

E-Learning Stata

Araceli Ortega Díaz
Instituto Tecnológico de Estudios Superiores de Monterrey,
Escuela de Gobierno y Transformación Pública
araceli.ortega@itesm.mx

Abstract

We developed a hybrid course motivated by the success of the Enspire modules that help students to learn complex subjects at their own pace. E-learning Stata combines econometric material taught in classroom supported by the platform Blackboard, and online videos which the students can revise at their own pace using the same class material and learning the solutions from scratch. We found that the best way that E-learning Stata works is when there are three components: classroom lectures with the teacher, supporting online material where classes are taught in blackboard, and link the application of the knowledge learnt with a real life final project where students test the skills learned.

Keywords: STATA, econometrics.

1. Introduction

Universities are more and more using Massive Open Online Courses (MOOCs) and students are demanding more flexible ways of learning. However, one concern is to track the impact these courses have in effective learning. Students are eager to learn new subjects and technologies, but massive MOOCs do not often provide interaction and personalise feedback to students.

The present paper explains the innovation we implemented to create a hybrid course where students can learn to use the software Stata in econometrics classes at postgraduate level (MSc and PhD) and last grade of the undergraduate in economics, including research assistants.

We based this implementation in the model done by Bjorn Billhardt called Enspire learning [1], who thought that the “on line education space is highly competitive, with many entrepreneurs recognising that Web-based learning would someday replace or supplement a significant part of many education programs”. The stage of our implementation is in its initial stages but we continue improving our hybrid course semester after semester. We also review several taught online econometric courses in the top universities like UCLA [2] and books with data bases like “econometric analysis” of Greene [3].

What was Enspire about? Enspire considers that every student is different and has different learning needs (visual, verbal, active, reflective, factual, theoretical, lineal and holistic). Enspire consists in online tutorials of 2-3 minutes explaining with several examples a complex

concept (statistics, mathematics, monetary policy, etc.). The student learn a concept, can see several specific examples, and the advantages are that the students can review several times the same lecture at their own pace, and receive feedback almost personalised at any time, just by clicking friendly buttons on a well-developed platform.

For teachers the advantage is that they can take several two-minutes tutorials and customise their own course, and allow students to see only what they customised for their class.

Enspire was developed stage by stage using student’s feedback to improve their teaching segments. In the following section we explain how the E-learning Stata course was developed.

2. Developing the course

2.1 Teaching Econometrics without any application

Econometrics and Mathematical Economics are the amazingly interesting areas for economist that like mathematics, and theoretical economist find this subjects respectful as long as everything is well explained in equations and theorems. But in real life problem solving, economists, policy makers, financiers and entrepreneurs among other are urged by the need of getting practical numbers that solve a difficulty.

As a lecturer, teaching from quantitative techniques for BSc in economics to theoretical econometrics for PhD in Financial Science, I learnt that pure theory helps students to acquire a great logical way of thinking to problem solving and this skill is a must in any student.

Notwithstanding when a student cannot quite grasp the theory or find its application, he may lose interest. Nudging that students solve a real life problem by applying the theory through using software increases its learning skills.

2.2 Teaching Econometrics with real life cases using a software

After four years of teaching only theoretical econometrics I decided to include real life exercises where students could see how theory was implemented. Given the complexity of equation solving we use three types of software 1) Interactive Time Series Modelling, 2) EViews and 3) Stata. The first one is useful mostly for analysis of single time series; the second one is used more for financial series which availability is more frequent; and the last one is used in large data base handling from economics, medicine, biology, chemistry, sociology, etc.

Therefore the course of econometrics and its applications to impact evaluation and experimental economics has a Stata laboratory. The barrier to overcome was that the theoretical part consumed 100% of the classes, so including a laboratory had to be done either by reducing the theoretical time, or by extending the class hours.

From August 2008 to December 2010 the hours were extended, some advanced volunteer students would learn in an individual classroom the basic commands of the software and I split my teaching session into two parts: lecture and lab class applied to real life cases. In 2011 two opposing factors changed the situation: 1) we had no more available volunteers as they could not receive compensation or help anymore and 2) a computer lab was finally created with Stata in each computer.

Therefore, from January 2011 to August 2013, I thought all sessions by myself but it was tiring for both students and I because sessions were extended, instead of 6-10pm to 10pm or longer and paying attention and knowledge assimilation was decreasing as time increased.

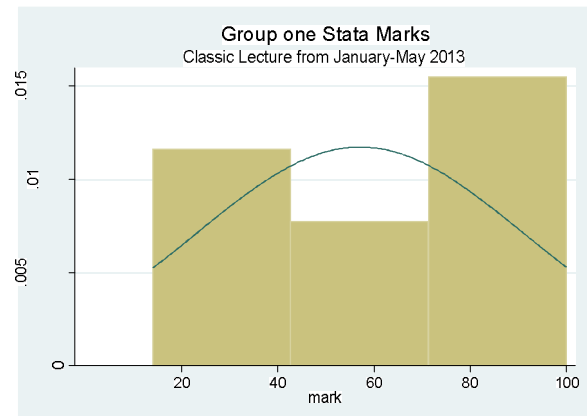
Learning that the creators of *Enspire Learning* managed to set an online course that was interactive, maintaining the high quality of contents and effective learning, I decided I could apply that methodology. Enspire Learning also provided with data on how much this type of courses can increase the number of students attending the course at marginal costs because

once the seed is done, it can be taught anywhere.

2.3 Designing the hybrid course

Before starting the hybrid course that would consist on students receiving the training online and clarifying their questions in sessions with the lecturer, we took a base line measure of performance. The last econometrics course taught using Stata in the classic way in a classroom was from January to May 2013 (see the syllabus in table A1 in the appendix). The course was uploaded in Blackboard, this included power point lectures, word lectures, data bases, and Stata programming files (do files and smcl files) and journal articles. The course consisted on classroom lectures and classroom laboratory, the Stata laboratory for the whole semester accounted for about 7 hours in classroom plus 21 hours of homework. The students were tested on paper at the end of the course and in their Stata knowledge was recorded, they were nine MSc in Economics students and obtained an average mark of 57.11 in a test of ten basic questions.

The graph 1 describes the distribution of their marks and we can observe they rank from 14 to 100 points.



Graph 1. Stata Marks of Group 1.

Having this low performance of students we set the goal of investigating if students learn more in long period exposed to the course of an intensive course; this would be the solution to improve their Stata knowledge.

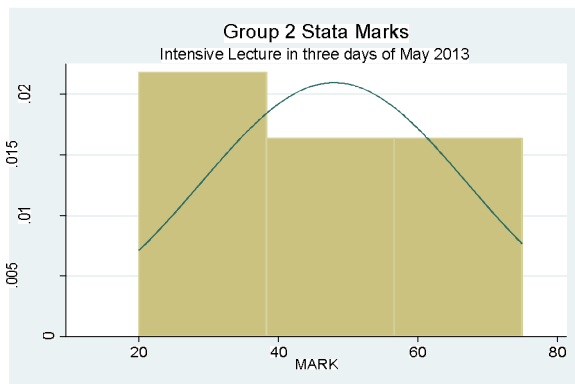
In May 2013 we taught a three days intensive course in the laboratory classroom in the use of econometrics using Stata. The course was uploaded in Blackboard using an attractive design that split in short modules the material (see figure 1), again with power point lectures,

word lectures, data bases, and Stata programing files (dat files, do files and smcl files) and journal papers but using less material than in the previous course (see table A2 in the appendix). The course included 23 hours using Stata, and some homework which students vaguely did.



Figure 1. Blackboard course initial screen.

The same test of group 1 was applied on paper to the 10 students of group 2 and the results were worst as average was 47.78 in a rank of 25 to 75 points (see graph 2). It is a decrease from an average of 57.11 to 47.78. This can suggest that an intensive course where students did not have a whole semester to grasp the tools and review several times the exercises to reinforce the knowledge is not the best way to learn.



Graph 2. Stata Marks of Group 2.

To move forward we took the advantage that this intensive course was taught in few days and was easy to be filmed, so we hired the virtual services to film it and ask them to split the film in small modules. The goal was to have in Blackboard small videos that could be linked to the class material and exercises so that students can repeat and repeat and learn at their own pace the software applied to real life data bases.

The result was 9 videos that were too long in time (about hours each), and we ask the virtual services to redo them as we ask for small module, but they came out with a self-started

DVD of 18 hours non-stop. That problem has to be solved immediately as semester August to December 2014 was about to start, and we planned to test the E-learning Stata in that semester. As the 9 videos were too heavy (4 Gigas each) to upload in Blackboard and to link them with small word and power point presentations, we upload them in YouTube.



Figure 2. Module 1 of Stata Course.

Class 1.1, Module 1. Basic commands of OLS regression models (1 hour 15 min), and use of the Mexican household survey ENIGH and poverty lines from time 01:15 to 02:22 hrs.

<http://www.youtube.com/watch?v=0IP0Q5DqVCg>

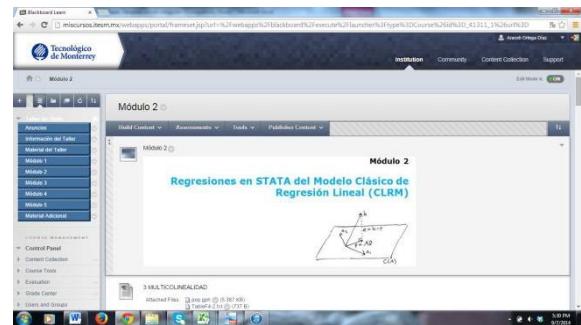


Figure 3. Module 2 of Stata Course.

Class 1.2, Module 2. Poverty lines tables from minute 0 to 14, and Hypothesis Test of Regression Analysis from minute 14 to 48.

<http://www.youtube.com/watch?v=lpwh1VmTicc>

Class 1.3, Module 2. Hypothesis testing from min 0 to 40 and faculty exercise from 00:40 to 01:55 hrs. Simple exercises with *mata* of non-linear hypothesis testing 01:55 to 02:18hrs.

http://www.youtube.com/watch?v=qEBJRT3C_gc



Figure 4. Module 3 of Stata Course.

Class 2.1, Module 2. Structural change tests from min 0 to min 30. Then Module 3 How to build a panel data base using Penn World Table [4] from minute 00:30 to 01:50 hrs.

<http://www.youtube.com/watch?v=0ADk1pdV4rq>

Class 2.2 Module 3 Fixed Effects vs Random Effects using real data from journal articles or cigarette consumption [5] minute 0 to 41. Afterwards we use real data from crime and policy [6] from minute 00:41 to 01:28 hrs.

http://www.youtube.com/watch?v=7X_x2MAcjhA

Class 2.3 Module 3. Dynamic panel data using GMM of Arellano & Bond [7] from minute 0 to 01:13 hrs. And explicit exercise using panel data 02:32. Module 4. Obtaining the propensity score command and the poverty data base minutes 02:35 to 02:57.

<http://www.youtube.com/watch?v=eTwsN9QcIxo>

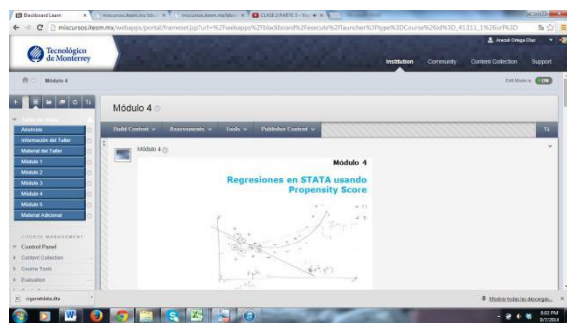


Figure 5. Module 4 of Stata Course.

Class 3.1 Module 4. Probit and logit models. Too heavy to be uploaded.

Class 3.2 Module 4. Poverty base data management form minute 0 to 18. Then Propensity Score [8] from 00:20 to 01:40.

<https://www.youtube.com/watch?v=PcqnXGjsOC8>

Class 3.3 Module 4. Then Propensity Score [9] from minute 0 to 14. Module 5. Quantile regression [10] from minute 15 to 01:24. And finally the test from minute 01:45 to end 2:20hrs.

<http://www.youtube.com/watch?v=tFqP50DEfog>

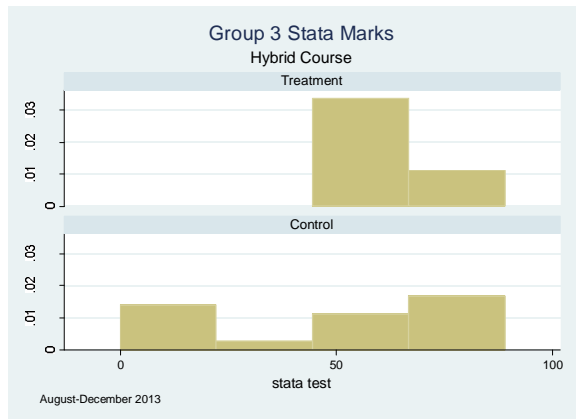


Figure 6. Module 5 of Stata Course.

We can observe that some videos were filmed out of focus, or typing command could not be seen clearly. Sometimes when students asked questions it was not easy to listen to the question and we can only listen to the teacher's answer. Moreover, as the introduction was missed in the first video, and all the videos were too long, we decided to acquire equipment to film our own short videos. We ask for the projector-data camera and audio, these took almost seven months to be approved and arrived.

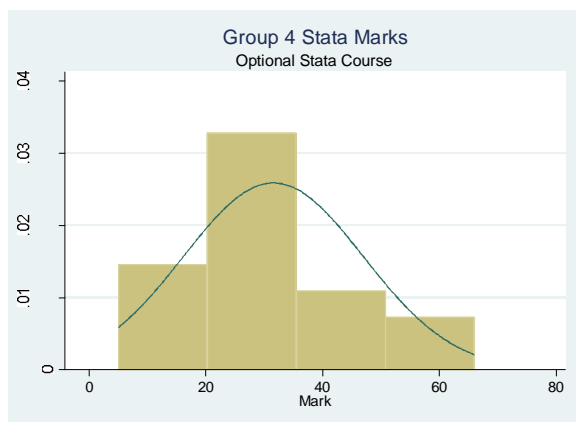
Therefore, for the semester August to December 2013 we decided to use a control and treatment group using differentiated teaching techniques. We taught one initial class to a group of 26 students which included theory and 30 minutes of Stata Lab, then 8 of them had sequential presence Stata classes and complete a 2.5 hours lab of Stata (see table A3 in the appendix). All students were told to study from the online Stata course which included the material and videos. But it was only compulsory for the treated 8 students to use Stata for their final project. At the end of the course we applied the same test as group 1 and 2 but online using the Blackboard exam tool. From the 26 students, 24 answered the test; 8 of the treated group and 16 of the control group.

The results show that using this hybrid method had positive results because the treatment group obtained a higher mean mark than groups 1 and 2, their mean was 62.88, and the control group had a mark mean 48.13 which was higher than group 2. This can give us insight that the hybrid course works better in longer periods where students can revise the material several times.



Graph 3. Stata Marks of Group 3.

As we kept waiting for the equipment to re-film the modules and make them more interactive, we decided that because the aim of the E-learning Stata is that students can learn the material in their own, we ask students in the semester January-May 2014 to learn the material on their own. We taught them in the classroom 3 hours of Stata in 5 short sessions (see table A4) but this time we did not link the learning of Stata and the material to any project. We apply the same test ad all three groups, but online as in group 4. This time the performance decrease as also, doing the test was optional and without consequences to reduce the final mark. The group had 25 students but only 18 answered the test, and the average mark of the group dropped to 28.56 (see graph 4).



Graph 4. Stata Marks of Group 4.

Observing the overall performance of how the hybrid course has been impacting student performance we noticed that groups with little time exposure to the material like group 2, control group 3 and group 4, had low performance and when there is no attached mark to the performance as in group 4, the performance is lower (see Table 1).

We found that the best performance happened when students had classroom classes combined or supported by the online videos, and that their mark is attached to use what they learnt in a real life project.

Table 1. Summary statistics of students marks

	Group 1	Group 2	Group 3	Group 3	Group 4
Marks	T	T	C	T	T
Average	57.11	47.78	48.13	62.88	28.56
Minimum	14.00	20.00	-	47.00	3.00
Maximum	100.00	75.00	89.00	76.00	66.00
Standard deviation	33.99	19.03	27.86	9.23	18.00
Observation	9	10	16	8	25
Missed values	0	0	2	0	7

By the end of July 2014 we received the equipment to start making the small 2 minutes videos and we will continue improving the material and we are currently making it available for students in the semester August-December 2014, and will measure the performance.

3 Conclusions and Further research

The present document explains the creation of a hybrid course to learn Stata software that is used for econometrical implementation and is very valuable to employers in finance, economics and biomedical sciences.

We developed the hybrid course motivated by the success of the Enspire modules that help students to learn complex subjects at their own pace.

E-learning Stata combines econometric material taught in classroom supported by the platform Blackboard, and online videos which the students can review at their own pace using the same class material and learning the solutions from scratch.

We found that the best way that E-learning Stata works is when there are three components:

1. Classroom lectures with teacher.
2. Supporting online material where classes are taught in Blackboard.
3. Link the application of the knowledge learnt whit a real life final project were students test the skills learned.

The challenge we have now is to remake the online videos that are too long (from 48 minutes to 2.5 hours) and difficult to link to each class material separately, and make 2 minutes videos with the equipment acquired and link it to more direct feedback and exercises.

4 Acknowledgements

I would like to thank to all students that had participated in the courses, also the NOVUS project for funding the E-learning Stata project and the Tecnológico de Monterrey for the facilities to teach the course.

References

- [1] R. Hallowell, "Enspire Learning", *Harvard Business Scholl, Case 9-802-001*, pp 1-23 (2002).
- [2] Institute for Digital Research and Education, "Home Page", <http://www.ats.ucla.edu/stat/>, visited on 08/06/2014.
- [3] Greene, W. H., *Econometric Analysis*, 7th Edition, Prentice Hall, New York (2008). <http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm>
- [4] Center for International Comparisons at University of Pennsylvania, "Home Page", <https://pwt.sas.upenn.edu/>, visited on 08/06/2014.
- [5] B. H. Baltagi, J.M. Griffin, and W. Xiong, "To Pool or Not to Pool: Homogeneous versus Heterogeneous Estimators Applied to Cigarette Demand", *The Review of Economics and Statistics*, Vol. 82, No. 1 (Feb., 2000), pp. 117-126
- [6] C. Cornwell and W.N. Trumbull, "Estimating the Economic Model of Crime with Panel Data", *The Review of Economics and Statistics*, Vol. 76, No. 2 (May, 1994), pp. 360-366.
- [7] M. Arellano and S. Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *The Review of Economic Studies*, Vol. 58, No. 2 (Apr., 1991), pp. 277-297.
- [8] P.R- Rosenbaum and D.B. Rubin, "The Central Role of the Propensity Score in Observational for Causal Effects", *Biometrika*, Vol. 70, No.12 (1983), pp. 41-55.
- [9] M. Arellano and S. Bond, "Estimation of average treatment effects based on propensity scores", *The Stata Journal*, pp. 1-19.
- [7] M. Arellano and S. Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *The Review of Economic Studies*, Vol. 58, No. 2 (Apr., 1991), pp. 277-297.

Appendix

Here we are including the syllabus of the courses taught.

Table A1: Syllabus of first group examined.

Lecture	Date	Topic	Minutes of stat lab
1	??????????	Lineal Regression Model (Brooks, Chap 1-4)	??
2	??????????	Lineal Regression Model (Brooks Chap 1-4))	??
3	??????????	Classical Lineal regression model (CLRM) (Greene, Chap. 6-8, Ed 4; Chap. 2-7 Ed. 6).	??
4	??????????	Other assumption on perturbances (Greene, Chap. 11-12 Ed. 4, Chap. 8 Ed. 6)	??
5	??????????	First term test	
6	??????????	Panel Data Models (Greene, Chap. 14 Ed. 4, Chap. 9 Ed.6, Baltagi Chap 1.)	??
7	??????????	Panel Data Models (Greene, Chap. 14 Ed. 4, Chap. 9 Ed.6, Baltagi Chap 1.)	??
8	??????????	Dynamic panel data (Greene, Chap. 17 Ed.4, Chap. 20 Ed. 6)	??
9	??????????	Instrumental Variables (Chap. 12, Ed. 6).	??
10	??????????	Second term test	
11	??????????	Eastern Holiday	
12	??????????	Models of Limited Dependent Variables (Greene, Chap. 19 Ed.4, Chap. 23 Ed. 6) (Ravallion (2001))	??
13	??????????	Models of Limited Dependent Variables (Greene, Chap. 19 Ed.4, Chap. 23 Ed. 6)	??
14	??????????	Multilevel models (Kim, J.Seon (2008))	??
15	??????????	Multilevel models (Kim, J.Seon (2008)) /Hand in final project	??
16	??????????	Presentation of Final Proyecto	
17	??????????	Final term test	
		Total time of laboratory of Stata in hours	7.0

Table A2: Syllabus of the second group examined.

Lab	Date	Topic in the intensive Course of Stata Laboratory	Hours
1	27-May.13	BASIC STATA COMMANDS	
		Introduction	1
		Data preparation	1
		Poverty example using household survey ENIGH-2010	1
		Hypothesis tests	1.5
		Structural Change tests	1.5
2	28-May.13	Classical Linear Regression Model (CLRM) (Greene, Chap. 6 Ed. 4, Chap 2-4 Ed. 6).	6
3	29-May.13	Panel Data Models (Greene, Chap. 14 Ed. 4, Chap. 9 Ed.6)	6
		Fixed effects and random effects	
4		Propensity Score (Rosenbaum & Rubin. "The central role of propensity score in observational studies for causal effects". <i>Biometrika</i> , 1983.).	3
5		Quantile Regression Models (Abadie, Angrist & Imbens. "An instrumental variable estimates of the Effect of Subsidized Training on the Quantiles Trainee Earnings". <i>Econometrica</i> . Vol. 70. January 2002.)	2
		Total time of laboratory of Stata in hours	23

Table A3: Syllabus of the third group examined.

Day	Date	Topic	Minutes Stata Lab Presence		Optional Online Stata Lab	
			T	C	T	C
Tuesday	27-Aug-13	Basic Elements of Impact Evaluation and randomization/ Targeting	30	30	180	180
Tuesday	3-Sep-13	Propensity Score Matching / Double Difference	30		180	180
Tuesday	1-Oct-13	Econometric techniques according to project	60		180	180
Tuesday	15-Oct-13	Econometric techniques according to project	60		180	180
Tuesday	5-Nov-13	Revision				
Tuesday	19-Nov-13	Final project presentation				
		Total of hours using Stata	3.0	0.5	12.0	12.0

T = treatment group, C= control group

Table A4: Syllabus of the fourth group examined.

Day	Date	Topic	Minutes Stata Lab Presence	Optional Online Stata Lab
Tuesday	25-Mar-14	Basic Elements of Impact Evaluation	30	180
Tuesday	1-Apr-14	Randomization/ Targeting	30	180
Tuesday	8-Apr-14	Propensity Score Matching	30	180
Tuesday	15-Apr-14	Eastern Holiday		
Tuesday	22-Apr-14	Double Difference	60	180
Tuesday	29-Apr-14	Poverty cases	60	180
Tuesday	6-May-14	Final Exam / No project in this topic		
Total hours of Stata			3.0	12.0