# University Attendance: An Econometric Analysis

## Vinay Nair – Junior Sophister

*In the past twenty years, the numbers of students attending Irish universities has increased at a dramatic rate. Vinay Nair utilises econometric analysis to examine some of the factors underlying this increase, measures their explanatory power and compares the empirical results to theoretical reasoning.*

### Introduction

### Full-time Undergraduate students

Since 1980, the number of full-time students attending Irish universities has increased by 270%, from around 23,000 to over 62,000 in 1997. The ethos of the populace in the early 1980s was that university education was merely for a selected few. Now, with an increasing amount of students staying on to further stages of education, coupled with a burgeoning economy, it appears that not only have numbers increased, but perspectives have also broadened. However, is this truly the case? In this essay, I hope to determine why the number of places has increased so rapidly and if (and to what degree) the aforementioned factors have affected it.

I will begin by specifying my dependent variable, my two independent variables and my dummy variable, outlining why I selected them in particular. I shall then estimate the model and evaluate the results of my regressions. Finally, I shall draw conclusions from my model, from the analyses that I have carried out.

### Specification

In order to determine the cause of the increasing numbers attending Universities in Ireland, I have selected the following variables.

*Dependent Variable (Y)*

For my dependent variable, variations in which I wish to explain, I have chosen *the number of students attending Irish universities (HEA Institutions).* My first statistic is for 1980/81. The statistic for the most recent academic year (i.e. the present one – 1998/1999) are still unavailable and hence the final statistic is for 1997/98.

*First Independent Variable ($X_1$)*

The first independent (explanatory) variable that I have chosen is the *number of students sitting the Leaving Certificate*. The vast majority of students (over 90%) going on to study in Irish Universities obtained the Irish Leaving Cert. From first principles, I would expect a high positive correlation between this and the dependent variable.

*Second Independent Variable ($X_2$)*

The second independent (explanatory) variable I have chosen is the *GNP per capita.* The level was adjusted for 1990 prices, to adjust for inflationary pressures in the economy. From economic theory, I would expect that there is a positive income effect for University places: as incomes go up, so too would the demand for the places. Therefore, one would again expect a high positive correlation (if perhaps not as high as for $X_1$.

*Dummy Variable (D)*

In early 1995, the Labour Minister for Education, Niamh Breathnach, introduced free fees for Third Level institutions. This radical proposal was implemented as half-fees for 1995/96, but free-fees have been the policy ever since. For the purposes of this study, since the principle was in effect, 1995 is considered a year of ‘free fees’. Hence, from 1980 to 1994, the dummy is allocated a value of "0", whereas 1995 to 1997 have been given a value of "1". Again, I anticipate a positive correlation between the number of students attending university and the introduction of free fees.

It is valuable to have expectations of one’s results before carrying out regressions. However, one must balance these expectations with the temptation to manipulate the data set so as to ensure these expectations are realised (a practice often carried out in econometrics). Indeed, this problem of ‘data mining’ by econometricians is one of the greatest reasons that business forecasts now often supersede econometric forecasts in many professional circles.

*Omitted Variables, Residual term*

In this analysis, considering the various restraints, some independent variables, such as the increasing number of non-nationals occupying places, had to be omitted. The analysis does include, however, a disturbance / stochastic term, which is effectively like a basket in which all effects on Y that are not explained by $X_1$ and $X_2$ are contained. These residuals could occur in three main ways:

1. Omission of the influence of innumerable chance events,
2. Measurement error,
3. Human indeterminacy.

**Line of best fit**

The method of estimation used in this analysis is Ordinary Least Squares. From the estimates, one can construct a line of best fit, based on the multiple regression model:

$$Y = b_1 X_1 + b_2 X_2 + b_3 D + m$$

This method allows us to see the relationship between the variables $X_1$ and $X_2$ and the dummy variable, by estimating the sign and size of $b_1$, $b_2$ and $b_3$ ($b_0$=0, due to inclusion of the dummy variable). The aforementioned stochastic term will be measured by m .

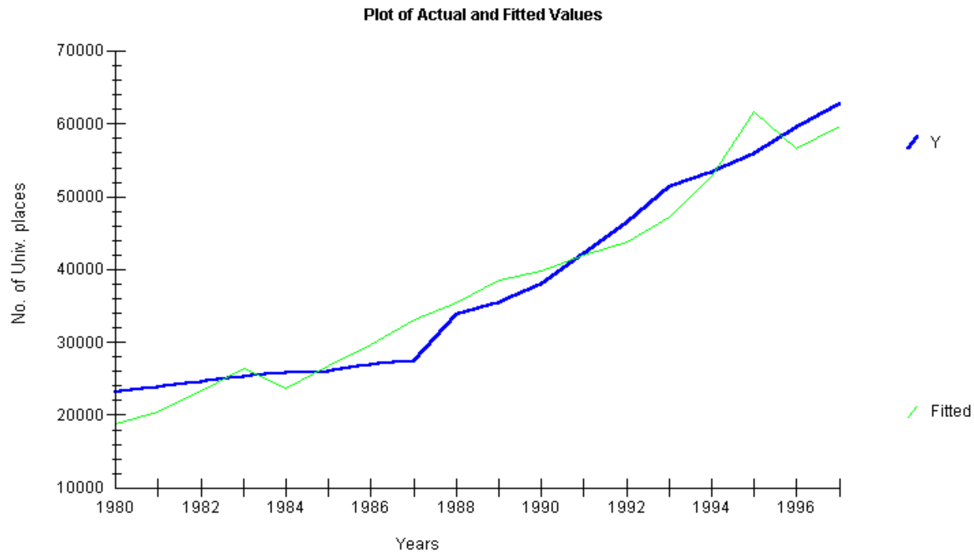
Using the Econometrics computer package, *Microfit*, my line of best fit was found to be:

$$Y = 0.77572\ X_1 - 4.5684\ X_2 + 8167.3\ D$$

**Estimation and Evaluation**

*$R^2$*

I found the correlation between the variables to be extremely high at 94% (correlation coefficient = 0.94663). An adjusted $R^2$, R-bar-squared, which is less biased than $R^2$, was found to be 0.93520; again, this is a very high figure. This high correlation is quite clearly seen from the following graph:




| Indep. Variable | Coefficient | T-statistic | Probability |
|---|---|---|---|
| $X_1$ | 0.77572 | 4.1462 | 0.001 |
| $X_2$ | -4.5684 | -1.9690 | 0.069 |
| D | 8167.3 | 2.6660 | 0.018 |

*$X_1$, $X_2$, and D*

Having seen the high correlation of all the variables, it was important to evaluate what were the *individual* explanatory powers of $X_1$ and $X_2$ (with D), as multicollinearity could still undermine the high $R^2$ (this will be discussed later).

Regressing Y on $X_1$: $Y = 0.42159 X_1 + 6515.2 D$

$R^2 = 0.93185$ t-statistic = 7.4958

Regressing Y on $X_2$: $Y = 4.6798 X_2 + 3309.9 D$

$R^2 = 0.88110$ t-statistic = 5.0794

*Possible Multicollinearity*

Multicollinearity is always a possibility in multiple regression. Particularly due to the fact that the sign of the coefficient of $X_2$ changed from the multiple to the single regression, I regressed $X_1$ on $X_2$, the standard check for multicollinearity. I found $R^2$ to be 0.71409. This is quite a high correlation for the two variables. Therefore, coupled with the changing sign, there is implied significant multicollinearity between my variables.

## Predictions vs. Outcomes of coefficients

My predictions were for the coefficients to be positive and large. Let me now evaluate how this compares to our results. From earlier, my line of best fit is:

$$Y = 0.77572\ X_1 - 4.5684\ X_2 + 8167.3\ D$$

Whilst we see that $X_1$ and D did indeed have their anticipated positive correlation, we duly note $X_2$ has a negative correlation in the multiple regression.

What does this mean? This means that increasing GNP per capita has had an inverse affect on the number of students attending university. As stated earlier, we expected a positive income effect, whereby an increase in the number of students would result from an increase in income. Hence, this negative value contradicts economic theory. However, in the simple regression, the coefficient is a positive value.

I believe that a large reason that this value in the multiple regression occurs is due to multicollinearity between the variables and/or a significant omission from my model.

*T- statistic*

The t-statistic (or "T-Ratio" in *Microfit*), is the "value of the parameter estimate divided by its estimated standard deviation (the standard error)". In the multiple regression case, $X_1$ is statistically significant at the 5% and 10% levels; indeed, it is even significant at the 1% level. However, $X_2$ is only statistically significant at the 10% level (value » 2). Again, the aforementioned multicollinearity may explain this result. The dummy variable is also statistically significant at the 5% and 10% levels.

*F-statistic*

The F-statistic of 82.7760 is high, at zero-probability. This again demonstrates that we can reject any hypothesis that this model has no explanatory power, i.e. that $b_1=0$, $b_2=0$ and $b_3=0$.

*Durbin-Watson test*

The DW-statistic for my model was 1.2194. This value falls between the $d_L$ and $d_U$ limits. This suggests that I cannot ascertain the level of autocorrelation. Naturally, I would have preferred no autocorrelation. Hence, in an attempt to overcome this, I lagged my $X_1$ variable by 3 years (the median length of a university degree). My DW-statistic fell dramatically to 0.65400.

## Forecasting powers

I omitted Y variables in 1996 and 1997, to see if my model had forecasting ability. Unfortunately, the forecasts of values of 58203 in 1996 (s of error 6045.7) and 58617 in 1997 (s of error 5268.8) were considerably off the actual values of 59651 and 62660 respectively.

**Trinity College**

Purely for the sake of interest, I conducted a simple regression of the Y variable (number of students attending Irish universities) on the number of new students enrolling in Trinity ($X_3$), to see if there was any explanatory power. Interestingly, I obtained a rather high $R^2$ value of 0.98697. This high value probably just shows how much of a pacesetter Trinity was and continues to be for the rest of the country!

## Conclusion

This model undoubtedly has sound explanatory variables. In analysing the number of students attending Irish universities, I evaluated the number of students sitting the Leaving Cert., the (increasing) GNP per capita and looked at the impact the introduction of free fees has made. There is a definite correlation between these explanatory variables and my dependent variable – approximately 95%. However, whilst Leaving Cert. students and free fees had the anticipated positive correlation, we saw that, whilst statistically significant at 10%, GNP per capita had a negative correlation to student attendance in multiple regression, yet a positive correlation in a simple regression. This implied two things: (i) the existence of multicollinearity between the variables and (ii) a possible significant omission from my model. Despite this, overall my model, (as shown by t-test and F-test), demonstrates the importance of the role Leaving Cert. students, GNP per capita and the introduction of free fees have had on the number of students attending Irish Universities.

## Bibliography

**Department of Education** (various years) *Statistical Summary* 1981-1998. CSO.

**Gujurati, Damodar** (1995) *Basic Econometrics*. McGraw-Hill: Singapore.

**Kennedy, Peter** (1992) *A Guide to Econometrics.* Blackwell: Oxford.

**Statistical Abstract** 1981-1997. CSO.

**TCD "Student Statistics"** 1981-1998. Records Office, TCD.

## Acknowledgements

Mr. Jim Dillon, C.S.O.

Mr. John McEvoy, C.A.O.

TCD Records Office.

Statistics Section, Department of Education.