The power of RCTs – Teacher Performance Pay in India

Econ 140, Section 6

Jonathan Old

Roadmap

1. Recap: Why RCTs? [SA7-Q1]
2. Teacher Performance Pay [SA7-Q2]
3. Conclusion
4. Other tables if there is extra time

Any comments about the midterm?

Recap: Why RCTs? [SA7-Q1]

RCTs solve selection bias [SA7-Q1a]

We had:

$$\begin{aligned}\Delta = E[\text{Score}_i(1) - \text{Score}_i(0) | \text{PPay}_i = 1] \quad + \\ E[\text{Score}_i(0) | \text{PPay}_i = 1] - E[\text{Score}_i(0) | \text{PPay}_i = 0]\end{aligned}$$

- The second line was selection bias: The potential Score of individuals with and without Performance Pay is different
- If the treatment (PPay) is independent of the potential outcomes, then:

$$\begin{aligned}\text{PPay}_i \perp (\text{Score}_i(1), \text{Score}_i(0)) \\ \Rightarrow E[\text{Score}_i(0) | \text{PPay}_i = 1] = E[\text{Score}_i(0) | \text{PPay}_i = 0]\end{aligned}$$

  and selection bias will be zero.
- The the difference is equal to the ATT and **also the ATE**:

$$\begin{aligned}\Delta &= E[\text{Score}_i(1) - \text{Score}_i(0) | \text{PPay}_i = 1] \\ &= E[\text{Score}_i(1) - \text{Score}_i(0) | \text{PPay}_i = 0]\end{aligned}$$

RCTs and omitted variable bias [SA7-Q1b]

Two ways to think about this

1. Potential outcomes framework: We can just compare the differences in means, and we do not need to control for anything $\Rightarrow$ No OVB!
2. OVB formula:

$$\beta_S = \beta_L + \delta\gamma$$

.

- The bias is $\delta\gamma$, where γ measured the association between X (treatment) and the omitted variable
- But in an RCT, this association is zero $\Rightarrow$ No OVB!

Do RCTs solve everything? [SA7-Q1c]

Discuss in groups of 4: Why don't RCTs solve everything?

Your concerns may be:

- ethical
- practical
- remaining econometric challenges
- others

Do RCTs solve everything? [SA7-Q1c]

Discuss in groups of 4: Why don't RCTs solve everything?

Your concerns may be:

- ethical
- practical
- remaining econometric challenges
- others

Example from my life: Job interview

RCTs: Remaining challenges [SA7-Q1c]

- Non-compliance
- Differential attrition
- Spillover effects / general equilibrium effects
- Practical concerns
- Ethical concerns
- External validity (Generalizability)
- Placebo effects / experimenter demand effects / Hawthorne effect

RCTs have revolutionized economics

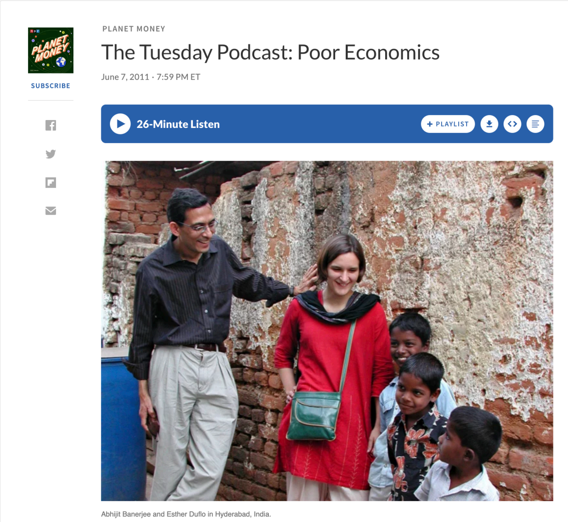
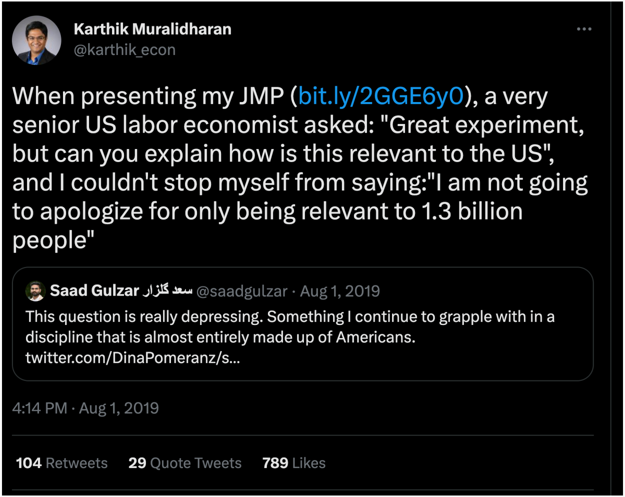


Figure 1: Abhijeet Banerjee and Ester Duflo. (Source)

Teacher Performance Pay [SA7-Q2]

Why does this matter?

Teacher absenteeism is a huge problem in many countries across the world!



Intervention [SA7-Q2b]

$$\text{Bonus} = \begin{cases} \text{Rs. } 500 \times (\%\text{ gain in average test scores} - 5\%) & \text{if gain} > 5\% \\ 0 & \text{otherwise.} \end{cases}$$

- How much bonus does a teacher get if their students' average test score changes by -10%, 0%, 5%, 10%, 50%?
- Is it ethical?
- What do you expect the effect to be? Which direction? Large or small? (teachers earn 180 USD without incentive)

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

	Incentives (Conditional on Improvement in Student Learning)		
Inputs	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?)

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

	Incentives (Conditional on Improvement in Student Learning)		
Inputs	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?)
  Much easier to implement, prevent undesirable effects within schools (spillovers, jealousy, etc.)

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

	Incentives (Conditional on Improvement in Student Learning)		
Inputs	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?)
  Much easier to implement, prevent undesirable effects within schools (spillovers, jealousy, etc.)
- What is the difference between the individual bonus and the group bonus?

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

INPUTS	INCENTIVES (Conditional on Improvement in Student Learning)		
	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?) **Much easier to implement, prevent undesirable effects within schools (spillovers, jealousy, etc.)**
- What is the difference between the individual bonus and the group bonus? **Incentive depends on own students (individual) or all students in the school (group-based)**

Balance table [Extra]

TABLE 2
Sample Balance across Treatments

	Control (1)	Group Incentive (2)	Individual Incentive (3)	p-Value (Equality of All Groups) (4)
	A. Means of Baseline Variables			
School-level variables:				
1. Total enrollment (baseline: grades 1–5)	113.2	111.3	112.6	.82
2. Total test takers (baseline: grades 2–5)	64.9	62.0	66.5	.89
3. Number of teachers	3.07	3.12	3.14	.58
4. Pupil-teacher ratio	39.5	40.6	37.5	.66
5. Infrastructure index (0–6)	3.19	3.14	3.26	.84
6. Proximity to facilities index (8–24)	14.65	14.66	14.72	.98
Baseline test performance:				
7. Math (raw %)	18.5	18.0	17.5	.69
8. Math (normalized; in SD)	.032	.001	−.032	.70
9. Telugu (raw %)	35.1	34.9	33.5	.52
10. Telugu (normalized; in SD)	.026	.021	−.046	.53

- Does this remind you of the problem set?
- If we test 100 baseline characteristics: How often would we expect p-values below 0.05 if our randomization worked?
- Do you think randomization worked in this case?

Regression specification (group work) [SA7-Q2c]

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{Incentives} + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk}. \quad (1)$$

The main dependent variable of interest is T_{ijkm}, which is the normalized test score on the specific subject, where i, j, k, and m denote the student, grade, school, and mandal, respectively. The term Y_0 indicates the baseline tests, and Y_n indicates a test at the end of n years of the program. Including the normalized baseline test score improves efficiency as a result of the autocorrelation between test scores across multiple periods.[24] All regressions include a set of mandal-level dummies (Z_m), and the standard errors are clustered at the school level. We also run the regressions with and without controls for household and school variables. The Incentives variable is a dummy at the school level indicating treatment status, and the parameter of interest is δ, which is the effect on test scores of being in an incentive school. The random assignment of the incentive program ensures that this is an unbiased and consistent estimate of the 1-year and 2-year treatment effects.

Regression specification (group work) [SA7-Q2c]

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{Incentives} + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk}. \qquad (1)$$

The main dependent variable of interest is T_{ijkm}, which is the normalized test score on the specific subject, where i, j, k, and m denote the student, grade, school, and mandal, respectively. The term Y_0 indicates the baseline tests, and Y_n indicates a test at the end of n years of the program. Including the normalized baseline test score improves efficiency as a result of the autocorrelation between test scores across multiple periods.[24] All regressions include a set of mandal-level dummies (Z_m), and the standard errors are clustered at the school level. We also run the regressions with and without controls for household and school variables. The Incentives variable is a dummy at the school level indicating treatment status, and the parameter of interest is δ, which is the effect on test scores of being in an incentive school. The random assignment of the incentive program ensures that this is an unbiased and consistent estimate of the 1-year and 2-year treatment effects.

For each of the four terms (yellow, red, blue, green), think:

- How is it measured?
- Why is it included?
- What does the coefficient on the term tell us?

Regression specification (group work) [SA7-Q2c]

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{Incentives} + \beta \cdot Z_m \qquad (1)$$
$$+ \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk}.$$

The main dependent variable of interest is T_{ijkm}, which is the normalized test score on the specific subject, where i, j, k, and m denote the student, grade, school, and mandal, respectively. The term Y_0 indicates the baseline tests, and Y_n indicates a test at the end of n years of the program. Including the normalized baseline test score improves efficiency as a result of the autocorrelation between test scores across multiple periods.[24] All regressions include a set of mandal-level dummies (Z_m), and the standard errors are clustered at the school level. We also run the regressions with and without controls for household and school variables. The Incentives variable is a dummy at the school level indicating treatment status, and the parameter of interest is δ, which is the effect on test scores of being in an incentive school. The random assignment of the incentive program ensures that this is an unbiased and consistent estimate of the 1-year and 2-year treatment effects.

Regression specification [SA7-Q2c]

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{ Incentives } + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk}.$$

- Outcome variable: $T_{ijkm}(Y_n)$: The test score of student i in grade k in school k in mandal (subdistrict) m, at the end of n years. **Normalized**.
- $T_{ijkm}(Y_0)$: test score at the *baseline* survey, for efficiency.
- *Incentives* is the variable of interest: an indicator (dummy) equal to 1 if a school was in the treatment group (and their teachers got incentive pay), and 0 otherwise. δ tells us the causal effect of performance pay for teachers on the test score of their students.
- Z_m is a set of fixed effects (dummy variables) for the Mandal, included for efficiency.
- All the ϵ terms are error terms.

Regression table [SA7-Q2d]

TABLE 3
IMPACT OF INCENTIVES ON STUDENT TEST SCORES
Dependent Variable: Normalized End-of-Year Test Score

	Year 1 on Year 0		Year 2 on Year 0	
	(1)	(2)	(3)	(4)
	A. Combined (Math and Language)			
Normalized lagged test score	.503***	.498***	.452***	.446***
	(.013)	(.013)	(.015)	(.015)
Incentive school	.149***	.165***	.219***	.224***
	(.042)	(.042)	(.047)	(.048)
School and household controls	No	Yes	No	Yes
Observations	42,145	37,617	29,760	24,665
R^2	.31	.34	.24	.28

- What are the rows and columns?
- Coefficient on normalized test score, significance?
- What is the coefficient on incentive school? Significant?
- Where do we see control variables in the table?
  How do they affect our coefficient of interest (δ)?

Regression table [SA7-Q2d]

- What are the rows and columns?

Regression table [SA7-Q2d]

- What are the rows and columns? The four columns are four different regression specifications (with and without controls, after one year and after two years). The rows give coefficients and their standard errors, and some additional information.
- Coefficient on normalized test score, significance?

Regression table [SA7-Q2d]

- What are the rows and columns? The four columns are four different regression specifications (with and without controls, after one year and after two years). The rows give coefficients and their standard errors, and some additional information.
- Coefficient on normalized test score, significance? The coefficient is around 0.5, with a standard error of 0.013. Significant at the 1% level. The coefficient means that students with a one standard deviation higher test score at the baseline are expected to have a 0.5 standard deviation higher test score at the endline, on average. Test scores are persistent, but correlation of test scores over time is not perfect.

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant?

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant? Around 0.15 without additional control variables, and 0.165 with control variables in year 1. The bonus incentive treatment has led to an increase in test scores by around 0.15 (0.165) standard deviations - a sizeable effect! In year 2, the treatment effect is even bigger (0.22 sd). The coefficient is always significant at the 1% level.

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant?
  Around 0.15 without additional control variables, and 0.165 with control variables in year 1. The bonus incentive treatment has led to an increase in test scores by around 0.15 (0.165) standard deviations - a sizeable effect! In year 2, the treatment effect is even bigger (0.22 sd). The coefficient is always significant at the 1% level.
- Where do we see control variables in the table?
  How do they affect our coefficient of interest (δ)?

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant?
  Around 0.15 without additional control variables, and 0.165 with control variables in year 1. The bonus incentive treatment has led to an increase in test scores by around 0.15 (0.165) standard deviations - a sizeable effect! In year 2, the treatment effect is even bigger (0.22 sd). The coefficient is always significant at the 1% level.
- Where do we see control variables in the table? How do they affect our coefficient of interest (δ)?
  Control variables are omitted from the table output, but we see that they slightly change the coefficient. The coefficient on the treatment becomes larger, but not by much. If including them changed a lot, we would question our assumptions about the RCT!

Benchmarking the effect [SA7-Q2e]

TABLE 10
IMPACT OF INPUTS VERSUS INCENTIVES ON LEARNING OUTCOMES
Dependent Variable: Normalized End-of-Year Test Score

	YEAR 1 ON YEAR 0			YEAR 2 ON YEAR 0		
	Combined (1)	Math (2)	Language (3)	Combined (4)	Math (5)	Language (6)
Normalized lagged score	.512***	.494***	.536***	.458***	.416***	.499***
	(.010)	(.012)	(.011)	(.012)	(.016)	(.012)
Incentives	.15***	.179***	.121***	.218***	.272***	.164***
	(.041)	(.048)	(.039)	(.049)	(.057)	(.046)
Inputs	.102***	.117***	.086**	.085*	.089*	.08*
	(.038)	(.042)	(.037)	(.046)	(.052)	(.044)
F-statistic *p*-value (inputs = incentives)	.178	.135	.298	.003	.000	.044
Observations	69,157	34,376	34,781	49,503	24,628	24,875
R^2	.30	.29	.32	.225	.226	.239

NOTE.—These regressions pool data from all 500 schools in the study: group and individual incentive treatments are pooled together as incentives, and the extra contract teacher and block grant treatments are pooled together as inputs. All regressions include mandal (subdistrict) fixed effects and standard errors clustered at the school level.

Benchmarking the effect [SA7-Q2e]

TABLE 10
IMPACT OF INPUTS VERSUS INCENTIVES ON LEARNING OUTCOMES
Dependent Variable: Normalized End-of-Year Test Score

	YEAR 1 ON YEAR 0			YEAR 2 ON YEAR 0		
	Combined (1)	Math (2)	Language (3)	Combined (4)	Math (5)	Language (6)
Normalized lagged score	.512***	.494***	.536***	.458***	.416***	.499***
	(.010)	(.012)	(.011)	(.012)	(.016)	(.012)
Incentives	.15***	.179***	.121***	.218***	.272***	.164***
	(.041)	(.048)	(.039)	(.049)	(.057)	(.046)
Inputs	.102***	.117***	.086**	.085*	.089*	.08*
	(.038)	(.042)	(.037)	(.046)	(.052)	(.044)
F-statistic *p*-value (inputs = incentives)	.178	.135	.298	.003	.000	.044
Observations	69,157	34,376	34,781	49,503	24,628	24,875
R^2	.30	.29	.32	.225	.226	.239

NOTE.—These regressions pool data from all 500 schools in the study: group and individual incentive treatments are pooled together as incentives, and the extra contract teacher and block grant treatments are pooled together as inputs. All regressions include mandal (subdistrict) fixed effects and standard errors clustered at the school level.

- What was the effect of the additional inputs?
- Which program did better? By how much?
- Can you tell whether the difference is significant?

Benchmarking the effect [SA7-Q2e]

- What was the effect of the additional inputs? Additional inputs increased test scores by around 0.1 standard deviations in the first year (a bit more for maths than for language). This is significant at the 1% level. Compared to year 0, test scores in year 2 were around 0.09 standard deviations higher, significant at the 10%-level.
- Which program did better? By how much? Comparing the two coefficients, incentives did better than inputs throughout. The difference is around 0.05 sd in the first year and 0.14 in the second year.
- Can you tell whether the difference is significant? The row "*F*-statistic *p*-value" gives the p-value for a statistical test whether the two coefficients are the same. We fail to reject this null hypothesis in year 1, but reject it at the 5% level in year 2.

Conclusion

What do we learn?

- We evaluated an RCT that is representative of schools in an Indian state with 80 million population
- The randomization design allows us to compare various treatment arms against each other
- The non-significant differences in the balance table indicate that randomization was successful
- We found sizeable effects of the pay-for-performance incentive scheme: The students in treated schools did significantly better on standardized tests
- The effect is larger than giving a comparable amount of money to the school directly
- **This study provides causal evidence that pay-for-performance for teachers increased students' outcomes**

Comments and criticisms?

- External validity: Would this also apply in other Indian states? Other developing countries? The world in general?
- Is this feasible in settings with no standardized tests?
- How do the costs compare against the benefits?
- . . .

Other tables if there is extra time

[Extra] Individual vs. group incentive effects

TABLE 8
GROUP VERSUS INDIVIDUAL INCENTIVES
Dependent Variable: Normalized End-of-Year Test Score

	Year 1 on Year 0			Year 2 on Year 0		
	Combined (1)	Math (2)	Telugu (3)	Combined (4)	Math (5)	Telugu (6)
Individual incentive school	.156***	.184***	.130***	.283***	.329***	.239***
	(.050)	(.059)	(.045)	(.058)	(.067)	(.054)
Group incentive school	.141***	.175***	.107**	.154***	.216***	.092*
	(.050)	(.057)	(.047)	(.057)	(.068)	(.052)
F-statistic *p*-value (testing group incentive school = individual incentive school)	.765	.889	.610	.057	.160	.016
Observations	42,145	20,946	21,199	29,760	14,797	14,963
R^2	.31	.299	.332	.25	.25	.26

NOTE.—All regressions include mandal (subdistrict) fixed effects and standard errors clustered at the school level.

[Extra] Teaching to the test: Repeat vs. non-repeat

TABLE 4
IMPACT OF INCENTIVES BY REPEAT AND NONREPEAT QUESTIONS
Dependent Variable: Percentage Score

	COMBINED		MATH		TELUGU	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Percentage score on non-repeat questions	.335***	.328***	.256***	.257***	.414***	.397***
	(.007)	(.007)	(.007)	(.008)	(.008)	(.007)
Percentage score on repeat questions	.352***	.42***	.252***	.386***	.452***	.468***
	(.006)	(.005)	(.007)	(.006)	(.007)	(.007)
Incremental score in incentive schools for nonrepeats	.030***	.039***	.033***	.046***	.027***	.033***
	(.009)	(.009)	(.009)	(.010)	(.010)	(.010)
Incremental score in incentive schools for repeats	.043***	.043***	.042***	.044***	.043***	.041***
	(.011)	(.011)	(.013)	(.012)	(.011)	(.013)
Test for equality of treatment effect for repeat and nonrepeat questions (F-statistic, p-value)	.141	.584	.374	.766	.076	.354
Observations	62,872	54,972	31,225	29,594	31,647	25,378
R^2	.24	.18	.26	.23	.29	.18

NOTE.—Repeat questions are questions that at the time of administering the particular test had appeared identically on any earlier test (across grades).

[Extra] Heterogeneous treatment effects A

TABLE 6
HETEROGENOUS TREATMENT EFFECTS
A. HOUSEHOLD AND SCHOOL CHARACTERISTICS

	Log School Enrollment (1)	School Proximity (8–24) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–4) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
				Two-Year Effect				
Incentive	−.198	−.019	.28**	.09	.224***	.226***	.233***	.219***
	(.354)	(.199)	(.130)	(.073)	(.054)	(.049)	(.049)	(.047)
Covariate	−.065	−.005	.025	.017	.068***	−.066	.029	.448***
	(.058)	(.010)	(.038)	(.014)	(.015)	(.042)	(.027)	(.024)
Interaction	.083	.018	−.02	.038**	−.003	−.013	−.02	.006
	(.074)	(.014)	(.040)	(.019)	(.019)	(.056)	(.034)	(.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243
				One-Year Effect				
Incentive	−.36	−.076	.032	.004	.166***	.164***	.157***	.149***
	(.381)	(.161)	(.110)	(.060)	(.047)	(.045)	(.044)	(.042)
Covariate	−.128**	−.016*	−.001	.017	.08***	.007	.016	.502***
	(.061)	(.008)	(.025)	(.013)	(.012)	(.035)	(.020)	(.021)
Interaction	.103	.017	.041	.042**	−.013	−.06	.002	.000
	(.081)	(.011)	(.031)	(.017)	(.016)	(.048)	(.025)	(.026)
Observations	42,145	41,131	41,131	38,545	38,525	42,145	39,540	42,145
R^2	.31	.32	.32	.34	.34	.31	.33	.31

[Extra] Heterogeneous treatment effects B

B. Teacher Characteristics (Pooled Regression Using Both Years of Data)

	Education (1)	Training (2)	Years of Experience (3)	Salary (Log) (4)	Male (5)	Teacher Absence (6)	Active Teaching (7)	Active or Passive Teaching (8)
Incentive	−.113	−.224	.258***	1.775**	.031	.15***	.084	.118
	(.163)	(.176)	(.059)	(.828)	(.091)	(.050)	(.054)	(.074)
Covariate	.003	−.051	−.001	−.034	−.084	−.149	.055	.131
	(.032)	(.041)	(.003)	(.066)	(.057)	(.137)	(.078)	(.093)
Interaction	.086*	.138**	−.009**	−.179*	.09	.013	.164*	.064
	(.050)	(.061)	(.004)	(.091)	(.069)	(.171)	(.098)	(.111)
Observations	53,737	53,890	54,142	53,122	54,142	53,609	53,383	53,383
R^2	.29	.29	.29	.29	.29	.29	.29	.29