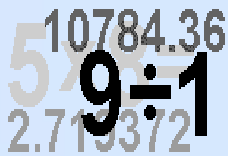
Gauging the Impact of Academic Support Programs: A Quasi Experimental Design Using Propensity Scoring

*****

http://www.unr.edu/ia/research



Serge Herzog, PhD
Director, Institutional Analysis
Consultant, CRDA StatLab
University of Nevada
Reno, NV 89557, serge@unr.edu

CAIR Conference 2014
Napa, CA, 19-22 November

What can we learn from education research?

- [Q]uestions of causality have been at the forefront of educational debates, in part because of the dissatisfaction with the quality of education research and recent federal initiatives designed to promote the accumulation of scientific evidence in education that rely on randomized controlled trials (RCTs). A common concern revolves around the design and methods used in education research, which many claim have resulted in fragmented and often unreliable findings.
  - Quoted in J.C. Smart (ed.) *Higher Education: Handbook of Theory and Research, vol. 24*, p. 47.

2

Purpose of Study

- Estimate the effect of participation in living-and-learning communities (LLC) and use of math support center on:
  - First-semester academic success (GPA, credits earned)
  - Enrollment persistence into second semester
  - Second-semester academic success
  - Enrollment persistence into second year
- Control for selection bias via counter-factual analytical framework
  - Compare parametric vs. non/semi-parametric estimates

4

Source and Data

- New full-time freshmen at the U. of Nevada, Reno
- Data Elements (Covariates):
  - Student demographics (age, gender, race, residency, parent education)
  - Student degree goal and employment plans when in college
  - High school academic experience (GPA, AP, admission scores/test date; class rank)
  - Fall semester academic experience (English, Math, dropped credits, distance ed., major, GPA, earned credits, F/I/W grades)
  - Financial aid profile: unmet need, aid type offered/received
  - Campus services used: math/diversity centers, campus jobs
  - LLC reference groups: off campus, non-LLC on campus
  - Math support ref group: did not use math support center

5

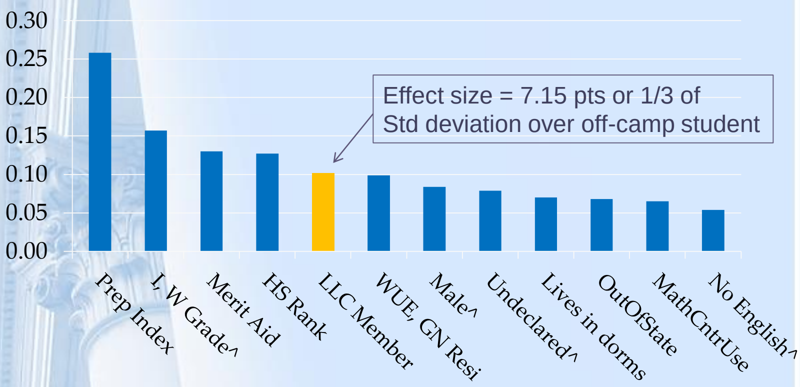
Data Sample and Statistics

- Fall semester cohorts 2011 and 2012
- Excluding:
  - Students with less than 12 attempted credits (PT)
  - Students with complete credit withdrawal
  - Statistical outliers (using Cook's, Mahalanobis')
- Effective sample: 4,871 (math), 3,138 (LLC) students
- Computed variables
  - Precollege preparation index (GPA-test score composite)
  - Delayed college entry: months to UNR matriculation
  - Academic momentum: index (GPA, credits earned comp.)
  - Imputation (mean, predicted value) of missing values
- Analysis:
  - Linear/logistic regression, weighted-sample analysis

Predicting First-Semester Academic Momentum

Baseline Model

Significant Predictors*



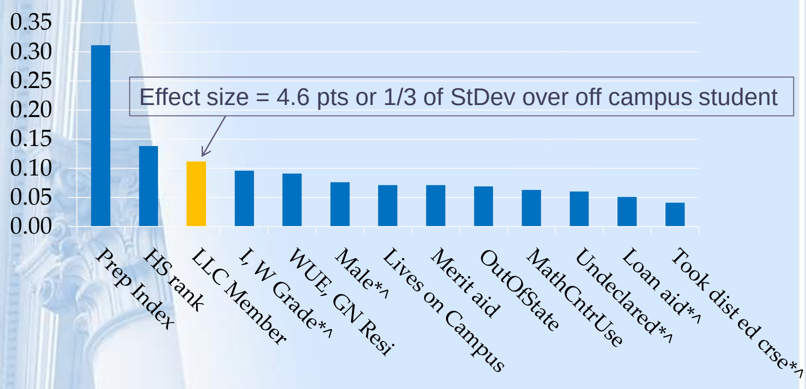
Ranked by Beta weight (standardized coefficient)

*Alpha <= 0.001; ^Negative; Adj R-square=.27; VIF highest = 2.08, all others < 2

Predicting Second-Semester Academic Momentum

Baseline Model

Significant Predictors*



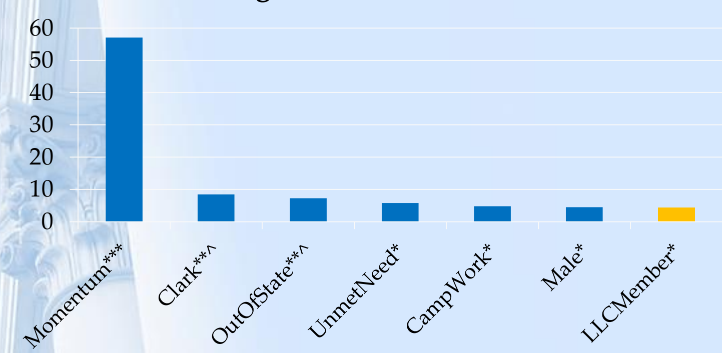
Ranked by Beta weight (standardized coefficient) 13

*Alpha <= 0.001; ^Negative; Adj R-square=.27; VIF highest = 2.09, all others < 2

Predicting Spring Enrollment

At End of Fall Semester

Significant Predictors*

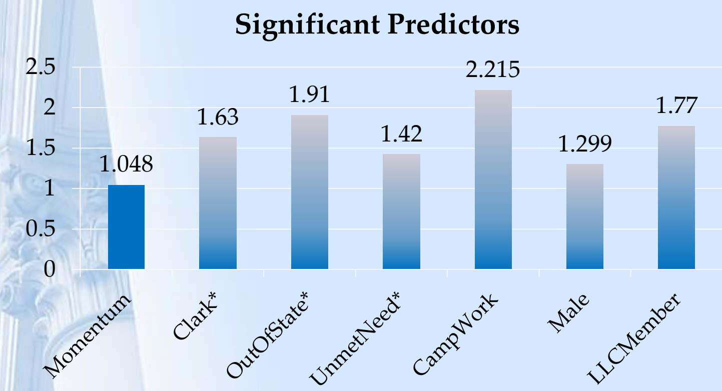


Ranked by Wald Significance Level 15

Alpha ***<=0.001, **0.01, *0.5; ^Negative; Nagelkerke R-square =.32; HL= 0.94

Predicting Spring Enrollment

At End of Fall Semester



Odds Ratio per unit change in IV

16

*Negative; Nagelkerke R-square =.32; HL= 0.94

Findings

- Impact on academic momentum (GPA, credits earned)
  - On average, LLC students are ~ *10 percentile* points *higher* in GPA and credits earned for both fall and spring semesters compared to off-campus students, controlling for all other covariates
  - On average, LLC students are ~ *5 percentile* points *higher* in GPA and credits earned for both fall and spring semesters compared to non-LLC on-campus students, controlling for all other covariates
  - Thus, compared to off-campus students, students living on campus are more likely to earn a higher GPA and more credits in their first year, an advantage that is magnified with participation in a living-and-learning community
- Impact on enrollment persistence
  - On average, LLC students are *9.5%* more likely to persist than off-campus students [((base-p / 1 – base-p)*LLC-OR = pp; ((pp / (1+pp)) = 0.95]
  - On average, non-LLC on-campus students are 9.4% more likely to persist than off-campus students [((.922 / 1-.922)*1.346 = 15.9; ((15.9 / (16.9)) = 0.94], however that result did not meet statistical significance (alpha <=.05)
  - The LLC participation benefit accrues net off all other covariates!!!

17

Caveats

- Findings estimate average effect using parameter data from all students, both 'treated' and 'untreated' (e.g., LLC and non-LLC students)
- Linearity assumption in regression fails if parameters in model (Xs) are highly nonlinear with outcome, distribution of Xs differs between groups of interest (e.g. LLC students vs others)
- Parameter models typically fail to answer to counterfactual H: The outcome for the 'treated' had they not received the treatment
- Probability of student selections/choices often correlate with outcomes of interest (selection/endogeneity bias)

Causal Inference Estimation

- Randomized Control Trials (RCT)
  - Costly, ethical issues, operational difficulties, disruption of natural (campus) setting
  - Replication not possible with RCT
- Nonexperimental designs in observational studies
  - Regression discontinuity
  - Instrumental variable (IV) techniques
  - Econometric models to adjust for selection bias
  - Propensity score methods
    - Inverse probability of treatment weighting (IPTW)
    - Subclass stratification
    - Propensity score matching (PSM)
    - Regression covariate adjustment (linear/nonlinear)

Causal Inference Estimation

- Propensity Scoring: *Rubin Causal Model*
  - Two potential outcomes: $Y_i(1)\ and\ Y_i(0)\ and\ let\ Z\ denote\ LCC\ (1)\ or\ No\ LLC\ (0)$
  - But only one is observed: $Y_i(Y_i = Z_i(1) + (1 - Z_i)Y_i(0)$
  - Thus, $Y_i(1) - Y_i(0) = effect\ of\ LLC\ and\ E[Y_i(1) - Y_i(0) = average\ LLC\ effect\ (ATE)$
  - Average LLC effect $(ATT)$for $Y_i = (1)$ is defined as $E[Y(1) - \mathrm{Y}(0)|\mathrm{Z} = 1]$ where $\Pr(Z_i = 1|X_i)$ where $X_i = pre\text{-}Z\ covariates$

- Propensity score (PS) accounts for covariates (characteristics) that predict treatment selection (participation in LLC or using math center)
- PS captures the conditional probability of treatment selection given observed (measurable) covariates
- Matching treated (LLC) with non-treated (no LLC) students on their PS enables 'counterfactual' analysis, i.e. the expected outcome values with and without LLC participation for those who *actually* participated!

Causal Inference Estimation

- The PS is estimated via logit, probit, log-linear link, Mahalanobis, neural net, classification trees etc.
- PS estimation included 23-30 baseline covariates measuring:
  - Advanced standing, undeclared, age, gender, ethnicity, residency, delayed entry, test date, Pell, unmet need, loans, merit aid, acad preparation, HS rank, first choice, edu goal, plan full-time/no work, mother/father education *(plus interaction terms for some models)*
  - Covariate selection: Must influence both LLC/math participation and outcome(s), and are fixed or measured prior to LLC/math experience
- Tested samples compare LLC with non-LLC on-campus students, and those using math center versus those who did not
- Reliable PS analysis requires
  - Stable Unit Treatment Value Assumption (SUTVA)
  - Strongly Ignorable Treatment Assignment (unconfoundedness)
  - Common support for PS matching to estimate difference in outcome
  - Balanced PS estimation: bias (support) versus variance
- Analysis done in *R* with *MatchIt, Matching, GenMatch* packages

Causal Inference Estimation

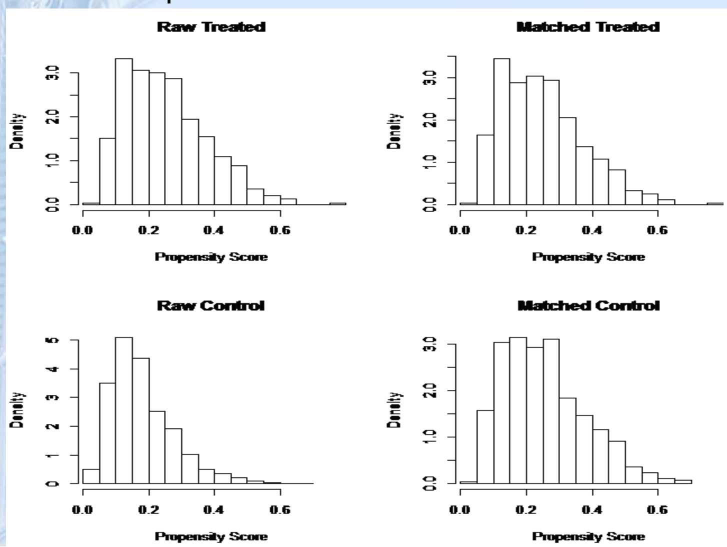
- IPTW with focus on (ATT) reweights ps of non-treated students with $w\big(Y_i(0)\big) = \frac{P_i(X)}{1-Pi(X)}$
- Stratification matches treated with non-treated on the propensity score within 6 subclasses (using support range of treated)
- PSM uses 7 algorithms with LLC/math-defined support

Computation Intensity

  - NN1:1 and K:1 nearest neighbor matching, with/no replacement, random, within 0.2 stddev of ps (caliper), equal weights of controls
  - Kernel1:1 seeks smallest ps distance (difference), no replacement
  - Kernel3:1 seeks smallest ps distance (difference), with up to 3 control (non-LLC/math) cases, weights based on ps distance
  - Full subclass weighted seeks minimal ps distance within max classes
  - Genetic seeks set of weights for each covariate to optimize balance, with replacement to result in 1:1 match, varying optimization iteration
- Regression adjustment uses matched data in linear/log models with LLC/math status and post-selection covariates

22

LLC Prop Score Balance Check: Kernel3:1



23

LLC Balance Check

Algorithm: Kernel3:1		Prop Score/Covariate Mean					
		After Match			**Before Match**		
	% Improve	*Treated*	*Control*	*Mean Diff*	*Treated*	*Control*	*Mean Diff*
distance	99.8	0.2517	0.2516	0.0002	0.2517	0.1795	0.0723
UNRFirst	98.7	0.7694	0.7705	-0.0011	0.7694	0.8534	-0.0841
PrepIndexImp	97.6	80.818	80.6819	0.1361	80.818	75.1213	5.6967
MoEd4yrColl	91.8	0.4959	0.5016	-0.0058	0.4959	0.4259	0.07
HSRankPctImp	89.8	77.646	77.232	0.414	77.646	73.5712	4.0748
EdGoalGrad	89.0	0.6524	0.6474	0.0049	0.6524	0.6073	0.0451
Loans	86.7	0.4465	0.4418	0.0047	0.4465	0.4816	-0.0352
PlanWorkFT	77.7	0.0313	0.0269	0.0044	0.0313	0.051	-0.0197
AdvStanding	70.8	0.0428	0.0368	0.006	0.0428	0.0221	0.0207
FathEd4yrColl	70.3	0.4959	0.483	0.0129	0.4959	0.4524	0.0435
WUEGN	67.6	0.3427	0.3545	-0.0118	0.3427	0.3062	0.0365
Pell:PlanWorkNo	64.3	0.028	0.0327	-0.0047	0.028	0.0411	-0.0131
Merit	62.0	0.8171	0.7825	0.0346	0.8171	0.7262	0.0909
Undeclared	55.5	0.089	0.0675	0.0214	0.089	0.1371	-0.0481
TestToUNR1	54.2	14.4655	14.347	0.1185	14.4655	14.207	0.2585
Male	49.6	0.3822	0.4311	-0.0489	0.3822	0.4793	-0.097
UnmetNeed	45.6	0.6524	0.6008	0.0516	0.6524	0.5575	0.0949
Ethmin	43.8	0.1878	0.2002	-0.0124	0.1878	0.2098	-0.022
Age19Plus	27.1	0.2537	0.2293	0.0244	0.2537	0.2872	-0.0335
Clark	21.1	0.3344	0.2872	0.0472	0.3344	0.3943	-0.0599
DelayedEntry6mo	14.0	0.022	0.0134	0.0086	0.022	0.0321	-0.01
Pell	9.2	0.2241	0.2452	-0.0211	0.2241	0.2473	-0.0233
OutOfState	-38.9	0.0725	0.1172	-0.0448	0.0725	0.1047	-0.0322
PlanWorkNo	-51.4	0.2422	0.2559	-0.0137	0.2422	0.2331	0.0091

Causal Inference Estimation

Average Outcomes and Naïve Estimator			
	On-Campus Students		
	Non-LLC	LLC	Difference
	(N=2531)	(N=607)	(ATT)
Fall momentum	79	88	**9**
Spring retention	93%	96%	**3%**
Spring momentum	85	91	**6**
Second yr retention	79%	90%	**11%**

Causal Inference Estimation

Estimation of LLC Participation Effect (ATT) Using Stratification, Matching, and Reweighting

	IPTW (N=3138)	*Subclass* (N=3084)	*NN1:1* (N=1200)	*NNk:1* (N=1105)	*Kernel1:1* (N=1214)	*Kernel3:1* (N=1628)	*Genetic* (N=1214)	*Avg Diff* to Naïve Est
Fall momentum	**4.6**	**4.8**	**4.5**	**3.4**	**4.4**	**5.0**	**3.8**	**-4.6**
Spring retention	**1.9%**	**1.8%**	**1.2%**	**1.7%**	**1.6%**	**2.6%**	**1.8%**	**-1.2%**
Spring momentum	**2.7**	**2.8**	**3.2**	**2.2**	**2.7**	**3.0**	**2.4**	**-3.3**
Second yr retention	**7.0%**	**7.2%**	**6.0%**	**6.9%**	**7.6%**	**8.1%**	**6.8%**	**-3.9%**

26

Causal Inference Estimation

Estimation of LLC Participation Effect in Weighted Regression Covariate Adjustment

	Unmatched (N=3138)	*IPTW* (N=3138)	*Subclass* (N=3084)	*NN1:1* (N=1200)	*NNK:1* (N=1105)	*Kernel1:1* (N=1214)	*Kernel3:1* (N=1628)	*Genetic* (N=1214)	*Avg Diff to Unmatched*
Fall momentum	**4.2**	**4.3**	**4.4**	**4.4**	**3.3**	**4.2**	**4.9**	**3.4**	-0.1
Spring retention	2.2%	1.1%	1.0%	0.6%	1.2%	1.2%	2.0%	1.5%	-0.9%
Spring momentum	**2.4**	**2.5**	**2.4**	**3.0**	**2.3**	**2.5**	**2.8**	**2.0**	0.1
Second yr retention	**6.9%**	**5.2%**	**6.0%**	**5.1%**	**6.0%**	**6.4%**	**6.0%**	**5.1%**	-1.2%

Note: Bolded=sig ≤0.05 alpha; SE bootstrapped 1000 replications (not listed)

Fall momentum post-treatment covariates: Acad ctr, div ctr, no Engl, no Math, online course

Other outcomes post-treatment covariates: as above plus fall unearned credits, flags for I/W/U/F grades

27

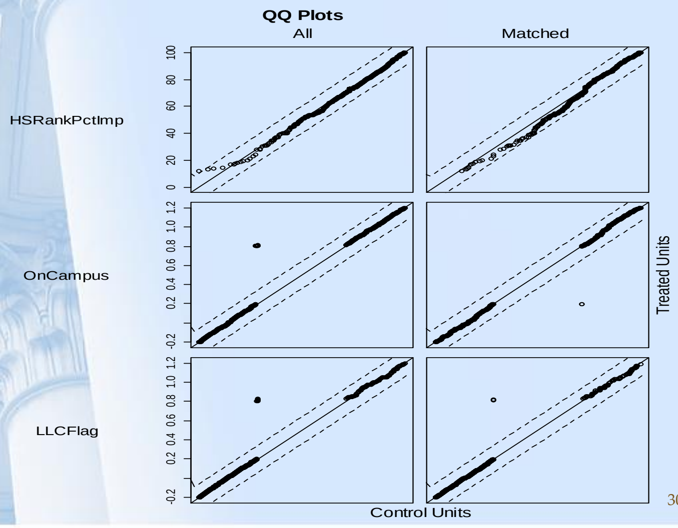
Findings on LLC Effect

- Impact on academic momentum (GPA, credits earned)
  - On average, inverse weighting, stratification, and matching reduce the LLC effect size by ~ 50%, 1:1 matching with optimization weights (genetic) showing the greatest reduction (60 to 70%)
  - Results are largely consistent for both fall and spring momentum
  - Weighted regression adjustment produces almost the same results after controlling for post-treatment selection covariates
- Impact on enrollment persistence
  - On average, inverse weighting, stratification, and matching reduce the LLC effect size by ~ 35 to 40%
  - Weighted regression adjustment reduces the effect size only for second-year retention (spring-fall) ~ 7 to 27% after controlling for post-treatment selection covariates
  - The LLC impact grows over time, suggesting varying time effects independent of measurement technique
- Asymptotically all PS estimators should yield the same results
- Sensitivity analysis of findings: Comparing LLC-eligible vs. ineligible (see slides 11, 15) indicates unconfoundedness is *more* plausible

Causal Inference Estimation

Average Outcomes and Naïve Estimator			
Math Support Center	Didn't Use	Used	Difference
	(N=4,107)	(N=780)	(ATT)
Fall momentum	78	83	**5**
Spring retention	92%	96%	**4%**
Spring momentum	84	86	**2**
Second yr retention	80%	86%	**6%**

Math Support Covariate Balance Check: Genetic P1000



30

Causal Inference Estimation

Estimation of Math Center Use (ATT) Using Stratification, Matching, and Reweighting

	OLS/Logit Regr	*IPTW*	*NNR:1*	*NN5:1^*	*Kernel1:1*	*Kernel3:1*	*Full:Opt Subcl W*	*Genetic*	*Gen:P1K*	*Avg Diff*
Treated	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	*to Regr*
Untreated	(N=4107)	(N=4107)*	(N=651)	(N=3538)	(N=780)	(N=2340)	(N=4107)	(N=680)	(N=677)	
Fall momentum	**4.1**	**4.0**	**3.8**	**3.8**	**8.7**	**5.8**	**7.7**	**4.0**	**4.5**	**1.1**
	(0.673)	(1.044)	(1.108)	(0.744)	(1.148)	(0.818)	(0.758)	(1.103)	(1.113)	
Spring retention	**4.7%**	**3.6%**	**3.2%**	**3.7%**	**6.9%**	**5.1%**	**6.1%**	**3.7%**	**4.6%**	**-0.1%**
	(1.2)	(1.2)	(1.3)	(0.9)	(1.3)	(0.9)	(0.9)	(1.3)	(1.3)	
Spring momentum~	**2.1**	**1.9**	**2.0**	**1.9**	**3.3**	**2.4**	**2.8**	**1.4**	**2.0**	**0.1**
	(0.468)	(0.673)	(0.692)	(0.503)	(0.746)	(0.544)	(0.508)	(0.705)	(0.707)	
Second yr retention	**7.1%**	**6.1%**	2.3%	**6.4%**	**11.0%**	**7.6%**	**10.6%**	**5.8%**	**5.4%**	**0.5%**
	(1.8)	(1.9)	(1.9)	(1.4)	(2.0)	(1.5)	(1.4)	(2.0)	(1.9)	

Bold=sig at 0.05 alpha; *Weighted N=780; ^Caliper 0.2 SD of TU; ~only retained students. SE are bootstrapped (1K replications)

Note: Retention logit coefficients are converted to Delta-p percentage points

31

Causal Inference Estimation

Estimation of Math Center Use (ATT) in Weighted Regression Covariate Adjustment

	OLS/Logit Regr	*IPTW*	*NNR:1*	*NN5:1^*	*Kernel1:1*	*Kernel3:1*	*Full:Opt Subcl W*	*Genetic*	*Gen:P1K*	*Avg Diff*
Treated	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	(N=780)	*to Regr*
Untreated	(N=4107)	(N=4107)*	(N=651)	(N=3538)	(N=780)	(N=2340)	(N=4107)	(N=680)	(N=677)	
Fall momentum	**3.4**	**3.3**	**3.5**	**3.5**	**8.2**	**5.5**	**7.2**	**3.8**	**4.1**	**1.5**
	(0.678)	(0.937)	(1.102)	(0.884)	(1.126)	(0.869)	(0.998)	(1.094)	(1.100)	
Spring retention	**3.9%**	**3.2%**	**3.0%**	**4.0%**	**7.1%**	**5.8%**	**7.4%**	**3.8%**	**4.5%**	**0.9%**
	(1.4)	(1.3)	(1.3)	(1.3)	(1.7)	(1.5)	(1.6)	(1.3)	(1.4)	
Spring momentum~	**1.7**	**1.5**	**1.6**	**1.6**	**1.7**	**1.5**	**1.5**	1.1	**1.3**	**-0.1**
	(0.426)	(0.551)	(0.604)	(0.461)	(0.647)	(0.471)	(0.488)	(0.603)	(0.616)	
Second yr retention	**4.9%**	**4.2%**	0.2%	**5.1%**	**6.7%**	**5.0%**	**8.2%**	3.4%	2.6%	**1.0%**
	(1.8)	(2.1)	(2.0)	(1.8)	(2.3)	(1.9)	(2.0)	(2.2)	(2.1)	

Bold=sig at 0.05 alpha; *Weighted N=780; ^Caliper 0.2 SD of TU; ~only retained students. SE are bootstrapped (1K replications)

First-semester post-treatment covariates: No English, No math, distance ed course, worked on campus, used diversity ct

Second-semester post-treatment covariates: as above plus first-semester credits dropped, first-semester I/W/F grades (flag)

Note: Retention logit coefficients are converted to Delta-p percentage points

32

Findings on Math Support Effect

- Impact on academic momentum (GPA, credits earned)
  - On average, PS matching increases the math support effect size in first semester by ~ 27%, full matching with optimal subclasses and kernel (1:1) showing the greatest rise in effect size (88 to 110%)
  - In contrast, the average PS-based effect size estimate does not differ much from standard OLS/logit estimates for second semester
  - Weighted regression adjustment produces very similar results, both full subclass and kernel (1:1) doubling the math support effect size
- Impact on enrollment persistence
  - On average, PS-based results show a greater effect size after controlling for post-treatment selection covariates, especially using kernel (1:1) and full subclass matching
- Given bias-variance tradeoff, weighted-distance matching (kernel, full subclass) may offer best estimate, assuming unconfoundedness holds.
- Asymptotically all PS estimators should yield the same results

33

Summary of Findings

- Living-and-learning community effect
  - Naïve estimators and non-weighted regression ***overestimate*** vis-à-vis propensity-score non/semi-parametric approaches
  - Assuming unconfoundedness, self selection may account for 50-70% of effect size on GPA/credits earned), up to 40% of effect size on enrollment persistence with non-weighted data
  - Unconfoundedness is more plausible given results on off-campus students (treatment 'ineligible')
- Math support center effect
  - ***Underestimation*** vis-à-vis propensity-score weighted data
  - Self selection may account for over 50% of effect on GPA/credits earned, ~ 25% of effect on persistence with non-weighted data
  - Bias-variance tradeoff in PS-weighted analysis suggests average PS-weighted results are lower-bound estimates

Further Research

- Expanded version of this study in

  Methodological Advances and Issues in Studying College Impact. N. A. Bowman and S. Herzog (eds.), New Directions for Institutional Research. (San Francisco: Jossey-Bass, forthcoming spring 2014).

- Foundational literature
  - *Higher Education: Handbook of Theory and Research, Vol. XXIV*, J. C. Smart (ed.), chapter by Reynolds and DesJardins
  - "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies," by P. Austin in *Multivariate Behavioral Research*, 46:399-424, 2011
  - "Some Practical Guidance for the Implementation of Propensity Score Matching", by M. Caliendo and S. Kopeinig in *Journal of Economic Surveys*, 22(1): 31-72, 2008

Link to presentation: http://www.unr.edu/ia/research