



Comment on Chen and Lin ‘Does Downloading Power-Point Slides Before the Lecture Lead to Better Student Achievement?’

Edmund Cannon

Abstract

Chen and Lin (2008) is an interesting attempt to measure the effect of students having access to PowerPoint slides before a lecture. They argue that such access leads to better learning outcomes as measured by higher exam marks. While their empirical results shed some light on students’ use of PowerPoint slides, I argue that they have not succeeded in isolating the treatment effect and the correlations cannot be interpreted as causal.

1. Introduction

Chen and Lin (2008a) is an interesting attempt to measure the benefits of a new teaching technology, namely the effect of downloading a lecture’s PowerPoint slides by students. To do this the authors report the correlation of downloading PowerPoint slides with exam performance. Their data set allows them to control for student and question fixed effects and they claim that this allows them to remove endogeneity bias and hence identify the treatment effect. They conclude that downloading the relevant PowerPoint slides results in a student being 3.48 per cent more likely to answer a right or wrong question correctly.¹

In this comment I argue that there are good reasons to believe that the treatment effect has not been isolated. An extension of the OLS analysis suggests that there is virtually no benefit in going to lectures for a student who downloads PowerPoint lecture slides: this (strange) result contradicts other analysis of lecture attendance (e.g. Stanca, 2006; Chen and Lin, 2008b) and is corroboratory evidence that there may be something wrong in using OLS. I suggest a possible test for endogeneity in this context and show that the null hypothesis of exogeneity is rejected. I speculate that the endogeneity of regressors then explains the OLS results.

In the next section I discuss formally the issue of endogeneity and suggest a test for its presence in this context. Then I re-analyse Chen and Lin’s data which they have kindly provided. The final section discusses the results and concludes.

¹ The parameter estimate on the indicator for downloading the PowerPoint slides is 0.0348 and most questions receive marks of either zero or one (some questions receive marks in between as I discuss below).

2. Analysis

A detailed description of Chen and Lin's data can be found in the original article and I do not repeat all of that information here. They have information on a variety of variables about each student (prior exam attainment, gender, family background) that are not used in the regression analysis of Chen and Lin (2008a). The reason for this is that student fixed effects are used in the analysis and hence the other variables could not be included since they would be perfectly collinear with the fixed effects. Since the information on the students is not really needed for this comment and since the authors are concerned about anonymity I have not been provided with these data.

The data which I have re-analysed consists of information on lecture attendance, PowerPoint slide download and exam performance at question level. There are a total of 126 students each of whom has been allocated marks for up to 40 questions.² For each question data are available on whether the student attended the lecture in which the material was taught and also whether the student downloaded the associated PowerPoint slide.

Changing Chen and Lin's notation slightly, I use $p_{ij} \in \{0,1\}$ to denote the binary variable showing that the PowerPoint slides have been downloaded by student i for the lecture corresponding to question j . Similarly I use $l_{ij} \in \{0,1\}$ to denote lecture attendance. The mark given to the question is denoted Y_{ij} and this can take on 18 different values, ranging from zero to one. Of the 5005 observations in the full data set, 1375 (27 per cent) of the marks are zero and 2573 (51 per cent) are one, so only 1057 (21 per cent) take on the intermediate values. For this reason, the ideal specification strategy would use ordered probit or censored regression. However, I was unable to obtain convergence using these estimation procedures and so my analysis follows Chen and Lin in using a linear probability model.

The specification that Chen and Lin estimate is

$$(1) \quad Y_{ij} = \theta p_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij} \quad \varepsilon_{ij} \perp \{p_{ij}, \alpha_i, \gamma_j\}.$$

Although this is followed by $i = 1, \dots, I, \quad j = 1, \dots, J$, the text makes it clear that the regression is estimated not for all $IJ = 5006$ observations but only for the 3675 observations where the student attended the lecture $l_{ij} = 1$. This specification can be contrasted with a more general specification

$$(2) \quad Y_{ij} = \theta p_{ij} + \phi l_{ij} + \psi p_{ij} l_{ij} + \lambda_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij} \quad \varepsilon_{ij} \perp \{p_{ij}, l_{ij}, p_{ij} l_{ij}, \lambda_{ij}, \alpha_i, \gamma_j\}$$

which can only be estimated for the full sample. In this specification I have allowed explicitly that lecture attendance can have an effect on exam performance (through the parameter ϕ) and that the effect of downloading PowerPoint slides depends on lecture attendance: the effect of downloading PowerPoint slides for those who attend lectures is $\theta + \psi$, whereas for those who are absent it is only θ .

I have also introduced a new explanatory variable λ_{ij} which is the unobserved question- j specific determinant for student i . There are several possible interpretations of this:

² This suggests that there should be 5040 observations but in fact there are only 5005, since not all students have marks for all 40 questions. The 126 student codes range from 2 to 150, but I was not told whether another 24 students were missing.

- (i) A student's prior ability varies *horizontally* across topics. Example: a student is innately good at interpreting supply and demand curves which are taught graphically, but struggles with price indices which are taught algebraically.
- (ii) A student's ability to answer questions depends *vertically* on the interaction of the difficulty of the question and the ability of the student. Example: there is an easy question which all students answer correctly, and a difficult question which only clever students can answer correctly.
- (iii) A student's allocation of effort between topics is chosen endogenously. Example: a student decides to skip one particular topic to save time and concentrate on studying something else.
- (iv) The student's total effort is chosen endogenously. Example: a student can either compensate for missing a lecture by working harder on the topic in private study or can take the time as leisure.

Either individually or in combination these interpretations of λ_{ij} suggest that it may be correlated with p_{ij} or l_{ij} , which are themselves correlated with each other, and hence OLS estimation of

$$(3) \quad \begin{aligned} Y_{ij} &= \theta p_{ij} + \phi l_{ij} + \psi p_{ij} l_{ij} + \alpha_i + \gamma_j + \xi_{ij} \\ \xi_{ij} &\equiv \varepsilon_{ij} + \lambda_{ij} \end{aligned}$$

will yield inconsistent estimates of the parameters θ , ϕ and ψ . In principle it might be possible to design tests which deal with the issues in (i) and (ii), but collecting information on students' total effort and allocation of effort is likely to be impossible. In any event, such information is not available in the data set used here.

What are the implications of using Chen and Lin's strategy of estimating η from the sub-sample of the data where $l_{ij} = 1$? If λ_{ij} were in fact orthogonal to the other regressors, then the parameter estimates should be consistent. In a conventional panel with small J , the consistency result would be of little help for the cross-section fixed effects, but here we have $J = 40$, which is just large enough to suggest that we might obtain useable parameter estimates for α_i . Then I estimate Chen and Lin's model twice, once for those observations where the student attended the relevant lecture and once where they did not, so that:

$$(4) \quad \begin{aligned} Y_{ij} &= \theta p_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij} & l_{ij} &= 1 \\ Y_{ij} &= \theta' p_{ij} + \alpha'_i + \gamma'_j + \varepsilon_{ij} & l_{ij} &= 0 \end{aligned}$$

where the first equation is Chen and Lin's model (and just repeats equation 1). If attendance at the lecture is genuinely exogenous then it should be uncorrelated with other explanatory variables. In particular whether or not a student attends should be uncorrelated with that student's ability and hence

$$(5) \quad \alpha_i = \alpha'_i \quad \gamma_j = \gamma'_j$$

Working on the slightly dubious assumption that $J = 40$ is a sufficiently large data set to obtain reliable estimates for the fixed effects, this can be tested with a Chow test to determine the validity of the assumption that $\xi_{ij} \perp \{p_{ij}, l_{ij}, p_{ij} l_{ij}, \lambda_{ij}, \alpha_i, \gamma_j\}$. Alternatively one might like to test whether

downloading PowerPoint slides is correlated with the fixed effects and hence consider an analogous Chow test splitting the data according to

$$(6) \quad \begin{aligned} Y_{ij} &= \phi l_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij} & p_{ij} &= 1 \\ Y_{ij} &= \phi' l_{ij} + \alpha'_i + \gamma'_j + \varepsilon_{ij} & p_{ij} &= 0 \end{aligned}$$

and conducting the same test as in (5).

The final consideration is the calculation of the standard errors. The exam taken by the students had 40 questions but there were only 12 lectures. Clearly this means that the material for more than one question was covered in each lecture, so on average about three or four questions belong to each lecture. So one would expect the disturbances for each group of questions to be correlated and hence the standard errors should be adjusted to reflect this. To complicate matters, the material for some questions was covered in more than one lecture, so it is necessary to separate questions covered in lecture 1, questions covered in lecture 2 and questions covered in both lectures 1 and 2. Using the underlying data sent to me by Chen and Lin I have attempted to do this and have identified 23 groups of questions (many of which only contain one question). With this many groups and large numbers of fixed effects the degrees of freedom are reduced considerably for some tests and so I report tests using both conventional and clustered standard errors. Note that the easiest way to implement the test in (5) with clustered standard errors in most software packages is to estimate

$$(7) \quad Y_{ij} = \theta p_{ij} + \phi l_{ij} + \psi p_{ij} l_{ij} + \alpha_i + (\alpha'_i - \alpha_i) l_i + \gamma_j + (\gamma'_j - \gamma_j) l_i + \xi_{ij}$$

and test $\widehat{\alpha'_i - \alpha_i} = \widehat{\gamma'_j - \gamma_j} = 0$, i.e. by testing the significance of the interaction of the student fixed-effects with the lecture attendance variable.

An obvious way to interpret the clusters is as topics. Clearly some students may be better at some topics than others and hence a further possibility is to consider fixed effects for each student on each topic. Denote the fixed effect for student i on topic k as δ_{ik} , $k = 1, \dots, 23$. Since the number of topics is fewer than the number of questions it might be possible to estimate the relationship

$$(8) \quad Y_{ij} = \theta p_{ij} + \phi l_{ij} + \psi p_{ij} l_{ij} + \delta_{ik} + \gamma_j + \xi_{ij}$$

In practice, however, I was unable to estimate this due to collinearity.

3. Results

Table 1 OLS estimates 1

	A	B	C
p_{ij}	0.0348 (0.0165) [0.0189]	0.0851 (0.0334) [0.0295]	0.0641 (0.0260) [0.0291]
l_{ij}			0.0228 (0.0186) [0.0149]
$p_{ij}l_{ij}$			-0.0194 (0.0296) [0.0316]
Sample used	$l_{ij} = 1$	$l_{ij} = 0$	Whole sample
Sample size	3,675	1,330	5,005

OLS regression results from regressions including student and question fixed effects. White's heteroskedasticity-robust standard errors in parentheses; standard errors also robust to clustering in square brackets.

Table 1 reports the results of my OLS regressions. Column A replicates the fixed-effects results of Chen and Lin (p. 16). However, when the standard errors are calculated correctly the t-statistic is only 1.84 resulting in statistical significance lower than conventionally used.

It is instructive to look at the results from using the data where students did not go to lectures, reported in column B. The effect here is more than twice as large. Taken at face value this suggests that students who do not go to lectures get much more benefit from downloading PowerPoint slides than students who do go to lectures.

Chen and Lin suggest two reasons why having PowerPoint slides may enhance students' understanding: it provides reference material in the lecture and it enhances note taking in the lecture. Clearly neither of these mechanisms are working when students are not in the lecture, so it is necessary to find alternative suggestions. One possibility is that students who do not attend lectures have difficulty reading textbooks – or indeed any conventional document – and find the summaries in PowerPoint slides a useful complement (or, more pessimistically, a useful substitute). However, to interpret the parameter estimates as reflecting this explanation alone would rely upon the joint combination of downloading PowerPoint slides and absence from lectures being exogenous.

Column C shows the results when the data are pooled. The standard errors of the parameter estimates for both lecture attendance and the interactive term are large: there is insufficient evidence that lecture attendance has any beneficial impact at all, since the test for joint significance of the two lecture variables is $F(2,22) = 1.34$ [$p = 0.28$]. The point estimate for the benefit of downloading slides conditional on attending the lecture is $\widehat{\theta + \psi} = 0.0433$ whereas the marginal effect of attending a lecture conditional on downloading the slides is $\widehat{\phi + \psi} = 0.0034$. Taken at face value the obvious conclusion from this is that lecture attendance is a waste of time: university administrators will be pleased to know that they can save money by scrapping costly lectures, selling the lecture rooms and relying upon students to download material from the web.

In the previous section I noted that testing the orthogonality of a student's ability to his or her exam performance is the same as testing $\widehat{\alpha'_i - \alpha_i} = \widehat{\gamma'_j - \gamma_j} = 0$ in either equations (4) or equation (6). Using tests based on clustering, these tests yield results of $F(22,22) = 24.91$ [$p = 0.000$] and $F(22,22) = 69.96$ [$p = 0.000$] respectively: the results are qualitatively the same with conventionally-calculated

standard errors or for tests of $\widehat{\alpha'_i - \alpha_i} = 0$ or $\widehat{\gamma'_j - \gamma_j} = 0$ individually. This suggests that unobserved variation is correlated with the explanators of interest and that the OLS estimates are inconsistent.

4. Conclusions

The arrival of technology such as PowerPoint and the web have excited teachers who believe that it may enhance student learning. To the extent that the latter is measured by exam performance this would mean that students exposed to new technology would get higher marks. However, one does not need to be a complete Luddite to be concerned that new technology might have negative effects which would attenuate or even overwhelm the benefits. One possibility is that students suffering from problems of motivation would attend lectures less when PowerPoint slides are available. And while there may be short-run benefits in students having structured PowerPoint notes available, their availability may reduce students' long-run progress in learning to make their own notes and reduce the benefits from lecture attendance by making the experience more passive. Macedo-Rouet *et al.* (2009) suggest that the benefits of new versus old technology depend upon the type of material being taught and this may explain why research hitherto has provided apparently inconsistent results.

Chen and Lin (2008) is an interesting attempt to measure the net benefits of students downloading PowerPoint slides. As always, the problem is identifying the treatment effect when there is unobserved heterogeneity. This issue is particularly acute because lecture attendance and PowerPoint download are endogenous choices of the students.

In this paper I have proposed a test for endogeneity and concluded that the data suggest that OLS is inconsistent. The OLS estimates themselves suggest that downloading PowerPoint slides has larger benefits for students who do not attend the lecture, so the data undermine the hypothesis that PowerPoint slides benefit students in lectures. Why do we obtain such strange OLS estimates?

Various narratives could be constructed that are consistent with this paradoxical result. Suppose that sometimes missing a lecture is an exogenous event (perhaps due to illness). Having missed a lecture for this reason, the students who are most likely to try to catch up on the work are those who are most hard-working (and this may correlate with other positive attributes). So the combination of missing a lecture and downloading the PowerPoint slides may merely indicate a 'good' student who will tend to do better anyway. But this line of argument relies on attendance being exogenous: probably many absences are endogenous choices of the students. Students' choices are partly irrational and partly rational – but with an objective that includes more than maximising exam mark. With such unobserved heterogeneity in both ability and behaviour, identifying treatment effects is very difficult.

Notes

This analysis was undertaken while I was visiting the University of Verona, whose financial support and hospitality are gratefully acknowledged. I thank the authors of the original paper for making data available. I have received useful comments from two anonymous referees. Any remaining errors are my own.

References

- Chen, J., and Lin, T-F. (2008a). 'Does downloading PowerPoint slides before the lecture lead to better student achievement?' *International Review of Economic Education*, Vol. 7(2), pp. 9–18.
- Chen, J., and Lin, T-F. (2008b) 'Class attendance and exam performance: a randomised experiment', *Journal of Economic Education*, Vol 39(3), pp.213–27.

Macedo-Rouet, M., Ney, M., Charles, S. and Lallich-Boidin, G. (2009). 'Students' performance and satisfaction with web versus paper-based practice quizzes and lecture notes', *Computers and Education*, Vol. 53, pp. 375–84.

Stanca, L. (2006). 'The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics', *Journal of Economic Education*, Vol. 37(3), pp. 251–66.

Author Biography

Edmund Cannon is Reader in Economics at the University of Bristol and a Deputy Director of the Economics Network. His PhD training at the University of Oxford was in applied economic growth and he continues to lecture this topic at the University of Verona: at Bristol he teaches econometrics and applied economics. His research is mainly on pensions and he is an expert on the UK annuity market.

Contact details

Edmund Cannon
University of Bristol,
School of Economics, Finance and Management,
8 Woodland Road,
Bristol.
BS8 1TN
edmund.cannon@bristol.ac.uk