

**ECON 2320 – Economics of Labor and Population**

**Problem Set 3**

Samuel Brown

Fall 2011

Due: November 14

- (a) The difference-in-differences estimation results can be seen in Table 1. With estimated 1988–1990 and 1988–1992 gain scores of 5.05 and 7.73, respectively, these results (spuriously) suggest that the P-900 program had a large and significant effect on test scores.

Table 1: OLS Results for 1988–1990 and 1988–1992 Average Gain Scores

	(1) 1988–1990	(2) 1988–1992
P-900	5.051*** (0.847)	7.729*** (0.869)
$R^2$	0.055	0.120
$F$	35.544	79.124
$N$	658	651

Huber-White standard errors in parentheses.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

- (b) The estimates in (a) may be biased by noise and mean reversion. Consider the following error-components model in which  $i$  indexes individuals and  $j$  indexes schools:

$$y_{ij}^{88} = \lambda_j + u_j^{88} + \alpha_i^{88} \quad (1)$$

$$y_{ij}^{90} = \lambda_j + u_j^{90} + \alpha_i^{90} \quad (2)$$

where  $\lambda_j$  is a permanent school component,  $u_j^t$  is a transitory school component in year  $t$ , and  $\alpha_i^t$  is an individual component in cohort  $t$ . Suppose that  $\lambda_j$ ,  $u_j^t$ , and  $\alpha_i^t$  are all independent of each other. Suppose further that  $\lambda_j \sim iid(0, \sigma_\lambda^2)$ ,  $u_j^t \sim iid(0, \sigma_u^2)$ , and  $\alpha_i^t \sim iid(0, \sigma_\alpha^2)$ . Under this model the mean test scores for each school are as follows:

$$\overline{y_j^{88}} = \lambda_j + u_j^{88} + \frac{1}{N_j^{88}} \sum_{i \in j} \alpha_i^{88} \quad (3)$$

$$\overline{y_j^{90}} = \lambda_j + u_j^{90} + \frac{1}{N_j^{90}} \sum_{i \in j} \alpha_i^{90} \quad (4)$$

where  $N_j^t$  is the number of individuals in school  $j$  in year  $t$ . As long as  $u_j^t$  and  $\alpha_i^t$  are nontrivial and transitory and school enrollments are finite, there will be a positive correlation between the P900 eligibility indicator and the error term in the difference-in-differences regression model.

The mean reversion bias function for 1988–1990 average test score gains can be derived as follows:

$$\begin{aligned} \rho &= \frac{\text{Cov}(\overline{y_j^{90}} - \overline{y_j^{88}}, \overline{y_j^{88}})}{\text{Var}(\overline{y_j^{88}})} \\ &= \frac{\text{Cov}(\overline{y_j^{90}}, \overline{y_j^{88}})}{\text{Var}(\overline{y_j^{88}})} - 1 \\ &= \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_u^2 + \frac{1}{N_j^{88}} \sigma_\alpha^2} - 1 \end{aligned} \quad (5)$$

The mean reversion bias is zero if the variation in mean test scores is due entirely to variation across permanent school characteristics ( $\sigma_\lambda^2$ ). In this case the difference-in-differences estimates are unbiased. The mean reversion bias is increasingly negative the larger are the variance of school- and individual-level shocks ( $\sigma_u^2$  and  $\sigma_\alpha^2$ ), and the fewer students there are enrolled in school  $j$  in the first year ( $N_j^{88}$ ).

- (c) Since the treatment assignment is discontinuous, we can identify the program effect off of discontinuities in outcomes at the threshold, while controlling for “smooth” functions of 1988 scores to eliminate mean reversion bias. The resulting estimates will be valid as long as (i) the control functions “absorb” all of the reversion bias, (ii) there are no other discontinuities determining outcomes at the threshold, (iii) there is perfect program compliance, and (iv) there is no endogenous sorting at the threshold.
- (d) The relationship between probability of treatment and program eligibility can be seen in Figure 1. Evidently program compliance is high, as there is a visibly “sharp” discontinuity in smoothed program assignment at the threshold. This suggests that program eligibility strongly (though not perfectly) predicts the probability of receiving treatment. Though non-random selection away from the threshold may confound estimates, the degree of non-compliance appears to be small.

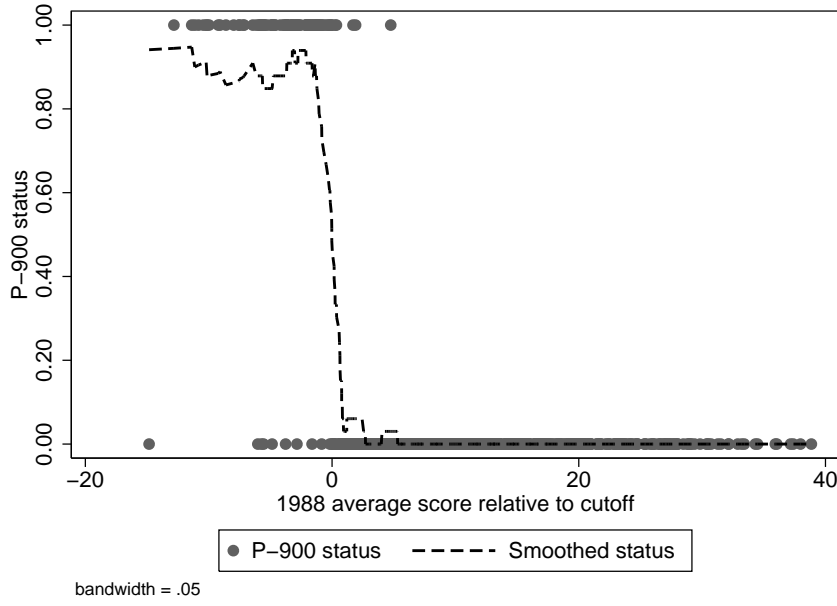


Figure 1: Program Allocation

- (e) The P900 eligibility indicator is a valid instrument as long as (i) it strongly predicts the probability of receiving treatment, and (ii) it is uncorrelated with the error term in the reduced-form regression model. *The latter condition is satisfied as long as mean reversion bias is controlled for.* The first stage and reduced-form equations of the two-stage least squares (2SLS) estimator are as follows:

$$P900_j = \delta_0 + \delta_1 \cdot ELIGIBLE_j + \psi(\cdot) + \nu_j, \text{ and} \quad (6)$$

$$\Delta y_j^t = \pi_0^t + \pi_1^t \cdot ELIGIBLE_j + \phi^t(\cdot) + \Delta \eta_j^t, \quad (7)$$

where  $t \in \{90, 92\}$  denotes the terminal year,  $\psi(\cdot)$  and  $\phi^t(\cdot)$  are bias control functions,  $\nu_j$  and  $\Delta \eta_j^t$  are error terms, and

$$\begin{aligned} ELIGIBLE_j &= \mathbf{1}(\overline{y_j^{88}} < y^*) \\ &= \mathbf{1}\left(\lambda_j + u_j^{88} + \frac{1}{N_j^{88}} \sum_{i \in j} \alpha_i^{88} < y^*\right) \end{aligned} \quad (8)$$

The coefficient  $\delta_1$  denotes the discrete “jump” in probability of treatment below the threshold, and  $\pi_1^t$  denotes the discrete difference in gain scores below and above the threshold. The 2SLS estimator uses the coefficient  $\delta_1$  to “inflate” the reduced-form coefficient, yielding an estimated effect of  $\pi_1^t/\delta_1$ . To reiterate, the instrument is valid as long as it is uncorrelated with the reduced-form error term  $\Delta\eta_j^t$ , that is to say,

$$E[ELIGIBLE_j \cdot \Delta\eta_j^t \mid \phi^t(\cdot)] = 0,$$

which holds if the control function  $\phi^t(\cdot)$  “absorbs” all of the bias.

- (f) The relationship between average 1988–1990 gain scores and program eligibility can be seen in Figure 2(a). Gain scores are “smooth” across the threshold, which suggests that the P-900 program had no effect on student performance between 1988 and 1990. The difference-in-differences estimates in part (a) above evidently “picked up” the mean-reversion effects characterized by the downward slope in Figure 2(a), rather than the true program effect.

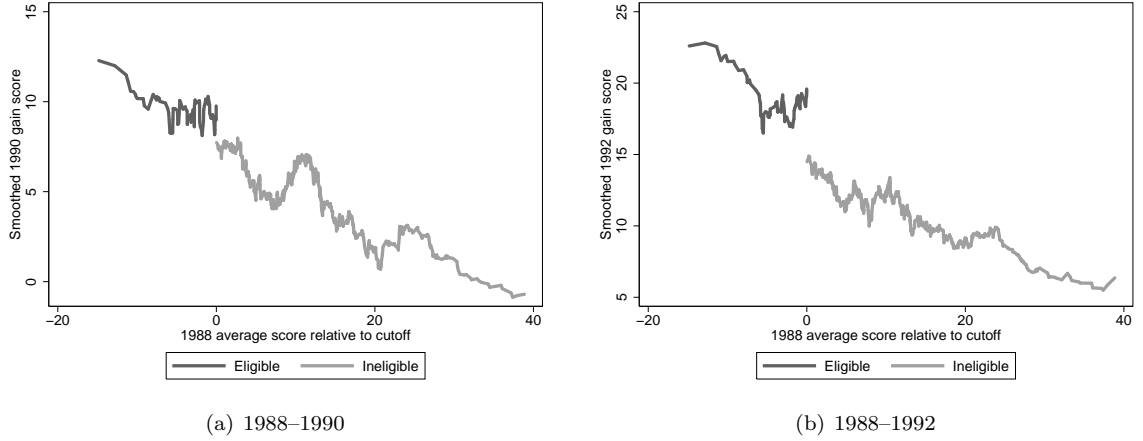


Figure 2: Gain Scores by 1988 Average Score Relative to the Regional Cutoff

The relationship between average 1988–1992 gain scores and program eligibility can be seen in Figure 2(b). Gain scores in this case exhibit a discrete “jump” across the threshold, which suggests that the P-900 program *did* have a positive effect on student performance between 1988 and 1992. The delayed program effect may be a result of changes in the flow of program resources over time, or to other, unobserved factors.

- (g) First-stage, reduced-form, and IV estimates of program effects can be seen in Table 2. In columns 1 and 4 there are no controls; columns 2 and 5 control for a cubic in the 1988 score; and columns 3 and 6 control for a cubic in the 1988 score, a quadratic in school enrollment, and the interaction between 1988 score and school enrollment.

In all cases the first-stage is “strong”, with program eligibility ensuring treatment roughly 85% of the time. This is unsurprising given the results in Figure 1 and the discussion in part (d). Thus as long as non-random selection away from the threshold is not a significant confounding factor, the instrument appears to be valid.

The reduced-form estimates in Columns 1 and 4 are comparable to those in part (a), presumably because mean reversion bias is still present, as would be expected given the discussion in part (e). When bias is controlled for in Columns 2 and 3, the reduced-form estimates of 1988–1990 program effects are reduced in magnitude by roughly a factor of 4 and are no longer significant. When bias is controlled for in Columns 5 and 6, the reduced-form estimates of 1988–1992 program effects are reduced in magnitude by roughly a factor of 2 but remain significant at the 95% level. These results are consistent with large mean reversion biases.

The 2SLS estimates in Panel C of Table 2 exhibit similar characteristics to those of the reduced-form estimates in Panel B, though they are slightly larger as would be expected.

Evidently controlling for bias eliminates a large portion of the estimated program effect, though it does not drive it to zero.

Table 2: 2SLS Results for Average Gain Scores, Using Eligibility for P-900 as an Instrument

	1988–1990			1988–1992		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: First-stage estimates (P-900)</i>						
Eligible	0.872*** (0.034)	0.842*** (0.068)	0.841*** (0.068)	0.872*** (0.034)	0.842*** (0.068)	0.841*** (0.068)
$\overline{y_j^{88}}$		0.025 (0.087)	0.018 (0.086)		0.025 (0.087)	0.018 (0.086)
$(\overline{y_j^{88}})^2$		−0.001 (0.001)	−0.000 (0.001)		−0.001 (0.001)	−0.000 (0.001)
$(\overline{y_j^{88}})^3$		0.000 (0.000)	0.000 (0.000)		0.000 (0.000)	0.000 (0.000)
$N_j^{88}$			0.001 (0.001)			0.001 (0.001)
$(N_j^{88})^2$			−0.000 (0.000)			−0.000 (0.000)
$\overline{y_j^{88}} \cdot N_j^{88}$			−0.000 (0.000)			−0.000 (0.000)
$R^2$	0.804	0.806	0.806	0.804	0.806	0.806
<i>Panel B: Reduced-form estimates (Average gain scores)</i>						
Eligible	4.975*** (0.825)	1.333 (1.450)	1.317 (1.452)	7.871*** (0.830)	3.453** (1.457)	3.420** (1.465)
$\overline{y_j^{88}}$		−0.415 (1.515)	−0.643 (1.551)		−1.488 (1.592)	−1.866 (1.608)
$(\overline{y_j^{88}})^2$		0.004 (0.026)	0.008 (0.027)		0.022 (0.027)	0.027 (0.028)
$(\overline{y_j^{88}})^3$		−0.000 (0.000)	−0.000 (0.000)		−0.000 (0.000)	−0.000 (0.000)
$N_j^{88}$			0.031 (0.045)			0.042 (0.039)
$(N_j^{88})^2$			−0.000 (0.000)			−0.000** (0.000)
$\overline{y_j^{88}} \cdot N_j^{88}$			−0.000 (0.001)			0.000 (0.001)
$R^2$	0.056	0.116	0.118	0.131	0.188	0.201
<i>Panel C: IV estimates (Average gain scores)</i>						
P-900	5.708*** (0.946)	1.583 (1.723)	1.567 (1.727)	9.030*** (0.953)	4.102** (1.731)	4.068** (1.743)
$\overline{y_j^{88}}$		−0.455 (1.493)	−0.671 (1.535)		−1.592 (1.567)	−1.939 (1.591)
$(\overline{y_j^{88}})^2$		0.005 (0.026)	0.009 (0.027)		0.024 (0.027)	0.029 (0.027)
$(\overline{y_j^{88}})^3$		−0.000 (0.000)	−0.000 (0.000)		−0.000 (0.000)	−0.000 (0.000)
$N_j^{88}$			0.029 (0.045)			0.036 (0.039)
$(N_j^{88})^2$			−0.000 (0.000)			−0.000** (0.000)
$\overline{y_j^{88}} \cdot N_j^{88}$			−0.000 (0.001)			0.000 (0.001)
$R^2$	0.056	0.116	0.118	0.131	0.188	0.201
$N$	658	658	658	651	651	651

Huber-White standard errors in parentheses.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

- (h) The 2SLS estimates using only schools within 7 points and 3 points of the eligibility threshold can be seen in Table 3. At 4.2 points, the estimated program effect within 7 points of the threshold is roughly equal to the analogous full-sample estimate of 4.1 points. However, when the bandwidth is restricted to  $\pm 3$  points, the estimated effect increases to 6.9 points. This suggests that there may be endogenous sorting at the threshold.

Table 3: 2SLS Results for 1988–1992 Average Gain Scores, within Narrow Bands of the Selection Threshold

	(1) Full Sample	(2) $\pm 7$ Points	(3) $\pm 3$ Points
P-900	4.068** (1.743)	4.168 (2.840)	6.916* (3.693)
$\overline{y_j^{88}}$	-1.939 (1.591)	66.462 (64.042)	251.391 (883.807)
$(\overline{y_j^{88}})^2$	0.029 (0.027)	-1.510 (1.465)	-5.426 (20.385)
$(\overline{y_j^{88}})^3$	-0.000 (0.000)	0.011 (0.011)	0.039 (0.156)
$N_j^{88}$	0.036 (0.039)	0.113 (0.157)	-0.778 (0.587)
$(N_j^{88})^2$	-0.000** (0.000)	-0.000 (0.000)	0.001* (0.001)
$\overline{y_j^{88}} \cdot N_j^{88}$	0.000 (0.001)	-0.002 (0.004)	0.013 (0.012)
$R^2$	0.201	0.113	0.106
$F$	30.553	4.672	.
$N$	651	245	102

Huber-White standard errors in parentheses.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

That there may be endogenous sorting at the threshold is supported by Figure 3. As can be seen in Figure 3(a), there is a visible “jump” in 1990 SES at the threshold. This suggests that at the time of program implementation high SES parents moved their children from eligible to ineligible schools, presumably because they saw P-900 eligibility as a signal of low school quality. There is not a corresponding “jump” in the 1990–1992 gain in SES shown in figure Figure 3(b), which suggests that sorting did not take place after 1990.

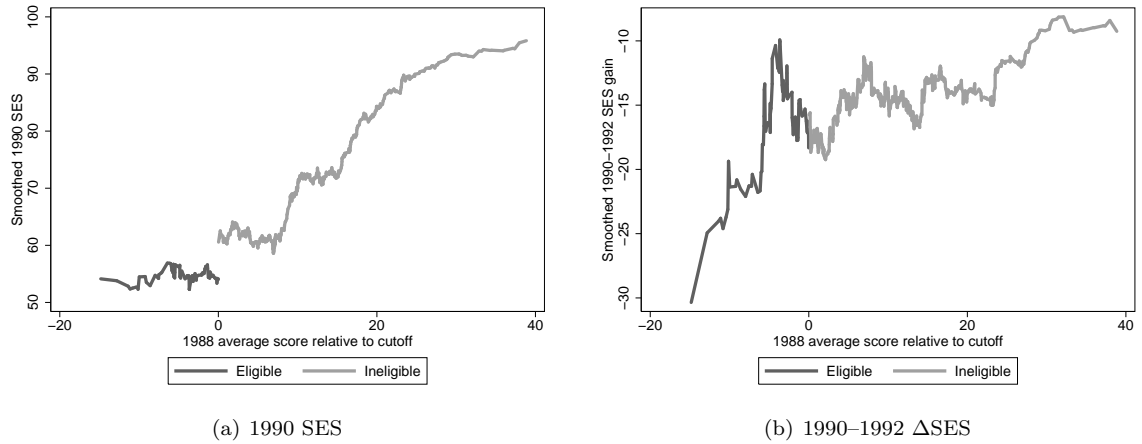


Figure 3: SES Status by 1988 Average Score Relative to the Regional Cutoff

The 1990 SES and the 1990–1992 change in SES in the two subsamples can be seen in Figure 4. These plots unambiguously show the same “jump” in 1990 SES at the threshold (Figures 4(a) and (c)), and the “smoothness” of the 1990–1992 gain in SES across the threshold (Figures 4(b) and (d)).

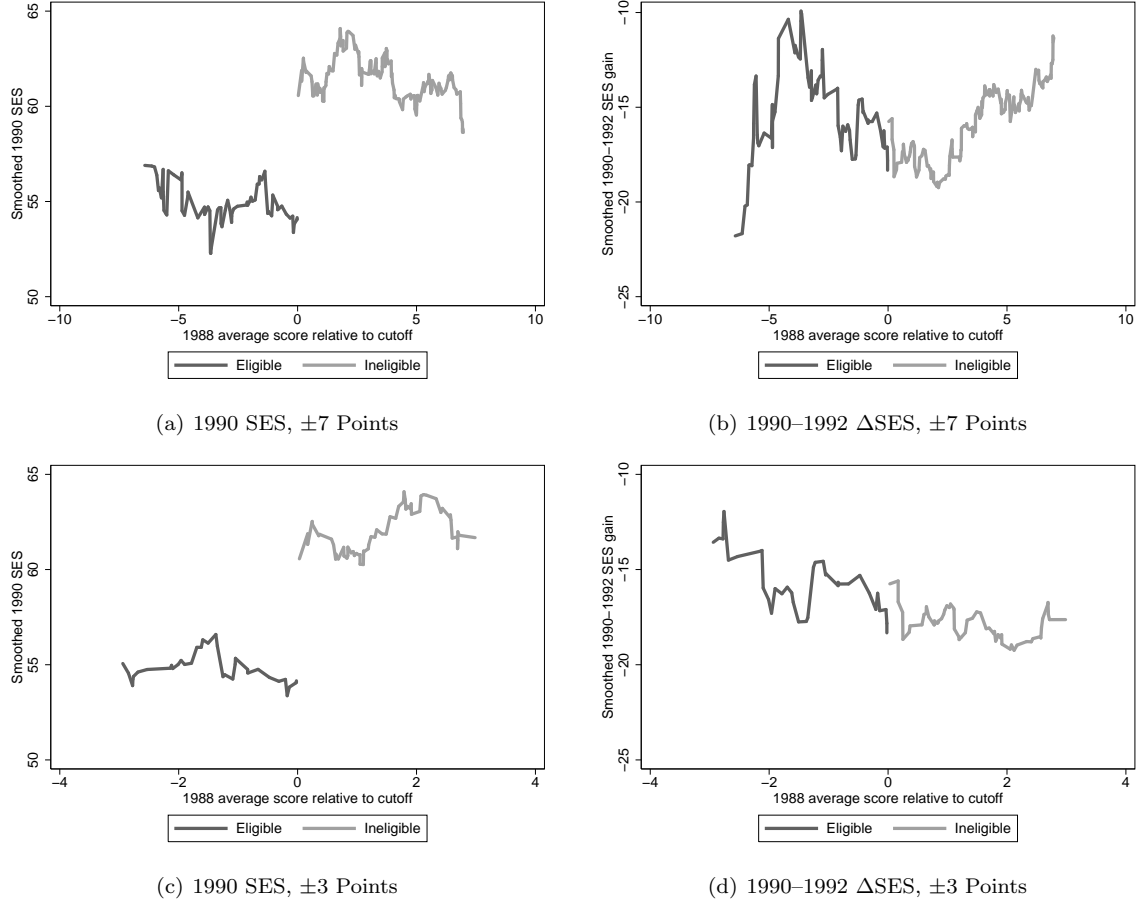


Figure 4: SES Status by 1988 Average Score Relative to the Regional Cutoff, within Narrow Bands of the Selection Threshold

Note that the sign of the estimated coefficients in the control function in Table 3 change as the sample of schools is narrowed. This is likely because in the smaller samples sorting bias dominates mean reversion bias, and these two effects work in opposite directions.

- (i) This problem set examines the effect of the P-900 program on average test scores in Chilean schools. It identifies the program effect off of discontinuities in outcomes at the program’s eligibility threshold. There several potential confounding factors: (i) **mean reversion** bias, (ii) non-random **selection** away from the threshold, and (iii) endogenous **sorting** at the threshold.

Standard difference-in-differences estimates overstate the program effect by confounding it with mean reversion. In other words, many schools that fell below the threshold in 1988 were small schools with “bad” draws of students, for whom an increase in test scores was to be expected in the proceeding years regardless of program enrollment. Failing to control for this increase overstates the estimated program effect.

A regression discontinuity design with controls for mean reversion can overcome the bias in the difference-in-differences estimates. This design is valid as long as there is no non-random selection away from the threshold. In other words, there must be a sufficiently high degree of program compliance. Moreover, in order to identify the estimated effect with the program effect, there must not be endogenous sorting at the threshold.

Though there is not perfect compliance, the data do not suggest a high degree of non-random selection away from the threshold. Controlling for mean reversion significantly decreases the estimated program

effect, and there is only evidence for such an effect in 1992 scores. There is evidence of sorting, though it appears to be limited to a narrow bandwidth around the threshold, and does not seem to affect estimates when the full sample is used.

In sum, failing to control for mean reversion bias significantly overstates the efficacy of the P-900 program, and the regression discontinuity design provides a “credible” way of removing this potential bias.