TEACHING TIME SERIES TO UNDERGRADUATE STUDENTS

Mike Forster
Department of Statistics
The University of Auckland, New Zealand
m.forster@auckland.ac.nz

Time series is a neglected area of statistics. Few secondary schools include it in their statistical curriculum. There is minimal time series, if any, taught at undergraduate level in tertiary education, except in econometrics. In graduate level statistics the frequency domain and/or the Box-Jenkins technique are most common. In 2006 we began a third year undergraduate Applied Time Series course to provide a comprehensive set of tools that enabled students to forecast future activity. In 2018 the course was taught twice, pulling in over 190 third and fourth year students. Some of the advanced techniques taught include ARCH & GARCH, Panel Data and Co-integrated Time Series models. Using modern computer technology, advanced topics can be taught, provided the students understand the basics of time series.

INTRODUCTION

Time series has been a neglected area in statistics with university statistics students only really having the opportunity to study the topic when they reach postgraduate level. The techniques that they then learn are usually Spectral Analysis and the Box-Jenkins differencing technique (SARIMA). Spectral Analysis is used by engineers, physicists, astronomers and other specialists who are all interested in isolating significant frequency signals. When analyzing real data the Box-Jenkins differencing technique is the *standard* methodology to turn non-stationary time series into stationary series so that the *preferred* stationary models can be fitted to the differenced data and forecasts done.

The other discipline that regularly uses time series is econometrics, and there, spectral analysis is of little help. Seasonality dominates most economic time series and in any spectral analysis it will be the main finding. Predictions with spectral analysis are also not possible. Until very recently, it was only banks, financial institutions, insurance companies and government that used time series analysis. Now, with modern computing technology, businesses of all types collect and save their data and hopefully *try* to use it when planning for future sales, stocks and raw material requirements, growth and expansion. The data they collect is non-stationary time series data as it most often contains trends, cycles and seasonality.

THE TEACHING OF TIME SERIES IN NEW ZEALAND

In New Zealand time series has been taught in secondary schools since 1996 (Passmore, 2018). In 2013 the syllabus was altered and the content shifted away from doing a moving average seasonal decomposition with an Excel spreadsheet to a module aimed at understanding through visualization techniques made possible by iNZight. While the popularity of the module remained at around 15,500 students annually, teachers were concerned about how they should award excellence when they faced a written or research component to assess. This led to some shying away from the module (Furushima, et al. 2018, Passmore, 2018).

At Auckland University, I first encountered time series in second year econometrics (1994), and then in a second year data analysis course (1996). However, that was the commerce version of the course. Something additional was needed there while the science students were introduced to R as commerce students had been in their version of the first year course.

My first real dabble into time series analysis, although I didn't realise it at the time, was doing exploratory time series analysis in the mid to late 1970s when I studied political science. I decided to draw a graph of the opinion poll results in the lead-up to the New Zealand election, even joining up the chronological poll results for each party with straight lines. While my main focus was on who was leading, I also paid attention to whether the gap between the two major parties was consistent and tried to determine what caused any changes in the size of any gap that I observed.

The mathematics required in the standard graduate time series course was demanding. Fourier transformations and imaginary exponentials were ever so slightly beyond the limit of my second year mathematical abilities. The non-trivial nature of the differencing technique required in the Box-Jenkins methodology affirms my belief as to why time series, as practiced by statisticians, has been left until graduate level studies. I included econometric courses in my postgraduate statistics study to try and understand that discipline a little better.

The major difficulty I encountered was the statisticians approach, let the data speak, whereas in econometrics, theoretical models were applied to data and if the model fit was wanting, the data was questioned, often for very good reasons. In a time series econometrics course which required a research essay as its only assessment, I modelled the US consumption function. Greene (2003, p171-3) used the US consumption function to illustrate non-linear regression techniques. While the model fitted well, it seemed a little odd to me as it suggested that the marginal propensity to consume from disposable income would be increasing and sooner or later would exceed unity. My hypothesis was that there was a break in the trend of the *essentially* linear US consumption function. After fitting a co-integrated time series model that included a highly significant break term, I rejected the null hypothesis of no break (Forster, 2003).

Subsequently, I updated and revised the time series component of the second year Data Analysis course in statistics (Forster, 2005, 2007) and then in 2006 began teaching a new course in Applied Time Series Analysis at third year undergraduate level (Forster, 2014). Third year applied time series courses in statistics are rare. Rob Hyndman has taught one at Monash University in Melbourne, Australia for many years and some of his ex-students have commenced similar courses elsewhere. Hyndman’s online book on forecasting is quite similar to the approach and pedagogy I promote. We differ on some aspects, but both agree that time series forecasting is a key component of successful business management (Hyndman, 2017).

We, as statistical educators now face a situation where many everyday companies (as well as government) have huge amounts of time series data. Unfortunately, they have no-one (or very few) who can model it and then use their model to predict future prospects. If companies could use all this data to forecast future sales, they would be able to make informed decisions on developments (increased warehouse space, greater stock of raw materials, additional labour, ….) rather than struggling to make changes forced on them by increased (or decreased) demand for their outputs, which is something that is often not realized until after the event.

TEACHING APPLIED TIME SERIES ANALYSIS TO THIRD YEAR TERTIARY STUDENTS

Typically, tertiary level time series courses start with stationary time series models. In my opinion, starting at the point the students have reached in their study of statistics in second year Data Analysis is a more sensible approach. Most students who have completed second year data analysis courses understand ANCOVA, and all that comes before it. Initially the course summarized the notes in a couple of lectures and then launched into a five week chapter that covered the special characteristics of time series data in the framework of non-stationary time series modelling using ordinary least squares regression.

In 2011, when experimenting with in-class clickers (Forster, 2014) it became apparent that the students were getting confused trying to digest time series peculiarities and modelling non-stationary series with regression in tandem. In 2012, when the course assessment included 10% clicker assessment, the second chapter of the course notes (Forster, 2018) was split into two chapters:

- Preliminaries or features peculiar to time series data, techniques rarely used with cross-sectional data and the basic ideas surrounding time series forecasting.
- Non-stationary Time Series modelling (Forster, 2014).

Since 2012 I have maintained the original structure for the second half of the course: Stationary Time Series models, ARIMA and Topics in Time Series. Additional material has been added into the earlier chapters but the most substantial change has been the introduction of more advanced topics such as ARCH and GARCH, Panel data, Random Walks, Spurious Regression and Co-integrated Time Series Modelling into the *Topics* chapter. These are topics that are sometimes covered in graduate level courses in statistics and econometrics. Optional additional advanced topics would include Cross-correlation, State-space models or Long-memory models with each contender extending an existing theme in the course.

PEDAGOGY: ENGAGEMENT, THINKING AND LEARNING IN CLASS

In teaching at tertiary level, both in economics and in statistics, I have *experimented* with numerous different ways to try and get students involved, thinking and learning while they are in class. My rationale has always been that if they enjoy class, they will attend, and if they attend, they will do well. The most successful pedagogy I have tried is in-class clicker assessment (Calverley, 2012; Forster, 2014). It worked incredibly well in second year data analysis, third year applied time series and in my latest trial, a third year course in applied multivariate analysis.

The most interesting pedagogical development while teaching the multivariate course was the inclusion of weekly online quizzes. This was included as a trial for a senior colleague. The students were given the best weekly score they obtained either with in-class clicker questions or the weekly online quiz. The students who regularly attended the three 8 am lectures were awarded the larger of their two scores. The students who did not attend or who rarely attended lectures had the weekly quiz score. There was a small group, clearly identified by me on a couple of occasions, who logged in using their smart phones and answered the clicker questions from the *bus stop*. Since we have shifted to the online version of Qwizdom, students who know the time of the lecture and the log-in code can log-in from anywhere. Once identified, I found that their performance in assignments and in the test was indistinguishable from that of the non-attenders, so I included them in the non-attenders group. The average coursework component marks for the two groups are shown below.

Table 1: Average coursework component scores for lecture attenders and non-attenders

	Assignments	Test	Quiz	Exam
Attenders	72.4 / 100	69.3 / 100	8.7 / 10	76.6/100
Non-attenders	55.1 / 100	58.3 / 100	6.2 / 10	64.6/100
Difference	17.3	11.0	2.5	12.0

The improvement in student performance when clickers are used to get the students involved, thinking and consequently, learning in class is clearly demonstrated, once again.

WHAT CAN TIME SERIES LITERATE STUDENTS DO?

Once students complete their undergraduate degree with a course in applied time series modelling, they are able to take any firm’s data, model the past and, provided the patterns of the past continue into the future forecast period, predict the company’s future sales, revenue, growth and a variety of other factors relevant to the individual firm. Currently, most firms that realize they have all this information have no-one who can use it to build models nor, more importantly, anyone who can build time series models and do forecasts.

Time series analysis has now moved from the preserve of graduate statistics and econometrics to mainstream data analysis simply because developments in modern computer technology allow us to do data analysis in a *timely* manner. While finance graduates used to fill the available positions in insurance, banking and other financial institutions, my experience, based on anecdotal evidence as a graduate officer and *concerned* teacher over two decades, is that the financial sector has moved towards hiring graduates in mathematics and statistics. They tend to be able to pick up finance (in essence applied statistics) with ease, but they also have *general modelling* capabilities that make them a better long-term employment prospect. Sensible finance and marketing students are now doing conjoint degrees in commerce and arts or commerce and science with mathematics, applied mathematics and statistics as their conjoint majors.

TECHNOLOGY – GOOD OR BAD FOR TEACHING DATA ANALYSIS?

The advent of modern computing has created what is essentially a new subject, data analysis. It relies on mathematics and computers, but is, in my view a tertiary level subject in its own right. Since the development of modern analysis programs, data analysis has become easier, less dependent on mathematical ability and largely independent of computer science.

CONCLUSION

Time series has been the forgotten topic within statistics. The coverage of time series in statistics has always been theoretical and focused on techniques that support areas of interest in science based subjects such as Physics and Engineering. The modelling of non-stationary time series has, until recently been the sole preserve of econometricians whose main interest and focus is forecasting. When statisticians come across non-stationary time series, they tend to difference to stationarity and apply their *preferred* stationary models.

Now that all businesses, including non-profit organizations and government collect and store their data on computers, modelling their time series data enables them to build models and forecast future values so that they can make informed evidence based decisions.

We, as statistical educators have a responsibility to ensure our graduates have the requisite skills to forecast the future with time series data when they enter the workforce.

REFERENCES

Calverley, A. (2012). Clickers: A study of student opinion on Audience Response Technology. (Unpublished honours dissertation), The University of Auckland.

Forster, M. (2003). The US Consumption Function (Unpublished dissertation). The University of Auckland.

Forster, M., Smith, D. & Wild, C. (2005). Teaching students to write about statistics. In Weldon, L. & Phillips, B. (Eds.), *Proceedings of the ISI/IASE Satellite on Statistics Education & the Communication of Statistics*, Sydney. Voorburg, The Netherlands: ISI.

Forster, M. & Smith, D. (2007). Assessing large second year undergraduate service courses in data analysis. In B. Chance & B. Phillips (Eds.), *Proceedings of the ISI/IASE Satellite on Statistics Education*, Guimares. Voorburg, The Netherlands: ISI.

Forster, M. & MacGillivray, H. (2010). Student Discovery Projects in Data Analysis. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*, Ljubjlana. Voorburg, The Netherlands: ISI.

Forster, M. (2014). Teaching Data Analysis in Large Classes Using Clicker Assessment. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.

Forster, M. (2018). Applied Time Series Analysis, STATS 326. The Department of Statistics, The University of Auckland.

Furushima, K., Passmore, R., & Forster, M. (2018). An Analysis of Student Performance in First Year Tertiary Statistics and Mathematics Courses. Submitted to the *Tenth International Conference on Teaching Statistics*, Kyoto.

Greene, W.H. (2003). *Econometric Analysis* (5th Edition). Upper Saddle River: Prentice Hall.

Hyndman, R.J., & Athanasopoulos, G. (2017). Forecasting: principles and practice. Otexts: Melbourne, Australia. http://otexts.org/fpp/. Accessed on 9 November 2017

Passmore, R. Time Series – Its Place in the Secondary School Curriculum. Submitted to the *Tenth International Conference on Teaching Statistics*, Kyoto.