as in equation (3). Standard errors are bootstrapped with 300 replications and are clustered at the market level. Risk-adjusted survival and readmission are reported in percentage points (e.g., a value of 0.1 is 10 percentage points); process of care and patient survey metrics are reported in standard deviation units (e.g., a value of 1 is one standard deviation).

Healthcare Exceptionalism: Production Function

Hospital Production Function:

$$y_p^s = a_h + \sum_k \lambda_k r_{pk} + \mu x_p + \xi_p$$

- a_h is hospital productivity (a FE) and variable of interest
- y_p is a patient outcome (survival-days, etc.)
- x_p are (log) hospital inputs
- r_{pk} are patient risk factors.
- This has interpretation as a production function. Why?

TABLE 7—ALLOCATION OF AMI WITH RESPECT TO AMI PRODUCTIVITY AND ITS COMPONENTS

Measure	Static allocation for AMI				Dynamic allocation for AMI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Productivity (fed $)	17.637				1.491			
	(1.118)				(0.420)			
Productivity (resources)		17.540				1.471		
		(1.013)				(0.386)		
Risk-adjusted ln(fed $)			1.447				0.246	
			(0.169)				(0.064)	
Risk-adjusted ln(resources)				0.620				0.468
				(0.406)				(0.162)
Risk-adjusted survival			17.940	19.789			1.479	1.559
			(1.192)	(1.297)			(0.446)	(0.441)
Hospitals	2,890	2,890	2,890	2,890	2,890	2,890	2,890	2,890

Notes: This table extends the analysis of Table 4 but is limited to static and dynamic allocation for AMI. It shows how allocation is related to AMI productivity or its two components (risk-adjusted survival and risk-adjusted log inputs). Productivity is defined as risk- and inputs-adjusted survival; see Section IIIC and equation (9). We consider two input measures, "federal expenditures" and "resources," also defined in the text. Standard errors are bootstrapped with 300 replications and are clustered at the market level. The standard deviation of productivity is 0.03 (Fed $ or Resources), of risk-adjusted log-inputs is 0.22 (Fed $) and 0.07 (Resources), and of risk-adjusted survival is 0.04—this number differs from that of Table 2 because it comes from estimating the joint distribution of survival and inputs, not survival alone.

Healthcare Exceptionalism: EB Adjustment

Table A2 - Sensitivity of Allocation Results to Empirical Bayes Adjustment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Static Allocation				Dynamic Allocation			
Condition	AMI	HF	Pneu	Hip/Knee	AMI	HF	Pneu	Hip/Knee
Panel A - Risk-Adjusted Survival								
Baseline (EB-Adjusted)	17.496	15.360	5.140		1.533	0.774	1.220	
	(0.995)	(1.320)	(0.777)		(0.379)	(0.501)	(0.354)	
Raw (No EB Adjustment)	6.833	3.761	1.957		0.645	0.084	0.340	
	(0.342)	(0.425)	(0.403)		(0.175)	(0.199)	(0.175)	
Hospitals	2,890	4,023	4,325		2,890	4,023	4,325	
Raw SD / Corrected SD	1.597	1.788	1.547		1.597	1.788	1.547	
Panel B - Risk-Adjusted Readmission								
Baseline (EB-Adjusted)	-9.162	-10.346	0.499	-21.037	-1.428	-2.300	-1.138	-1.112
	(1.621)	(1.782)	(1.575)	(2.027)	(0.611)	(0.651)	(0.679)	(0.836)
Raw (No EB Adjustment)	-1.699	-1.043	0.755	-6.492	-0.307	-0.217	-0.189	-0.431
	(0.395)	(0.346)	(0.427)	(0.727)	(0.197)	(0.162)	(0.195)	(0.382)
Hospitals	2,322	3,904	4,264	2,632	2,322	3,904	4,264	2,632
Raw SD / Corrected SD	1.870	2.132	1.864	1.794	1.870	2.132	1.864	1.794

This table shows the sensitivity of the allocation results of Table 4 to the empirical Bayes adjustment procedure. In each panel, we first repeat the baseline allocation results in which the quality metric is empirical-Bayes-adjusted. We then show the same allocation models using the raw quality metric without empirical Bayes adjustment. Lastly, we show the ratio of the raw standard deviation of the quality measure to its standard deviation after correcting for measurement error (see Appendix Section C.3.1). Standard errors are bootstrapped with 300 replications and are clustered at the market level.

Empirical Bayes

What is Emprical Bayes?

- Priors can be an important modeling choice
- But what makes a good prior?
  - Sufficiently diffuse
  - As non-informative as possible
  - Don't tip the scales
  - Don't rule out the truth
- Idea: can we use the data itself to construct a prior?
  - If everything is a function of data, are we back in frequentist paradigm?
  - Can we get benefits of Bayes estimation without unpalatable assumptions?

A (famous) Baseball Example

Suppose we want to estimate batting averages (AVG) for some baseball players

- $AVG = \frac{\#\text{hits}}{\#AtBats}$
- Use data on the first $n = 45$ at bats and hits x_i for the 1970 season.
- Predict the batting average μ_i for the end of the season ($n = 400 - 500$ at bats).
- Obvious estimate is batting average after 45 at bats: $\widehat{\mu}_i^{MLE} = x_i/45$.
- Is there a better estimate?

A Baseball Example

Table 1.1: Batting averages $z_i = \hat{\mu}_i^{(\text{MLE})}$ for 18 major league players early in the 1970 season; μ_i values are averages over the remainder of the season. The James–Stein estimates $\hat{\mu}_i^{(\text{JS})}$ (1.35) based on the z_i values provide much more accurate overall predictions for the μ_i values. (By coincidence, $\hat{\mu}_i$ and μ_i both average 0.265; the average of $\hat{\mu}_i^{(\text{JS})}$ must equal that of $\hat{\mu}_i^{(\text{MLE})}$.)

Name	hits/AB	$\hat{\mu}_i^{(\text{MLE})}$	μ_i	$\hat{\mu}_i^{(\text{JS})}$
Clemente	18/45	.400	**.346**	.294
F Robinson	17/45	.378	**.298**	.289
F Howard	16/45	.356	**.276**	.285
Johnstone	15/45	.333	**.222**	.280
Berry	14/45	.311	**.273**	.275
Spencer	14/45	.311	**.270**	.275
Kessinger	13/45	.289	**.263**	.270
L Alvarado	12/45	.267	**.210**	.266
Santo	11/45	.244	**.269**	.261
Swoboda	11/45	.244	**.230**	.261
Unser	10/45	.222	**.264**	.256
Williams	10/45	.222	**.256**	.256
Scott	10/45	.222	**.303**	.256
Petrocelli	10/45	.222	**.264**	.256
E Rodriguez	10/45	.222	**.226**	.256
Campaneris	9/45	.200	**.286**	.252
Munson	8/45	.178	**.316**	.247
Alvis	7/45	.156	**.200**	.242
Grand Average		.265	**.265**	.265

A (famous) Baseball Example

Probably we can do better than the MLE here:

- Thurman Munson wins Rookie of the Year and ends up batting $\mu_i = .316$. If he batted .178 all year, his career would not have lasted long.
- Clemente's $.400$ seems unlikely to hold up. Last player to hit $> .400$ was Ted Williams $.406$ in 1941.
- But how?

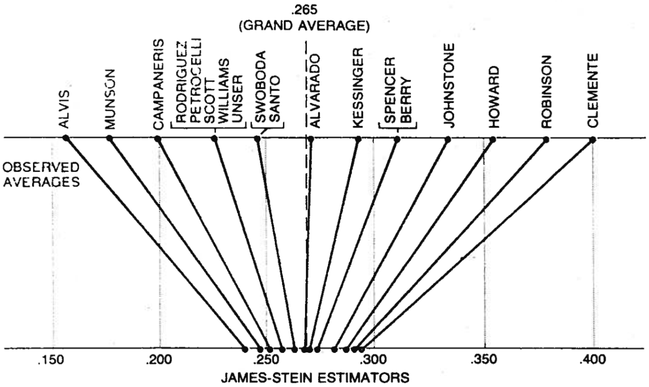
Bayesian Shrinkage

Idea is to take an average between the observed average y_i and the overall mean $\overline{y}$:

$$\widehat{\mu}_i^{JS} = (1-\lambda)\cdot\overline{y} + \lambda\cdot y_i, \quad \lambda = 1 - \frac{(k-3)\sigma^2}{\sum_i (y_i - \overline{y})^2}$$

- This has the effect of shrinking y_i towards the prior mean $\overline{y}$.
- In this case the prior mean is just $\overline{y}$ the grand-mean of all players
- How can information about unrelated players inform us about μ_i?
- Also consider proportion of foreign cars in Chicago as an additional y_i, can this help too?
- The shrinkage factor λ depends on sample size and variance, but how is it chosen?

A Baseball Example



JAMES-STEIN ESTIMATORS for the 18 baseball players were calculated by "shrinking" the individual batting averages toward the overall "average of the averages." In this case the grand average is .265 and each of the averages is shrunk about 80 percent of the distance to this value. Thus the theorem on which Stein's method is based asserts that the true batting abilities are more tightly clustered than the preliminary batting averages would seem to suggest they are.

Thanks!