“Educational Production”
Lazear (2001)*

Economics of Education (ECO383)
Fall 2010

This version: October 2, 2010

1 Preamble

This note provides a **Reader’s Guide** to Lazear’s interesting model of education production. In it, I want to explain the key features of the model clearly, in a way that does justice to Lazear’s own treatment and also draws attention to some of the relevant complications. If anything is unclear, I will be very happy to discuss further.

2 Overview

Lazear’s paper is noteworthy (for our purposes) in two respects:

- First, it provides an analytical model that links student disruption to class size. (Models are useful, generally speaking, in making potential causal mechanisms very explicit.) Specifically, it shows how optimal class size varies with an underlying ‘disruption’ parameter: more disruptive students leads the optimal class size to shrink. The mechanics of this we will see below.
- Second, the model helps explain (for reasons that I hope were clear from the class discussion) why one might not see the expected negative relation ‘in the data’ between smaller classes and higher student performance. (This is mentioned at the top of page 783 of Lazear’s paper.) Intuitively, where we see smaller classes, the students are more likely to be more difficult to educate, and this effect is likely to counteract any direct benefit of a smaller class on student learning.

*Note by Robert McMillan, email: mcmillan@chass.utoronto.ca.

3 The Model

Let's go over the key elements of the model. There are really two main pieces that we need to understand: the specification of the disruption function (and its implied properties), and the administrator's problem. Analysis of the administrator's problem will make clear the conditions needed for Proposition 1 – to the effect that optimal class size is increasing as students become less disruptive. So, in turn:

3.1 Specification of the disruption function

Suppose that individual students can be in one of two states: a disruptive state, or a non-disruptive state. We will parameterize the proportion of time a student is non-disruptive by p. Learning (as we will see below) is assumed to occur only in the non-disruptive state – that is, when *no* student is being disruptive.

Consider education over a single period, and assume (for simplicity) that all students in a school share the same non-disruption propensity, p.[1]

If there was just one student in a class, then for proportion p of the period, there would be no disruption, and for proportion $1 - p$ of the period, there would be disruption.

What happens to the proportion of time when there is absolutely *no* disruption in the classroom as the number of students, denoted by integer n, increases? It turns out that this is the case that is most relevant to Lazear's setup. Let us illustrate the general logic by considering a two-student example.

Two-student case:

Let us refer to the two students by the indices i and j. And assume that each student's propensity to be disruptive is independent of (i.e. unrelated to) the behaviour of the other student. Then four cases arise:

1. Both student i and student j are undisruptive.
2. Student i is undisruptive and student j is disruptive.
3. Student i is disruptive and student j is undisruptive.
4. Both student i and student j are disruptive.

[1] Clearly, realism would suggest considering a mixture of types of student in a given school, differing in their degree of disruptiveness.

Note that these four cases are mutually exclusive and exhaustive.

Take the first case – that both student i and student j are undisruptive. Given that behaviour is independent, the corresponding event occurs with probability equal to the proportion of the time student i is undisruptive times the proportion of time student j is undisruptive, which is equal to a proportion $p \times p = p^2$ of the time.

The second and third cases each occur a proportion $p(1-p)$ of the time. Thus the second and third cases *together* occur a proportion $2p(1-p)$ of the time. The fourth case – that both students are disruptive – occurs a proportion $(1-p) \times (1-p)$ of the time, or $(1-p)^2$.

[Aside: As a check that we have the right answer, the proportions associated with each of the four exhaustive cases should sum to 1, which they do.]

In sum: the proportion of time when no student is disruptive is given by p^n; and the proportion of time that *some* disruption is occurring is given by the sum of the proportion of time associated with the second, third and fourth cases above. Given the exclusive and exhaustive nature of the options, this is one minus the proportion of time that no student is disruptive, or $1-p^2$ of the time.

$n-$student case:

Suppose we now extend consideration to the $n-$student case: here, the proportion of time that no disruption occurs is equal to p^n, and the proportion of time that some disruption occurs is equal to $1-p^n$. This gives us the first piece we need: the model tells us how much totally undisrupted time (as well as partially and totally disrupted time) occurs, given p and n.

Graphical representation – an exercise

We will see that the focus in Lazear's model is on completely *undisrupted* time, on the basis that this is assumed necessary for learning to occur.[2] Let undisrupted time be denoted by Q. We have established that undisrupted time

$$Q(p,n) = p^n. \tag{1}$$

We can easily represent this function, which maps p and n into Q, noting that the relevant ranges of p and n are: $0 \leq p \leq 1$ (given p is a proportion) and $n \geq 1$.

[2]This is admittedly an extreme rendering. We might think that more learning might occur if just one student was being disruptive out of a class of $n >> 1$ students than if all n were being disruptive. But that's not how things are developed in the paper.

Exercise: Plot $Q = p^n$ with p on the horizontal axis taking values ranging between 0 and 1 (say, 0.1, 0.2, 0.3 etc.), considering the effects of different integer values of $n \geq 1$. Three features of your plot should become apparent:

- First, you can see that, for a given n, as p rises, so the proportion of time that no disruption occurs rises. No mystery to this: p measures the propensity *not* to be disruptive for the typical student, so if this goes up, we'd expect disruption to fall.
- Second, for a given p, an increase in class size n *reduces* the proportion of time in the classroom that is undisrupted (assuming all students are the same). So bigger classrooms, for a *given* p, are associated with more disruption, except in the one case where $p = 0$, in which there is total mayhem all the time, regardless of n.
- Third, note that the effect of increased class size on the total quantity of non-disruption (which we have called Q) rises then falls in p. Mathematically, this is summarized by the cross-partial derivative of Q with respect to p and n. This fact makes things less clean, analytically, than they could be.

Quiz questions:

1. If $n = 3$, what is the proportion of time in which at least two students are disruptive?
2. Calculate the cross-partial derivative $\frac{\partial^2 Q}{\partial p \partial n}$. (Hint: $\partial/\partial n(p^n) = p^n ln(p)$, applying the chain rule.[3]

Quiz solutions (and more):

1. If $n = 3$, the proportion of time in which at least two students are disruptive $=$ $3p(1-p)^2 + (1-p)$. That is, there are three possible cases where different pairs of students are disruptive and one isn't, each with probability $p(1-p)^2$ and one case where all are disruptive, with probability p^3.

2. We can derive the analytical properties of this $Q(p, n)$ 'undisrupted time' function.

First, what is $\frac{\partial Q}{\partial p}$?

$$\frac{\partial Q}{\partial p} = p^n n/p > 0, \tag{2}$$

[3]Note that, definitionally, $p \equiv e^{ln(p)}$, which is just to say that "p is identical to e raised to the power by which e needs to be raised in order to equal p." Then $p^n = e^l n(p)^n = e^{nln(p)}$. Thus $(\partial/\partial n)(p^n) = (\partial/\partial n)e^{nln(p)}$ and recall that $(\partial/\partial n)e^{f(n)} = e^{f(n)}(\partial/\partial n)f(n)$ by the chain rule, where $f(n)$ is some general function of n. Letting $f(n) = nln(p)$, we have the result.

using the fact that $Q = e^{nln(p)}$, and applying the chain rule.

Second, what is $\frac{\partial Q}{\partial n}$?

$$\frac{\partial Q}{\partial p} = p^n ln(p), \tag{3}$$

according to the reasoning in the footnote, and this must be negative, given that $p \in [0, 1]$ – specifically, $p^n > 0$, and $ln(p) < 0$ over this range.

Third, what is $\frac{\partial^2 Q}{\partial p \partial n}$?

$$\frac{\partial Q}{\partial p} = p^{(n-1)}(1 + nln(p)). \tag{4}$$

What sign does the right-hand side term have? The expression outside the brackets is positive. The expression inside can be either positive or negative. Why? Recall: as p gets close to zero from above, so $ln(p)$ tends to negative infinity, and the bracketed term will be negative. Recall also that $ln(p = 1) = 0$. Thus as p rises towards 1, so the bracketed term will become positive.

Thus we have established the following three results:

1. For a given n, as p rises, so Q (the proportion of total time that no disruption occurs) rises.
2. For a given p, an increase in class size n *reduces* Q.
3. The effect of increased class size on the total quantity of non-disruption (or $\frac{\partial Q}{\partial n}$) starts negative but becomes positive over the relevant range $0 \leq p \leq 1$.

3.2 The School Administrator's problem

The next part of the model addresses the issue of how a school administrator would pick the optimal class size, given that student disruption is likely to inhibit learning (as we'd expect).

The approach in the paper is to draw on the theory of the firm: Think of a school as a profit-maximizing organization, charging school fees to students who enrol. The school's revenue consists simply of the fee per student multiplied by the number of students enrolled, with total enrollment being given by Z. For simplicity, assume that Z is fixed.[4]

[4] In practice, enrollment will tend to be increasing in a given school's quality, holding fixed the quality of other schools.

The benefit from attending the school (at a given fee) is related to the quality of education provided. This depends on how much students can learn, in turn related to how much disruption there is in the classroom, as per the formulation we have considered above. Let V be the value of a unit of undisrupted learning time. So the amount a given student will learn is equal to Vp^n, and the total value of the educational experience is ZVp^n.

The school's financial costs are equal to the number of classrooms (equal to the number of teachers hired) multiplied by the wage paid to each teacher. Clearly, the number of teachers (and classrooms) is equal to Z/n. Note that increasing n helps spread costs.

Thus we can write down total profits as

$$\Pi(n;p) = ZVp^n - WZ/n, \tag{5}$$

where the problem is to choose n to maximize profits, given p. (We will consider the effect of changing p on the optimum n in a minute.) Because enrollment is assumed to be fixed, for simplicity, we can equivalently focus on per student profits in the sense that we get exactly the same answer, hence considering a different objective function, by dividing the previous equation through by Z. This yields

$$\Pi(n;p)/Z = Vp^n - W/n. \tag{6}$$

Complication 1: does this have what's called an interior maximum? Not always. To see this, let's plot it, given p, against n. One needs to have p in the right range, close to one. We will assume that obtains, henceforth. [Aside: What is the relevant necessary condition here, precisely, for a given n?]

Given there is an interior maximum, let's figure out the optimum n. This requires that we consider the relevant costs and benefits. To that end, we should derive the first-order condition, with respect to n:

$$\frac{\partial}{\partial n}(Vp^n - W/n) = Vp^n ln(p) + W/n^2 = 0. \tag{7}$$

This equation, which is also equation (2a) in Lazear's paper, implicitly defines optimal class size, n^*, in the sense that n^* must satisfy this condition. Rearranging, we have the important (and more familiar) result:

$$W/n^2 = -Vp^n ln(p). \tag{8}$$

This equation also implicitly defines the optimal class size, n^* – see below.

What is the interpretation? In short, equation (8) says that the marginal benefit of raising n (the left-hand side of the expression) should equal the marginal cost of doing so (the right-hand side) at the optimum.

In more detail, think about the marginal benefit: if we raise class size n in a school of size Z, the cost of hiring a given teacher goes down, as the teacher's wage is spread across more students. That is the *benefit* of a bigger class.[5] On the cost side, as n goes up, then we already know that the amount of undisrupted time goes down, in turn resulting in lost educational value. The right-hand side (giving the marginal cost) provides a precise measure of this loss, associated with a slight increase in n.

We can graph the marginal benefit and cost curves (see exercise below), with n on the horizontal axis. Having done so, n^* is found at the intersection of the two.[6]

Given that n^* exists, the next step – indeed, the *key* step for us – is to see how this optimum value changes as p rises, noting that p takes on values between zero and 1 inclusive, or $p \in [0, 1]$.

To that end, formally speaking, we need to totally differentiate the first-order condition with respect to p and n, and re-arrange, to form $\frac{dn^*}{dp}$. (If you are not familiar with this type of operation, then I'm happy to explain further.) Intuitively, the effect is going to hinge on whether the first derivative of the marginal cost is increasing or decreasing in p, given that the marginal benefit is not a function of p at all – see the left-hand side of (8). If the marginal cost shifts down, then we are in business, as you can see by sketching the relevant curves: this will lead n^* to increase.

Complication 2: The marginal cost need not shift down as p rises!

But fortunately, for p quite close to one, it is possible to verify that it does shift down. Hence

Exercise: based on equation (8), sketch the marginal cost and marginal benefit curves, with n on the horizontal axis. (You can do this very easily using Excel.)

Exercise: starting with p values fairly close to 1, see what happens to the marginal cost curve when you raise p further. And, in turn, see what happens to optimal n^*, given by the intersection.

The answer to this latter question is that n^* will increase as p rises, which is the key result in Proposition 1. (If you are interested, one can prove this result analytically also.)

[5] Note the slight oddity: the extra benefit is associated with more widely spread financial costs.

[6] A couple of remarks here: first, note that the marginal benefit curve slopes down in n, and so does the marginal cost curve, so in some cases there may not be an intersection; second, we need the marginal cost curve to cut the marginal benefit curve from below at the optimum.