Instructors' Teaching Performance and Students' Grade Expectations: An Application of the Stochastic Frontier Model

Myungsup Kim[*] Michael A. McPherson[†]

December 2012

[*] Associate Professor at the University of North Texas
[†] Associate Professor at the University of North Texas, email address: mcpherson@unt.edu, Telephone number: 1-940-565-2573, Fax number: 940-565-4426, Mailing Address: Department of Economics, University of North Texas, 1155 Union Circle, #311457, Denton, TX 76203, USA

Abstract

Student evaluation of teaching (SET) scores have been used widely to assess instructors' performance in teaching, and it has been argued that instructors may "purchase" better evaluation scores by inflating students' grade expectations. In this paper, using a stochastic frontier model, we explicitly control for the grade expectation not only as a regressor but also as a factor in instructors' inefficiency in reaching the frontier. This method permits a separation of unobserved factors that are distributed randomly ("luck") from an instructor's unobservable inherent teaching skill (efficiency or lack thereof). After controlling for various instructor, student and course characteristics we find that, with students' higher grade expectations, instructors earn higher SET scores and instructors' inefficiencies decrease.

Keywords: Student evaluation of teaching; Stochastic frontier model; Time-varying efficiency

JEL codes: A20; D24

Instructors' Teaching Performance and Students' Grade Expectations: An Application of the Stochastic Frontier Model

INTRODUCTION

Student evaluation of teaching (SET) scores have been used extensively in assessing the quality of teaching and are an important component in evaluating instructors for promotion and tenure. The use of SET scores as a measurement of teaching effectiveness (Gramlich and Greenlee 1993) and an instructor’s likely (Krautmann and Sander 1999) or unlikely (Seiver 1983) "purchase" of SET scores by inflating students' expected grade have been discussed in the literature. McPherson, Jewell, and Kim (2009) discuss factors contributing to SET scores and the adjustment of SET scores. The purpose of this paper is to analyze the role of students' grade expectation in instructors' teaching evaluations with a stochastic frontier model so that both evaluation scores and inefficiencies can be a function of expected grade. With the inefficiency estimates, instructors can be ranked based on their inabilities in achieving higher SET scores.

The stochastic frontier model is a popular models in evaluating decision-making units' outputs. Some examples of decision-making units analyzed in this way include municipalities (Lorenzo and Sánchez 2007), fishermen (Alvarez and Schmidt 2006), students (Dolton, Marcenaro, and Navaro 2003), as well as firms. The frontier model has been used to examine street lighting service by municipalities as a function of human resources, capital goods and environmental variables, fishermen's skill and luck in explaining fish catches, and students' time allocation problem in maximizing exam performance. In this paper, instructors are treated as producers, and they produce teaching service whose quality is measured by SET scores. This methodology allows us to separate unobservable or unobserved factors that are distributed randomly, such as the disproportionate presence of enthusiastic and appreciative students in a particular class or an instructor’s physical attractiveness from an instructor’s unobservable inherent teaching skill, how charming he might be, his organizational talent, etc. One might think of the former category as luck; the latter may be thought of as efficiency (or lack thereof). The frontier is set by an instructor with the highest evaluation score, and the frontier is allowed to be stochastic due to statistical noise. The difference between the highest evaluated instructor and another instructor is the latter instructor's relative inefficiency in producing teaching service. In one of the models that we consider, the inefficiency is a function of students' expected grade which may be controlled by an instructor.

While there have been a number of studies that have examined the effects of various factors on SETs, these have generally conflated efficiency and luck. This paper represents the first effort to estimate the determinants of SETs while separating these effects statistically.

This paper is organized as follows. We begin with the discussion of the stochastic frontier model and inefficiency in the setting of evaluating instructors. Next, we provide data description, followed by estimation results, and conclusions.

MODEL

The stochastic frontier model proposed by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977) has the following form, $y = x\beta + v - u$, where y is the decision making units' output levels, x is the vector of inputs, v is a symmetric error term of random factors, and u is a positive one-sided error term representing efficiency so that $-u$ represents shortfall or inefficiency from the stochastic frontier of ($x\beta + v$).

The idiosyncratic error term, v is a typical regression error term representing random factors affecting SET scores such as the luck of having in one’s class generous or parsimonious students or a random discovery of good instructional materials to be used in class. The inefficiency term in our setting represents instructors' inherent skill in teaching which may limit them from realizing their potential in teaching or earning higher evaluation scores. An instructor with some inefficiency does not reach the SET score frontier. The shortfall may be due to unobservable factors such as differences in inherent teaching ability, the students' views on an instructor's charm, or an instructor's responsiveness to students' emails or requests to meet. However, at least in our data set inefficiency can be also affected by an observable factor; this is a factor over which instructors can exercise some control. Also, when instructors are observed over time, it is possible to control for the fact that inefficiencies may change over time. This could be the result, for example, of learning by instructors.

In this spirit, we consider two frontier models in which the inefficiency is modeled as a function of time and/or students' expected grade.[1] The first one is the Battese and Coelli (1992) model (in short, BC92). For an unbalanced panel data with i $=1, \ldots, N$ and $t = 1,\ldots,T_i$, the stochastic frontier model is

$$y_{it} = x_{it}\beta + v_{it} - u_{it} \tag{1}$$

and the one-sided error term is defined as

$$u_{it} = \eta_{it} u_i = \exp\{-\eta(t - T_i)\} u_i \tag{2}$$

where η is an unknown scale parameter, the u_i's are assumed to be independent and identically distributed (iid) non-negative truncations of $N(\mu, \sigma_\mu^2)$, and the v_{it}'s are iid N(0,

σ_v^2). This model also assumes that u_i and v_{it} are independent of each other and of the regressors in the model.

When η is positive, u_{it} decreases over time; with a negative η, u_{it} increases over time; with η = 0, u_{it} remains constant. It is a somewhat rigid parameterization in that the rate of change for η_{it} with respect to *t* is non-decreasing regardless of the sign of η since $\partial^2 \eta_{it} / \partial t^2 = \eta_{it}\eta^2 \geq 0$. However, it is a simple function of time and allows a multiplicative decomposition of u_{it} which has the so-called "scaling property" (Wang and Schmidt 2002). *u* is a product of a function of exogenous variables, z and the part of *u* that does not depend on z, *u**, so that *u* = f(z; δ)*u**.

Battese and Coelli (1992) showed that the minimum mean squared error predictor of technical efficiency, exp(-u_{it}) is

$$TE_{it} = E\left[\exp(-u_{it})\middle|\varepsilon_{it}\right] = \left[\frac{1-\Phi\left(\eta_{it}\sigma_i^* - \mu_i^* / \sigma_i^*\right)}{1-\Phi\left(-\mu_i^* / \sigma_i^*\right)}\right]\exp\left(-\eta_{it}\mu_i^* + 0.5\eta_{it}^2\sigma_i^{*2}\right) \qquad (3)$$

where $\varepsilon_{it} = v_{it} - u_{it}$, $\mu_i^* = \left(\mu\sigma_v^2 - \sum_t \eta_{it}\varepsilon_{it}\sigma_u^2\right) / \left(\sigma_v^2 + \sum_t \eta_{it}^2\sigma_u^2\right)$,

$\sigma_i^{*2} = \left(\sigma_v^2\sigma_u^2\right) / \left(\sigma_v^2 + \sum_t \eta_{it}^2\sigma_u^2\right)$, and Φ(·) represents the cumulative distribution function of the standard normal random variable.

The second model is the Battese and Coelli (1995) model (hereafter, BC95) in which the one-sided term is defined as

$$u_{it} = z_{it}\delta + w_{it} \qquad (4)$$

where u_{it} follows a truncated normal distribution with the mean of $z_{it}\delta$ and the variance of σ_u^2. The technical efficiency is $TE_{it} = \exp(-u_{it})$. This specification is flexible because the inefficiency term can be a function of various factors including a time trend. Given the possibility that instructors, good teaching can be rewarded by higher evaluation scores and instructors can influence students' grade expectations, grade expectations are also included in z as a regressor. A time trend and instructor's age are also included in z because inefficiency can change over time or instructors may become more experienced as they age.

DATA

We use the panel data collected from the University of North Texas (UNT) Academic Records office and from the UNT Department of Economics as used in McPherson, Jewell, and Kim (2009). UNT is a comprehensive state university with more than 36,000 students. The Department of Economics has approximately 250 undergraduate majors but serves many thousands of students from other departments in its various course offerings. The UNT Economics department is broadly similar to programs at other large, state universities; thus, our data set is representative of that group. These data represent 24 consecutive semesters between January 1994 and December 2005. Our data comprise 602 individual principles of economics classes and 379 individual upper-level classes. The dataset has a total of 63 different instructors, 19 of whom are female and 22 of whom are non-white. On average, an instructor teaches about 2.10 courses with a standard deviation of 0.89. The variables used in this study are discussed below, and their summary statistics are presented in Table 1.

[Insert Table 1 about here]

The dependent variable is the logarithm of an instructor's SET scores (*eval*).[2] The department's SET is calculated by averaging responses to four statements: "I would take another course that was taught in this way;" "The instructor did not synthesize, integrate, or summarize effectively;" "Some things were not explained very well;" and "I think that the course was taught quite well." The average evaluation scores can range from one to four, with a four representing the best possible SET score. Following earlier studies employing these data, we utilize the department's chosen measure. The average SET score in principles classes is 3.325; the comparable statistic for upper-level classes is slightly higher at 3.491.

Following the literature, the determinants of SET scores are likely to fall into several categories. The first group of SET score determinants comprises characteristics specific to instructors. In order to control for observable characteristics, we include the gender (*male*), race (*white*), total semesters of university teaching experience (*exper*), whether the instructor is a teaching fellow or adjunct (*adjunct*), and the instructor's age (*age*). Even if race and gender do not have an impact on teaching ability, an instructor's observable characteristics (including race, gender, and age) can still affect SET scores if some bias exists in the evaluation process. For instance, research exists suggesting that students perceive female instructors differently than men. Experience should be positively related to SET scores, since more time in the classroom should increase the quality of one's teaching. Instructors who are adjuncts or a teaching fellows (*adjunct*) have no research and limited service responsibilities and consequently may focus on teaching. As a result, we expect that such faculty will have higher SET scores, all else equal. Holding constant the effect of experience, we expect SET scores to fall as instructors age. There are several reasons to expect such an effect. First, faculty members may spend less time on teaching activities and allocate more time towards research or administrative duties, as they age. Second, students may simply prefer courses taught by younger instructors. Third, as an instructor ages, she becomes further removed from her graduate education. Without additional training, an instructor's human capital, in terms of her knowledge of the current state of the discipline, will inevitably depreciate.

A second group of factors that may influence SET scores are characteristics of the students in each class; these include the proportion of students participating in the evaluation exercise that major in economics (*pctmajor*), the proportion of students in each class that is female (*pctfemale*), the average grade expected by students in the course (*expgrade*), and the percentage of students enrolled in the class that participate in the evaluation exercise (*response*). The proportion of students majoring in economics may affect evaluation scores in that economics majors are presumably more interested in economics in general and may be more likely to recognize quality teaching in economics. The gender composition of the respondents may impact SET scores if there are differences in the standards used by male and female students in evaluating teaching. *expgrade* is measured on the usual four-point scale, averaging 2.912 for principles courses in the data and 3.222 for upper-level courses. *A priori* one would suspect that students expecting higher course grades are more likely to give instructors high evaluation scores. The effects of the response rate on SET scores may be less clear. On the one hand, *response* may be correlated with student enthusiasm for the course so that higher response rates lead to higher evaluation scores. On the other hand, the response rate may be higher in courses in which attendance is required or perceived as important. This might cause the relationship between *eval* and *response* to be an inverse one. Finally, we include a time trend variable (*semester*) in order to control for changes in the composition and preferences of students over time.

A final type of variable that may determine SET scores is characteristics of the course. We include a series of dummy variables indicating the type of the course. For principles classes, *prncpl-micro* equals one if the course is a principles of microeconomics section; *prncpl-macro* = 1 if a principles of macroeconomics section. Upper-level courses are divided into five categories: *intermediate* equals one for intermediate-level theory courses required of all Economics majors, including Intermediate Microeconomics, Intermediate Macroeconomics, and Money and Financial Institutions; *elective3* equals one if the course is a junior-level elective course; *elective4a* equals one if the course is a senior-level elective course without an intermediate theory prerequisite; *elective4b* equals one if the course is a senior-level elective course that does have an intermediate theory prerequisite; and *quantitative* equals one if the course is a statistics or econometrics course.

In order to fit the stochastic frontier model with time-varying efficiency, some variables take different forms for estimation in the BC92 and BC95 models. Typically an instructor teaches more than one course per semester. Given multiple SET scores of an instructor in each time period, we use an average of SET scores from all the courses an instructor teaches in each semester in defining the dependent variable for the BC92 and BC95 models. This gives us a single observation for an instructor in each time period. Accordingly, characteristics of the students in each course such as *pctmajor*, *pctfemale*, *response* and *size* are also averaged. All the variables measuring course characteristics in the models are modified to count the number of courses an instructor teaches each semester for each course category, unlike the dummy variables used in the first model we estimate, M1. For example, *prncpl* in the BC92 and BC95 models counts the total number of principles courses an instructor teaches each semester, and the same applies to other course variables.

ESTIMATION RESULTS

Table 2 presents the estimation results for three different models. The estimation is done by Stata and Coelli's Frontier software. In the first model labeled as M1, the one-sided error term is time-invariant: that is, both u and η in the BC92 model are set to zero in Ml. In the BC92 model, inefficiency is a function of time, and a positive and significant $\hat{\eta}$ indicates that the non-negative inefficiencies would decrease over time. In the BC95 model, students' expected grade is found to be a significant factor in u_{it}. Given statistically significant parameter estimates in the one-sided error term across the three estimation outputs, the frontier model specification is preferred to the regression model without the inefficiency term.

[Insert Table 2 about here]

Figure 1 presents SET score trend for faculty members with the three highest overall SET scores and three lowest. It seems to indicate that there is no noticeable upward or downward trend in their SET scores, although there are ups and downs between semesters. This is confirmed by the insignificant estimates on the trend term (*semester*) in the three models.

[Insert Figure 1 about here]

However, there are several factors that do affect evaluation scores. Students' expected grade plays the largest role in affecting SET scores in all three specifications. In the BC95 results, if a student's grade expectation changes from a C to a B (that is, 50 percent increase from a grade of 2 to 3), instructor's evaluation scores would go up by 5.845 percent. At the mean value of evaluation scores, this translates to about 0.2 higher evaluation score (= 3.389 x 0.05845) which is not a trivial amount given that the sample standard deviation is 0.311. This effect is smaller than an increase between 0.34 and 0.56 reported in the paper by Krautmann and Sander (1999). This is because in the BC95 model expected grade affects evaluation scores through two different channels: one is through a regressor, and the other is through an inefficiency term. The expected grade is also a significant factor in reducing instructor's inefficiency even after controlling for trend and age in the BC95 specification.[3] These effects support the arguments that instructors may "purchase" higher evaluation scores by inflating students' grade expectation since with a smaller inefficiency amount, the instructor can be closer to the frontier of teaching evaluation. This finding is consistent with much of the earlier literature, including McPherson (2006), Isely and Singh (2005), Krautmann and Sanders

(1999), and Dilts (1980). Other significant factors are instructor and course characteristics.

In the BC95 results, male instructors receive about three percent higher SET scores than female instructors. Being white and experienced also positively affect SET scores. As an instructor gets older, SET scores would decrease by about 0.34 percent per year, holding other factors constant. This may be because students prefer younger instructors. Adjuncts also earn higher evaluation scores, as expected. In terms of course characteristics, SET scores are negatively affected when an instructor teaches more principles and intermediate courses.[4] This is to be expected since students with non-economics majors are taking the required principles and intermediate courses and may not value such courses as highly as students who are or will become economics majors. In the case of intermediate courses, these are also required of economics majors, and so SET scores me be lower relative to upper-division classes that students elect to take. As instructors teach more upper level courses, their SET scores will be higher. For example, teaching an additional quantitative course would raise SET scores by 2.69 percent.

Unlike the instructor and course characteristics, student characteristics in classes are not significant factors. This may be due to the frontier model's separation of efficiency from luck in the composite error term. In a study without that separation, McPherson, Jewell, and Kim (2009) found that some student characteristics do affect SET scores. In particular, they find that principles of economics classes with higher proportions of economics majors and of female students tend to assign instructors higher SET scores. In other economics classes, however, these effects are negative (although only significant in a statistical sense in the case of the proportion of economics majors).

McPherson (2006) found a positive effect of *pctmaj* in upper-level economics courses, and McPherson and Jewell (2007) found that the gender composition of graduate-level classes has no effect on SET scores.

We also compute tenure-track (instructors with *adjunct* = 0) faculty members' efficiency estimates in all three models and list them in comparison with SET scores in Table 3. Each instructor's technical efficiencies averaged over semesters are listed in the three columns before the last: $\hat{T}E_{it,BC95}$ is calculated based on estimates in BC92; $\hat{T}E_{it,BC92}$ is based on estimates in BC92; and $\hat{T}E_{it,M1}$ is based on estimates from M1. Highly ranked instructors tend to be more efficient than their lower-ranked colleagues. Instructor 35 who is male and white with 22 semesters of teaching has earned the highest overall SET score of 3.6924 and is ranked in the top half in terms of efficiency; instructor 30 who is male and non-white with 12 semesters of teaching has the lowest average SET score of 3.1628 and is the least efficient. Given the BC95 results, instructor 75 is most efficient but does not have the highest evaluation scores, and instructor 35 has the highest overall evaluation scores but is ranked sixth in terms of efficiency. This indicates that instructor 75 can earn higher evaluation scores by becoming more efficient; this is even more the case with instructor 35. This might be accomplished by these instructors by means of laudable methods such as the mastering of more effective pedagogies or by becoming better organized or more responsive to students' needs. However, it is also true that instructors could also reduce their inefficiency measures by using the less laudable method of inflating student grade expectations.

[Insert Table 3 about here]

Based on the BC95 estimates, technical efficiencies for faculty members with top three and bottom three evaluation scores are presented in Figure 2. Overall, highly evaluated faculty members are more efficient, but not always. Instructors 16 and 21 who are two of the bottom three are, in some semesters, just slightly less efficient than one of the top three instructors. Figure 2 also indicates that instructors' efficiency levels are stable over time, which is confirmed by the insignificant coefficient estimate on the trend term.

[Insert Figure 2 about here]

CONCLUDING REMARKS

We analyze instructors' performance on teaching using the stochastic frontier model with time-varying technical inefficiency. In the frontier model, the unobservable luck or random factor in the SET score determination is separated from an instructor's efficiency term that affects SET scores. Efficiency estimates represent the part of the composite error term that is free from random factors or luck, and we model efficiency as a function of expected grade. Also, students' expected grade in all models is shown to be a significant factor in explaining SET scores while controlling for various instructor, student, and course characteristics. The significant parameter estimates in the one-sided error term support the presence of technical inefficiencies in teaching. Higher grade expectation reduces instructor's inefficiency in reaching the SET score frontier. This supports an argument that an instructor may "buy" better evaluation scores by inflating students' grade expectations. Also, the comparison of SET scores and technical inefficiencies shows that highly evaluated instructors tend to be more efficient than instructors with low evaluation scores, but not always.

What is not considered in this paper is that possible positive or negative effect of faculty members’ research or service activities on teaching: students may value faculty members with active research agenda. On the other hand, such faculty members may have less time to prepare for classes and to engage students, which may negatively affect SET scores. This paper is an attempt to provide additional evidence on the link between evaluation scores and grade expectation where opposing views exist. Perhaps an improved way of evaluating instructors should somehow take into account of grade distribution in classes to offset seemingly higher evaluation scores generated by students' higher grade expectations.

NOTES

1. Ignoring an exogenous variable in inefficiency term can lead to inconsistent estimates if it is correlated with a regressor. For example, consider the simple linear regression model with inefficiency, $y = \alpha_1 + \alpha_2 x + \varepsilon$ where $\varepsilon = v - u$. Suppose that $u = \alpha_3 z + w$ where *z* is an exogenous variable correlated with *x*, and *w* is uncorrelated with *v* and *x*. The least squares estimator of slope coefficient, $\hat{\alpha}_2$, is inconsistent. Specifically,

   $\text{plim}\,\hat{\alpha}_2 = \alpha_2 - \alpha_3 \frac{Cov(x,z)}{Var(x)}$.

2. SET scores result from survey near the end of the semester. SET forms are distributed without previous announcement and are anonymous.

3. When only one of the trend and age terms besides expected grade is included, the estimated coefficient on expected grade is still significant.

4. The course variables are measured in counts in BC92 and BC95, unlike in dummy variables in M1.

REFERENCES

Aigner, D.J., C.A.K. Lovell, and P. Schmidt, 1977, Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 2l-37.

Alvarez, A., and P. Schmidt, 2006, Is skill more important than luck in explaining fish catches?, *Journal of Productivity Analysis* 26, 15-25.

Battese, G.E., and T.J. Coelli, 1992, Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India, *Journal of Productivity Analysis* 3, 153-169.

_______, l995, A model for technical inefficiency effects in a stochastic frontier model for panel data, *Empirical Economics* 20, 325-332.

Dilts, D.A., 1980, A statistical interpretation of student evaluation feedback, *Journal of Economic Education* 11, 10-15.

Dolton, P., O.D. Marcenaro, and L. Navaro, 2003, The effective use of student time: a stochastic frontier production function case study, *Economics of Education Review* 22, 547-560.

Gramlich, E.M., and G.A. Greenlee, 1993, Measuring teaching performance, *Journal of Economic Education* 24, 3-13.

Isely, P. and H. Singh, 2005, Do higher grades lead to favorable student evaluations? *Journal of Economic Education* 36, 29-42.

Krautmann, A.C., and W. Sander, 1999, Grades and student evaluation of teachers, *Economics of Education Review* 18, 59-63.

Lorenzo, J.M.P., and I.M.G. Sánchez, 2007, Efficiency evaluation in municipal services: an application to the street lighting service in Spain, *Journal of Productivity Analysis* 27, 149-162.

McPherson, M.A., 2006, Determinants of how students evaluate teachers, *Journal of Economic Education* 37, 3-20.

McPherson, M.A. and R.T. Jewell, 2007, Leveling the playing field: should student evaluation scores be adjusted?, *Social Science Quarterly* 88, 868-881.

McPherson, M.A., R.T. Jewell, and M. Kim, 2009, What determines student evaluation scores? A random effects analysis of undergraduate economics classes, *Eastern Economic Journal* 35, 37-51.

Meeusen, W., and J. van den Broeck, 1977, Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18, 435-444.

Seiver, D.A., 1988, Evaluations and grades: a simultaneous framework, *Journal of Economic Education* 14, 32-38.

Wang, H.J., and P. Schmidt, 2002, One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels, *Journal of Productivity Analysis* 18, 129-144.

Table 1: Summary Statistics

Variable	**Mean**	**S.D.**	**Min**	**Max**
eval	3.389	0.311	1.45	4
expgrade	3.032	0.311	2.214	4
semester	12.276	6.789	1	24
male	0.624	0.485	0	1
white	0.74	0.439	0	1
age	38.152	8.48	23	72
exper	15.084	11.237	1	45
adjunct	0.59	0.492	0	1
pctmajor	0.163	0.24	0	1
pctfemale	0.519	0.119	0	1
response	0.684	0.127	0.267	1
size	53.205	36.357	5	289
prncpl-micro	0.259	0.438	0	1
prncpl-macro	0.355	0.479	0	1
intermediate	0.164	0.371	0	1
elective3	0.044	0.205	0	1
elective4a	0.072	0.259	0	1
elective4b	0.074	0.263	0	1
quantitative	0.032	0.175	0	1
sample size			981	
prncplt	1.295	1.255	0	4
intermediatet	0.346	0.639	0	3
elective3t	0.092	0.297	0	2
elective4at	0.153	0.436	0	2
elective4bt	0.157	0.398	0	2
quantitativet	0.067	0.266	0	2
sample size			465	

Note: S.D. = standard deviation

Table 2: Estimation Results

		Dependent variable: log(eval)			
	M1	BC92		BC95	
log(expgrade)	0.2363***	0.2698***	0.1169***	expgrade	-1.1261***
	(0.0249)	(0.0389)	(0.0394)		(0.4107)
semester	0.0003	-0.0001	0.0007	semester	0.0052
	(0.0003)	(0.0008)	(0.0005)		(0.0033)
male	0.0200***	0.0120	0.0291***	age	-0.0030
	(0.0047)	(0.0135)	(0.0066)		(0.0047)
white	0.0170***	0.0063	0.0295***	intercept	3.0033***
	(0.0050)	(0.0138)	(0.0061)		(1.012)
age	-0.0029***	-0.0013	-0.0034***		
	(0.0003)	(0.0008)	(0.0004)		
exper	0.0010***	-0.0002	0.0007**		
	(0.0002)	(0.0006)	(0.0003)		
adjunct	0.0153*	-0.0195	0.0335***		
	(0.0060)	(0.0138)	(0.0092)		
pctmajor	-0.0140	0.0067	0.0079		
	(0.0199)	(0.0256)	(0.0215)		
pctfemale	-0.0040	-0.0114	-0.0023		
	(0.0154)	(0.0248)	(0.0219)		
response	-0.0128	-0.0189	-0.0274		
	(0.0142)	(0.0234)	(0.0207)		
size	-0.0000	-0.0001	0.0002		
	(0.0001)	(0.0001)	(0.0001)		
prncpl-macro	0.0040				
	(0.0048)				
prncplt		0.0086*	-0.0092**		
		(0.0041)	(0.0039)		
intermediate	0.0121	-0.0121*	-0.0114**		
	(0.0084)	(0.0054)	(0.0045)		
elective3	0.0313**	0.0071	0.0148**		
	(0.0116)	(0.0103)	(0.0071)		
elective4a	0.0471***	0.0168*	0.0167***		
	(0.0132)	(0.0080)	(0.0062)		
elective4b	0.0491***	0.0115	0.0184**		
	(0.0144)	(0.0101)	(0.0074)		
quantitative	0.0511**	0.0138	0.0269**		
	(0.0198)	(0.0161)	(0.0127)		
intercept	1.1091***	1.0458***	1.2306***		
	(0.0308)	(0.0605)	(0.053)		
sample size	981	465	465		
$\hat{\sigma}_u$	0.1284***	0.2513		$\hat{\sigma}^2 = \hat{\sigma}_u^2 + \hat{\sigma}_v^2$	0.0363**
$\hat{\sigma}_v$	0.0179***	0.0464***		$\hat{\gamma} = \hat{\sigma}_u^2 / \hat{\sigma}_v^2$	0.9852***
$\hat{\mu}$		-0.6531			
$\hat{\eta}$		0.0391***			
χ^2	321.7444	94.4174			

Note: standard errors in parentheses; *** for p-value ≤ 0.01, ** for p-value ≤ 0.05, * for p-value ≤ 0.10

Table 3: Average Technical Efficiencies for Tenure-track Faculty Members

Instructor	Gender	Race	T_i	$\overline{\hat{TE}}_{i,BC95}$	$\overline{\hat{TE}}_{i,BC92}$	$\overline{\hat{TE}}_{i,M1}$	$\overline{eval}_i$
75	F	W	6	0.9810	0.9785	0.9599	3.6611
49	M	W	12	0.9801	0.9867	0.9554	3.6222
18	F	W	6	0.9792	0.9837	0.9679	3.6402
25	M	NW	3	0.9784	0.9723	0.9525	3.4928
8	M	W	7	0.9720	0.9832	0.9484	3.6692
35	M	W	22	0.9706	0.9823	0.9530	3.6924
10	M	W	23	0.9669	0.9795	0.9472	3.6005
69	F	W	22	0.9602	0.9513	0.9243	3.5679
37	M	NW	23	0.9574	0.9275	0.9066	3.4500
42	M	W	24	0.9514	0.9533	0.9186	3.5065
67	M	W	8	0.9514	0.9312	0.9053	3.3590
23	M	W	19	0.9472	0.9396	0.9198	3.5237
16	M	NW	4	0.9464	0.9025	0.8833	3.3083
21	F	W	3	0.9256	0.8762	0.8511	3.2204
55	M	W	19	0.9132	0.9169	0.8967	3.3165
30	M	NW	12	0.8937	0.8904	0.8679	3.1628

Note: "—" represents the average over time. For example, $\overline{eval}_i = T_i^{-1}\sum_{t=1}^{T_i} eval_{it}$

Figure 1: SET Scores for Top 3 and Bottom 3 Instructors

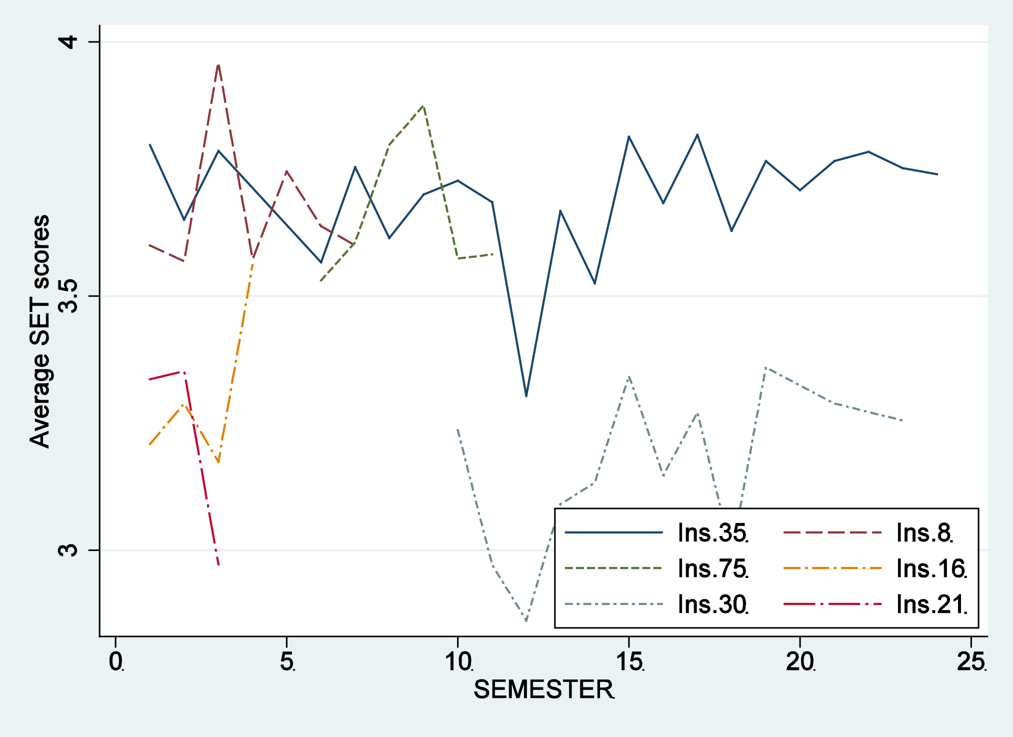
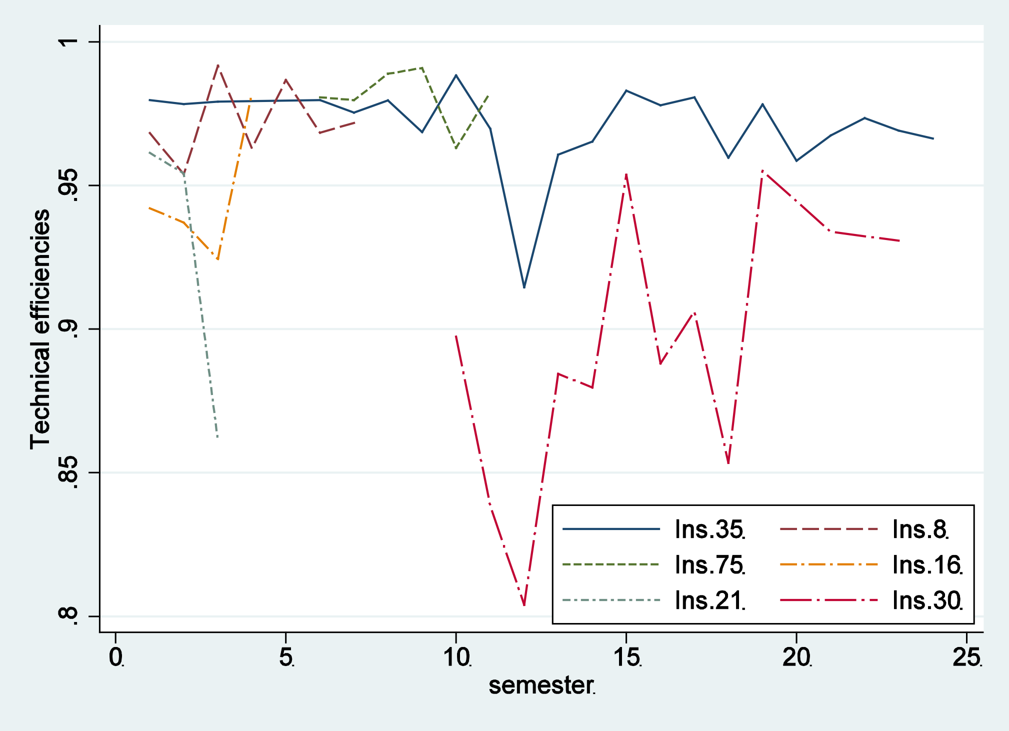


Figure 2: Time-Varying Technical Efficiencies (BC95) for Top 3 and Bottom 3 Instructors


Technical efficiencies
semester
Ins.35
Ins.8
Ins.75
Ins.16
Ins.21
Ins.30