engineering proceedings

MDPI

Proceeding Paper

A Negative Binomial Regression Model of Student Absenteeism in the Principles of Microeconomics †

Chen Chen [1], Shih-Shen Chen [1,*] and Chin W. Yang [2]

[1] College of Economics and Management, Fuzhou University of International Studies and Trade, Fuzhou 350200, China; chenchen@fzfu.edu.cn
[2] Department of Economics, Clarion University of Pennsylvania, Clarion, PA 16214, USA; cwyang@clarion.edu
* Correspondence: chensishen@fzfu.edu.cn
† Presented at the 3rd IEEE International Conference on Electronic Communications, Internet of Things and Big Data Conference 2023, Taichung, Taiwan, 14–16 April 2023.

Abstract: We investigated the determinants of absenteeism in the Principles of Microeconomics at a rural state university in Pennsylvania. We apply Poisson statistics and a negative binomial of both Type 1 and 2 models to estimate absenteeism behavior. All cases have GPA as a significant factor. A higher GPA reduces the absence rate by 3.2574 times in the NB1 model. It was found that business majors cut classes significantly less (26.24%) than education majors.

Keywords: attendance; economics education; learning performance factors; academic achievement; university outcome assessment; regression (statistics)

Citation: Chen, C.; Chen, S.-S.; Yang, C.W. A Negative Binomial Regression Model of Student Absenteeism in the Principles of Microeconomics. *Eng. Proc.* **2023**, *38*, 82. https://doi.org/10.3390/engproc2023038082

Academic Editors: Teen-Hang Meen, Hsin-Hung Lin and Cheng-Fu Yang

Published: 14 July 2023

1. Introduction

Absenteeism is a serious problem for universities and often leads to lower grades and hence, a high dropout rate. In a state-supported university, increasing rates of class absenteeism have been observed. Smaller classes may not be a problem for a typical freshman when the grades of popular education courses are good to obtain. A student cruises through less challenging courses without much effort in the wake of ubiquitous grade inflation due to pressure from the administration. One of the challenging courses for a sophomore business major is the Principles of Microeconomics, in which absenteeism is detrimental to a student's grade. Those who excessively decrease classes end up obtaining a letter grade of D or E. The subsequent "repeat of the course" or "take it at a community college during summer" attitudes give rise to an uneasy feeling in the Department of Economics. Before implementing any policy regarding absenteeism, it is important to analyze student characteristics concerning class attendance. Studies on college student absenteeism are scarce, as class attendance is generally not mandatory. Despite the proliferation of the applications of count regression (e.g., ship accidents by [1]; doctor visits by [2]), there seems to be a lack of application to economic and business education. An interesting application is found in the study who employed a Probit model in relation to Principles of Finance classes [3]. Another found a significant but small impact of class attendance on the performance of an introductory statistics class [4].

The most cited study is reported an average absenteeism rate at three major US universities of 1/3, which was comparable to that reported by Rogers and Rogers (2003) in Intermediate Microeconomics at an Australian University [5]. Romer found a positive and significant attendance–performance relationship. Before this, Schmidt (1983) identified a positive correlation between exam performance and time spent in lectures of an Information Processing class [6]. Park and Kerr (1990) also detected an inverse relationship between absenteeism and course grade in a Money and Banking course [7]. In a similar vein, Durden and Ellis (1995) indicated that attendance had a large and significant effect on performance when absenteeism is excessive [8]. Devadoss and Foltz (1996) employed a large sample

(n = 400) in an Agricultural Economics class and found that perfect attendees had higher average scores than those who cut classes by more than 50% [9].

Marburger (2001) identified a significant correspondence between incorrectly answered questions and missed class on a given date [10]. Kirby and McElroy (2003) examined the determinants of class attendance (n = 368) and found that working hours and travel time had a positive but diminishing marginal effect on course grade. In his detailed model [11], Stanca (2006) found neither a proxy variable nor instrument variable approach that offered a satisfactory solution to the omitted variable in terms of estimating the effect of class attendance on course performance [12]. He proposed a panel data approach (fixed-effect model) to control for the unobservable characteristics of class attendance. Even though class attendance impacted course performance positively and significantly, the magnitude was small.

Chen and Lin (2008) applied a randomized experiment approach to a Public Finance course and found between a 9.4 and 18% improvement in exam performance on average due to attending lectures [13]. The randomized design has the advantage of providing treatments (cause) and responses (effect) understood by social scientists. It also has non-response and selection bias for the experiment. Arulampalam et al. (2012) provided a quantile regression model in an attempt to differentiate the unobserved characteristics relating to class attendance [14]. They found that students in upper quantiles prevailed in terms of the negative absenteeism–performance relationship. When students in UK universities cannot understand the lecture, the difference in learning outcome, regardless of attending class or not, is small.

The above-mentioned result has a common problem. Absenteeism is rarely exogenous despite the efforts to alleviate the instrument, proxy variable approach, panel data, or quantile regression. To solve the problem, Self (2012) applied a negative binomial model to a Principles of Macroeconomics course at a Midwestern university [15]. The results indicated certain factors, including seniors, morning classes, and male students. An attendance policy has a significant impact on the absence rate. However, the model specification test did not directly estimate the impact of GPA or other variables on the absence rate. In this paper, we employ three types of count regressions: Poisson, negative binomial Type I, and Type II to focus on the impact of different majors and GPA on the absence rate.

2. Data and Methodology

To reduce the variations due to faculty- and text-specific factors, we employed a data set from one of the authors who had taken attendance in Principles of Microeconomics (or ECON 212) and used the same text as that of Clarion University, one of fourteen state-supported university systems in Pennsylvania. The school of business has been accredited by the AACSB since 1998, with an enrollment of approximately 650, about 1/10 of the university's enrollment. It serves the educational needs of residents who are largely from the region. We used the data of 604 students, computerized student transcript records of Clarion University from a course in the sample period to minimize exogenous factors consistently. Only those students who received final grades in ECON 212 were included in the sample.

For the number of absences in ECON 212, count regressions are the ideal analysis method. That is, the rate of absence is no longer assumed to be exogenous. The most commonly used count regression models are Poisson and negative binomial types [7]. We first considered the Poisson regression model in which y_i (the number of absences) is drawn from a Poisson distribution.

$$P(Y_i = y_i/\mathbf{x_i}) = e^{-\lambda_i}\lambda_i^{y_i}/y_i! \tag{1}$$

where the expected number of events (absence) given $\mathbf{x_i}$ is λ_i. The most common form for λ_i is as follows:

$$E(y_i/\mathbf{x_i}) = \lambda_i = e^{\mathbf{x_i'}\boldsymbol{\beta}} \tag{2}$$

$$\text{or } \ln \lambda_i = \mathbf{x_i'}\, \boldsymbol{\beta} \tag{3}$$

Using Equation (3), we formulate the Poisson model as below.

$$\ln E(y_i/\mathbf{x_i}) = \ln \lambda_i = \mathbf{x_i'}\, \boldsymbol{\beta} \tag{4}$$

where $\mathbf{x_i}$ contains GPA, GENDER, MAJOR1, MAJOR2, and MAJOR3. GPA is a continuous variable on a 4-0 scale. GENDER = 0 for male students, and it is 1 for female students. MAJOR1 = 1 for business majors, and it is zero otherwise. MAJOR2 = 1 for arts and science majors, and it is zero otherwise. MAJOR3 = 1 for business information majors, and it is zero otherwise. The reference group for the three MAJOR dummies is education majors.

Given that the log-likelihood function is concave and that rapid convergence is achieved, the Poisson regression model is computationally efficient with the asymptotic covariance for parameter estimates of

$$\left[\sum_{\mathrm{i}} \hat{\mu}_i \mathrm{x_i} \mathbf{x_i'}\right]^{-1} \tag{5}$$

where $\hat{\mu}_i$ is the expected value of the Poisson distribution.

Thus, it has become a popular choice in modeling count regressions in accident analysis and the number of times a patient visits a doctor. A Poisson model has a restrictive assumption in which the average and variance of y_i must be the same. When the variance exceeds the average, a case of overdispersion occurs. Cameron and Trivedi (1998) suggested the use of negative binomial (NB) models for robust estimators and bootstrap approaches. Table 1 suggests the following results based on the Poisson regression model. First, GPA is negatively related to absenteeism, with a p-value of 0.000. Second, different from Self (2012), gender is not a significant determinant in explaining absenteeism (p = 0.275) [15]. In general, female students tend to cut fewer classes than male students ($\hat{\beta}_2 = -0.078$). Third, the students of a business major cut classes significantly less than those of education (p = 0.000) and arts and sciences majors (p = 0.107).

Table 1. Estimated results of the Poisson regression model.

Variable	Estimate	Standard Error	t-Statistic	*p*-Value
C	3.54371	0.158480	22.3606	[0.000]
GPA	−0.639844	0.057290	−11.1684	[0.000]
GENDER	−0.077989	0.071515	−1.09052	[0.275]
MAJOR1	−0.333574	0.087194	−3.82565	[0.000]
MAJOR2	−0.135093	0.083797	−1.61215	[0.107]
MAJOR3	0.069834	0.121339	0.575528	[0.565]

n = 604, mean of dependent variable $\hat{\lambda}$ = 5.09768; standard deviation of dependent variable = 4.66226; Log-likelihood = −487.896.

The negative binomial model type 1 (NB1) assumes the variance w_i is a linear function or a multiple of the mean λ_i or

$$w_i = (1+\alpha)\lambda_i \tag{6}$$

The negative binomial model type 2 (NB2) assumes the following quadratic relation.

$$w_i = \lambda_i + \alpha\, \lambda_i^2 \tag{7}$$

As both NB1 and NB2 are considered to be a generalization of the Poisson model and are available in many econometrics, a common practice is to run the Poisson, NB1, and NB2 models before choosing the optimal one. Tables 1–3 report the estimated results for the three models.

Table 2. Estimated results of the NB1 regression model.

Variable	Estimate	Standard Error	t-Statistic	*p*-Value
C	3.57134	0.162156	22.0241	[0.000]
GPA	−0.650996	0.056559	−11.5100	[0.000]
GENDER	−0.084488	0.067383	−1.25385	[0.210]
MAJOR1	−0.304356	0.083130	−3.66121	[0.000]
MAJOR2	−0.134290	0.082046	−1.63677	[0.102]
MAJOR3	0.042738	0.111408	0.383616	[0.701]
ALPHA α	2.75437	0.250689	10.9872	[0.000]

n = 604, $\hat{\lambda} = 5.09768$, standard deviation of dependent variable = 4.66226, Log-likelihood = -1561.95.

Table 3. Estimated results of the NB2 regression model.

Variable	Estimate	Standard Error	t-Statistic	*p*-Value
C	3.72674	0.196303	18.9846	[0.000]
GPA	−0.710946	0.066194	−10.7403	[0.000]
GENDER	−0.080952	0.073221	−1.10558	[0.269]
MAJOR1	−0.336069	0.089712	−3.74609	[0.000]
MAJOR2	−0.090736	0.091299	−0.993831	[0.320]
MAJOR3	0.090447	0.127622	0.708713	[0.479]
ALPHA α	0.542145	0.046782	11.5888	[0.000]

n = 604, $\hat{\lambda} = 5.09768$, standard deviation of dependent variable = 4.66226, Log-likelihood = -1571.54.

3. Results and Discussions

The results in Table 1 suggest that the average and variance of the dependent variable (numbers of absence in ECON 212) differ significantly. The average of $\hat{\lambda}$ = 5.09768 is much less than variance $\hat{w}$ = 21.7367. The larger variance suggests that several students in the Economics Department reduce classes excessively. Grade inflation allows a student to maintain a GPA of 2.0 to stay in university. The overdispersion is analogous to heteroscedasticity in the OLS model. For a known variance specification, we correct the problem by using the NB1 model, where the variance of the dependent variable is a multiple of its average. Table 2 shows that the results are similar to those of the Poisson model. GPA remains significant in the NB1 model (p = 0.000). The negative coefficient of -0.651 indicates that GPA and the number of absences is significantly and inversely related. That is, the higher the GPA, the less class absence. The magnitude of the relationship for an average student can be derived from Equation (2) or

$$\frac{\partial E(y_i/\mathbf{x_i})}{\partial x_i} = \lambda_i \hat{\beta}_i = 5.09768 * (-0.639) = -3.2574 \quad (8)$$

If GPA is increased by one unit, the expected number of skipped classes is reduced by 3.2574. A student with GPA = 2 cuts 3.2574 classes more than a student with GPA = 3. Missing classes for over one week and a half have a noticeable impact on performance in Principles of Microeconomics.

Female students have a decreased tendency to reduce classes (p = 0.21). The same trend is also observed in the Poisson model (p = 0.275). We surmise that female students at the university are more conscientious in the Principles of Microeconomics course than male students. In the Poisson model, MAJOR1 (business major) is significant in the NB1 model. A business major has less tendency to reduce classes than the reference group (education majors) (p = 0.00). It means that students of an education major feel difficulty and lose interest in the Principles of Microeconomics. The same trend is observed for the students of arts and science majors. The students of business information science seem to have a similar trend as those of education majors in terms of absenteeism.

The difference in the expected rate of absence between majors is evaluated by using Equations (2) and (6).

$$\frac{\lambda_i(\text{MAJOR1} = 1)}{\lambda_i(\text{MAJOR1} = 0)} = \frac{e^{\mathbf{x}'\boldsymbol{\beta}(\text{MAJOR1}=1)}}{e^{\mathbf{x}'\boldsymbol{\beta}(\text{MAJOR1}=0)}} = \frac{e^{\mathbf{x}'\boldsymbol{\beta}-0.304356}}{e^{\mathbf{x}'\boldsymbol{\beta}}} = \frac{1}{e^{0.304356}} = 0.7376 \tag{9}$$

where -0.304356 (Table 2 of the NB1 model) is the estimated coefficient on the dummy variable MAJOR1. In total, 73.76% of the students of business majors did not take Principles of Microeconomics, while those of education majors take 26.24%.

Similar for arts and sciences majors, the ratio of the absence rate is as follows:

$$\frac{\lambda_i(\text{MAJOR2} = 1)}{\lambda_i(\text{MAJOR2} = 0)} = \frac{1}{e^{0.13429}} = 0.8743 \tag{10}$$

The absenteeism rate for the students of arts and sciences majors is 87.43% of those of education majors.

The overdispersion is tested with the estimated coefficient on α (Wald test). In the case of the NB1 model, a p-value of 0.000 leads to the rejection of $H_0 : \alpha = 0$. In addition, the value of α is 2.75437 implies considerable overdispersion (Cameron & Trivedi, 1998, pp 78–79). It is tested by using the LR test, and the result shows the difference in the fitted log-likelihoods of the Poisson and NB1 or $-2 * (-1561.95 + 487.896) = 2148.1$, which far exceeds the critical value of x^2. Both test results indicate that the overdispersion problem precludes the Poisson regression model due to a small group of outliners (excessively cutting classes).

Table 3 presents that GPA is a factor in explaining the absenteeism problem but remains an insignificant predictor. MAJOR1 remains the same as in NB1. MAJOR2 becomes significant (p-value of 32%), and MAJOR3 remains insignificant. The Wald test α indicates the NB2 is appropriate for the Poisson model in terms of the overdispersion problem. The choice between the two negative binomial models is important for the LR test as the Wald test gives indistinguishable results. Since the log-likelihood function of the NB1 model gives a slightly larger value than that of the NB2 model ($-1561.95 > -1571.54$), we opt for the NB1 model in analyzing the student absenteeism problem at this university.

4. Concluding Remark

Decreasing the attrition rate in public universities is of primary concern to their survival. Declining enrollment, especially in rural areas, causes many problems. A proper policy cannot be effectively implemented without understanding the factors that affect absenteeism. As a majority of college professors do not take attendance, the issue remains unresolved. We applied the Poisson, NB1, and NB2 regression models to analyze the problem. All of the models indicate that GPA is inversely related to absenteeism in Principles of Microeconomics classes. In particular, one unit of GPA is expected to reduce absences by 3.2574 in the NB1 model. In the NB1 model, the students of business majors reduce 26.24% more classes than those of education majors. Due to an overdispersion problem, the Poisson model is inappropriate. A large variance in the number of absences exceeds its corresponding average (21.7367 > 5.09768). The overdispersion problem leads to the question of whether one should implement a mandatory attendance policy. Low-students become disruptive and hence jeopardize the overall welfare of the entire class. The students ought to have the freedom to weigh the positives and negatives between the marginal gain in course performance and attending the class. The overdispersion (excessive cut) problem is found in the NB1 and NB2 models. The LR test indicates that the NB1 model has a slight advantage over the NB2 model. The explanatory variables that determine the number of cut classes are GPA, MAJOR1, and MAJOR2 (marginally significant). Female students tend to cancel the Principles of Microeconomics noticeably. The efforts should thus be focused on those with low GPAs in business information science and education majors.

Author Contributions: Conceptualization, S.-S.C. and C.W.Y.; methodology, C.C.; software, C.C.; validation, S.-S.C.; formal analysis, S.-S.C.; investigation, C.C.; resources, C.W.Y.; data curation, C.W.Y.; writing—original draft preparation, C.C.; writing—review and editing, S.-S.C.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McCullagh, P.; Nelder, J. *Generalized Linear Models*; Chapman and Hall: New York, NY, USA, 1983.
2. Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*; Cambridge University Press: Cambridge, UK, 1998. [CrossRef]
3. Chan, K.; Shum, C.; Wright, J. Class attendance and student performance in Principles of Finance. *Financ. Pract. Educ.* **1997**, *7*, 58–65.
4. Rodgers, J.R. A Panel-Data Study of the Effect of Student Attendance on University Performance. *Aust. J. Educ.* **2001**, *45*, 284–295. [CrossRef]
5. Romer, D. Do students go to class? Should they? *J. Econ. Perspect.* **1993**, *7*, 167–174. [CrossRef]
6. Schmidt, R.M. Who maximizes what? A study in student time allocation. *Am. Econ. Rev.* **1983**, *73*, 23–28.
7. Park, H.; Kerr, P. Determinants of academic performance: A multinomial logit approach. *J. Econ. Educ.* **1990**, *21*, 101–111. [CrossRef]
8. Durden, G.C.; Ellis, L.V. The effect of attendance on student learning in Principles of Economics. *Am. Econ. Rev.* **1995**, *85*, 343–346.
9. Devadoss, S.; Foltz, J. Evaluation of Factors Influencing Student Class Attendance and Performance. *Am. J. Agric. Econ.* **1996**, *78*, 499–507. [CrossRef]
10. Marburger, D.R. Absenteeism and undergraduate exam performance. *J. Econ. Educ.* **2001**, *32*, 99–109. [CrossRef]
11. Kirby, A.; McElroy, B. The effect of attendance on grade for first year Economics students in University College Cork. *Econ. Soc. Rev.* **2003**, *34*, 311–326.
12. Stanca, L. The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics. *J. Econ. Educ.* **2006**, *37*, 251–266. [CrossRef]
13. Chen, J.; Lin, T.-F. Class Attendance and Exam Performance: A Randomized Experiment. *J. Econ. Educ.* **2008**, *39*, 213–227. [CrossRef]
14. Arulampalam, W.; Naylor, R.A.; Smith, J. Am I missing something? The effects of absence from class on student performance. *Econ. Educ. Rev.* **2012**, *31*, 363–375. [CrossRef]
15. Self, S. Studying absenteeism in Principles of Macroeconomics: Do attendance policies make a difference? *J. Econ. Educ.* **2012**, *43*, 223–234. [CrossRef]