

# Predicting scientific research output at the University of KwaZulu-Natal

**AUTHOR:**  
Mike Murray<sup>1</sup>

**AFFILIATION:**  
<sup>1</sup>School of Mathematics,  
Statistics and Computer Science,  
University of KwaZulu-Natal,  
Westville Campus, Durban,  
South Africa

**CORRESPONDENCE TO:**  
Mike Murray

**EMAIL:**  
murraym@ukzn.ac.za

**POSTAL ADDRESS:**  
School of Mathematics,  
Statistics and Computer Science,  
University of KwaZulu-Natal,  
Westville Campus, Durban 4001,  
South Africa

**DATES:**  
**Received:** 10 Oct. 2013  
**Revised:** 19 Dec. 2013  
**Accepted:** 01 Jan. 2014

**KEYWORDS:**  
publishing record; zero inflation;  
Poisson; science faculty; UKZN

**HOW TO CITE:**  
Murray M. Predicting scientific  
research output at the  
University of KwaZulu-Natal.  
S Afr J Sci. 2014;110(3/4),  
Art. #2013-0321, 4 pages.  
[http://dx.doi.org/10.1590/  
sajs.2014/20130321](http://dx.doi.org/10.1590/sajs.2014/20130321)

Academic members of staff at the University of KwaZulu-Natal (UKZN) are expected to publish in research journals that have been accredited by the South African based Department of Higher Education and Training. However, some members of staff have chosen to focus solely on the teaching aspect of their careers and as a result they have no publication record. In this study, a set of per annum productivity unit counts was calculated for every academic at UKZN. Because it is possible for a publishing academic to also record a zero count for a given year, it is necessary to develop an appropriate methodology that can distinguish this zero count from one that will always be recorded by a non-publishing academic. By fitting a zero-inflated Poisson model to the data, specific factors can be identified that separately drive the non-publishing and publishing processes at UKZN. In particular, having a PhD and working in a large school has a significant impact on improving the research output of a publishing academic. If UKZN wants to become a research-focused university, non-publishing academics should be encouraged to undertake a PhD degree.

## Introduction

In South Africa, the Department of Higher Education and Training allocates funds to universities based on a formula that rewards student throughput as well as the research output of publishing academics.<sup>1,2</sup> In this paper, an attempt is made to identify possible demographic and academic factors that help to improve the research output of academics at the University of KwaZulu-Natal (UKZN). Similar studies have been conducted, mainly on US data (for example Xie and Shauman<sup>3</sup>, Aksnes et al.<sup>4</sup> and Kyvik and Teigen<sup>5</sup>), in which it was found that for almost every age group in their respective data sets, men publish more than women. Barjak<sup>6</sup>, Gonzalez-Brambila and Veloso<sup>7</sup> and Kyvik<sup>8</sup> have found that research productivity tends to increase with age, reaching a peak before tapering off towards retirement. What distinguishes this study from those mentioned above is that it analyses a set of per annum based publication counts for all members of staff at UKZN. Some staff members will have a zero value for a particular year because they did not publish anything during that year even though they have a record of previous publications, whereas others may have a zero value because they have chosen to focus entirely on the teaching aspect of their careers and therefore never publish. Essentially, a per-annum publication-based output variable is observed, but an accompanying variable denoting whether or not the person recording this zero outcome is (or will eventually become) a publishing academic (or not) cannot be known. A zero-inflation based modelling approach helps to overcome this problem of not being able to identify the true identity of a given record as belonging to that of a publishing or non-publishing academic by introducing into the modelling process an underlying process – process Z – that generates two types of zeroes: a structural zero when dealing with a non-publishing academic and a true zero when dealing with a publishing academic.

## Research methodology

A zero-inflation model assumes two possible sources from which a zero observation can arise. Academics who have made a conscious decision to never publish will generate what is called a 'structural' zero. Those who have a record of previous publications but have not published in a particular year generate what is called a 'true' or 'sampling' zero.

Focusing on the structural zero recorded every year for a non-publishing academic, a zero inflation model links the probability  $p_0$  of being able to observe such an outcome with the covariate profile  $x$  associated with a particular academic using the following logistic regression model<sup>9</sup>:

$$\ln\left(\frac{p_0(x)}{1-p_0(x)}\right) = x^T\gamma = > p_0(x) = \frac{e^{x^T\gamma}}{1+e^{x^T\gamma}} \quad \text{Equation 1}$$

Focusing on the per annum based research records generated by a publishing academic for whom, in a particular year, a true zero value may have been recorded, a Poisson distribution with a covariate dependent intensity parameter  $\lambda(x)$  that is linked to a set of covariates  $x$  using the following function

$$\ln \lambda(x) = x^T\theta = > \lambda(x) = e^{x^T\theta} \quad \text{Equation 2}$$

is used to model the output generated. With these model choices in hand, a final model formulation – a zero-inflated Poisson model – is obtained for our response variable  $Y$  that sets<sup>13</sup>

$$P(Y=0) = p_0 + (1-p_0)e^{-\lambda(x)}$$

and

$$P(Y=k) = (1-p_0) \frac{e^{-\lambda(x)} \lambda(x)^k}{k!}$$

for  $k=1,2,\dots$

© 2014. The Authors.  
Published under a Creative  
Commons Attribution Licence.

Parameter estimates for  $\theta$  and  $\gamma$  can be obtained using the method of maximum likelihood, details of which can be found in Lambert<sup>9</sup>. Because the mean of a Poisson distribution equals the intensity parameter  $\lambda(x)$  that is helping to characterise that distribution, the above expression linking  $\lambda(x)$  to the covariates in  $x$  implies for any positive estimate that occurs in  $\theta$  that the resulting covariate to which this estimate refers has a positive effect, that is, increases the number of per annum based research outputs produced by that academic at UKZN. Similarly, those covariates that have a negative parameter estimate associated with them are associated with a reduced level of research output.

### The response variable

At UKZN, the productivity unit counts that appear in Table 1 have been used to unbiasedly apportion a 'unit of worth' to a published piece of work. This approach differs from that of most other studies in which each article that has been published is usually weighted by the journal impact factor which measures the average number of citations a journal paper will receive in the first 2 years following its year of publication. Because the focus in this study was to identify specific factors that affect the production of research rather than on the quality of this research, a 'unit of worth' was chosen to use as a measure, but it can be noted that the method to be outlined can just as easily be applied as an impact factor adjusted response variable. This response variable was then rescaled using the following rule so that it better fitted into the model paradigm outlined in the previous section for a zero-inflated Poisson model:

Y	Number of productivity units achieved
0	No productivity units in that given year
1	1–15 productivity units in that given year
2	16–30 productivity units in that given year
3	31–45 productivity units in that given year
4, 5, etc.	46–60 productivity units, 61–65 units, etc.

In addition to this response variable, the following covariates that help to distinguish one academic from another were also collected:

- a 0/1 indicator variable denoting whether the academic member of staff was female or male
- a set of separate 0/1 indicator variables denoting whether the academic was a lecturer, senior lecturer or professor
- a set of separate 0/1 indicator variables denoting the racial group to which the academic belongs (African, Indian or white)
- a 0/1 indicator variable denoting whether the academic has a PhD or not
- a variable denoting the number of academics in the school in which the academic resides (size)
- an age-based category variable taking on a value 0 if the academic (in that particular year) is in their twenties, a value 1 if they are in their thirties, etc.

**Table 1:** Productivity unit counts used to record the research output generated by an academic in the Faculty of Science and Agriculture at the University of KwaZulu-Natal

Research activity	Productivity unit count
Journal article <sup>†</sup> (sole author)	60
Entire book	100
Chapter in book	15
Graduating MSc student	15
Graduating PhD student	60

<sup>†</sup>Co-authored articles were assigned productivity unit counts on a pro-rata basis; for example, a count of 30 was assigned for an article with two authors.

### Results

Data was collected for the period from 2004 to 2008 and consisted of a total of 1236 year-on-year productivity unit counts. A breakdown of the data set according to age, race, gender, qualification and job position is given in Table 2.

**Table 2:** A demographic breakdown of the academics within the data set for this study

	2004	2005	2006	2007	2008
<b>Age</b>					
20–29	11	9	15	16	13
30–39	60	70	68	67	70
40–49	92	80	87	91	86
50–59	70	74	72	70	71
60+	4	8	11	12	9
<b>PhD</b>					
Yes	195	195	197	198	188
No	42	46	56	58	61
<b>Gender</b>					
Male	186	187	191	186	187
Female	54	58	65	73	65
<b>Race</b>					
African	30	30	44	45	45
Indian	56	54	56	52	53
White	151	157	153	159	147
<b>Academic position</b>					
Professor	93	93	93	87	81
Senior lecturer	61	64	70	67	68
Lecturer	83	84	90	102	100

Parameter estimates for the zero inflation model were obtained using Stata. The strongly skewed nature of each of the plots that appear in Figure 1 seem to support the idea that the observed counts are being generated from a combination of two different sources – a non-publishing academic who has chosen to focus on the teaching aspect of their careers and will therefore never produce a research article throughout their entire academic career or a research active academic who in a so-called 'dry year' may have recorded a zero record for that particular year.

Table 3 contains a set of parameter estimates for  $\theta$  and  $\gamma$ . Any significantly positive value that occurs in  $\gamma$  can be associated with a factor that will serve to increase the log-odds ratio in favour of generating a structural zero (i.e. a person who will never publish a research article). Any significantly positive value obtained for  $\theta$  should be interpreted as identification of a factor that will help to increase the expected per annum based research output of that individual when compared with a person from the baseline category (a white female who holds a senior lecturer position and has no PhD qualification).

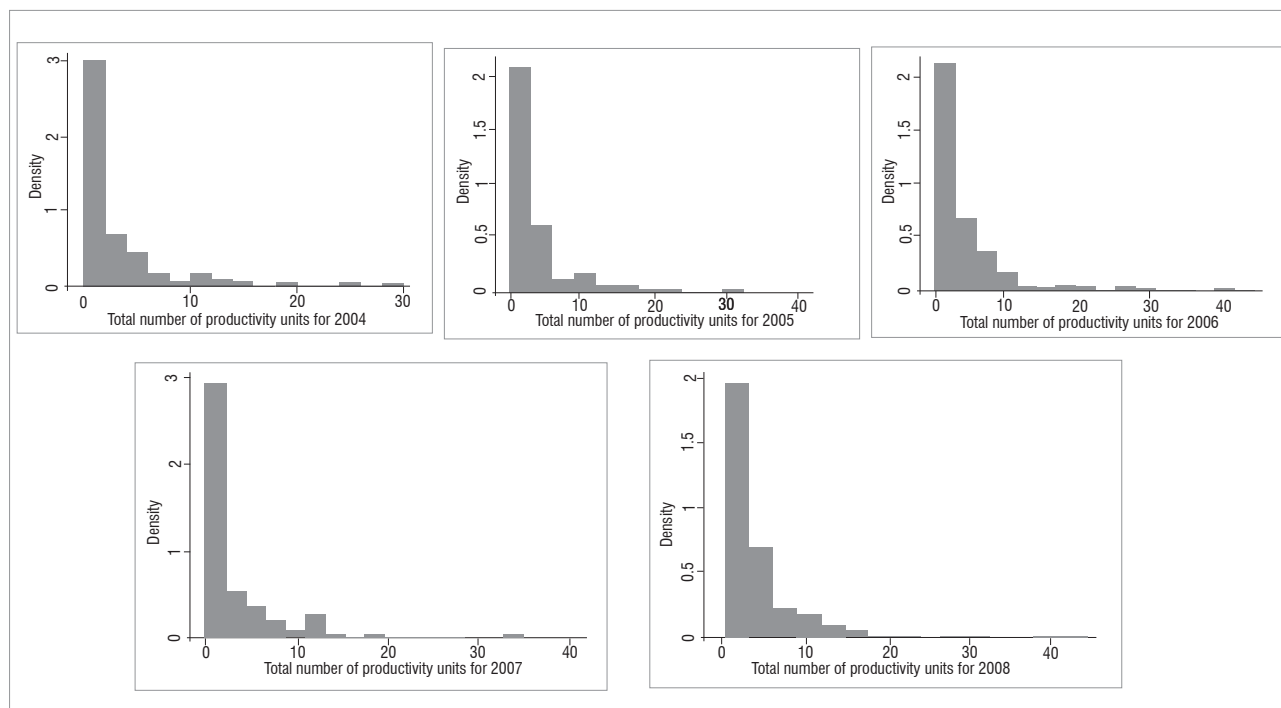


Figure 1: A set of observed sampling distributions for the number of productivity units generated in each year from 2004 to 2008

Table 3: Parameter estimates for  $\theta$  and  $\gamma$  in the zero-inflated Poisson model

	Parameter estimate $\theta$ (SE)	Parameter estimate $\gamma$ (SE)
Intercept term	1.12340(0.1540)*	0.2201(0.4599)
School size	0.0192(0.0015)*	-0.0138(0.0058)
Lecturer	-0.2878(0.0659)*	0.6630(0.1837)*
Professor	0.7627(0.0041)*	-0.9732(0.1706)*
Male	0.1015(0.0399)*	0.1078(0.1588)
African	-0.1148(0.0540)*	-0.0592(0.1913)
Indian	-0.2219(0.0475)*	0.5344(0.1628)*
Age	-0.1866(0.0211)*	0.1407(0.0832)
PhD	0.3934(0.0972)*	-0.8353(0.1917)*

\*Denotes significance at the 5% level

Focusing on those factors that help to distinguish someone who does publish from someone who does not, the estimates for  $\gamma$  that appear in Table 2 suggest that having a PhD, being a professor (rather than a lecturer) and being of non-Indian origin are the only variables that significantly increase the odds ratio associated with becoming an academic who does publish. More importantly, however, are the factors that do not seem to be affecting the odds ratio associated with becoming a research productive academic in the Faculty of Science and Agriculture at UKZN. In particular, age and school size do not seem to play a significant role in whether or not someone becomes a more research productive academic.

Focusing now on those academics who appear to be research active, the estimates for  $\theta$  that appear in Table 3 suggest, at a 5% level of significance, that holding the position of professor, having a PhD qualification, being male and working in a large school all help to increase the research productivity of academics in the Faculty of Science and Agriculture at UKZN when compared with someone from the baseline category (a white female who holds a senior lecturer position in the faculty and

has no PhD qualification). However, staff who are older seem to be less productive when compared with younger staff. Similarly, African and Indian researchers are not producing as much research output as their otherwise identical white counterparts.

## Conclusion

The excess number of zero values observed in our data set may be attributable in part to a large number of academics within this data set who have chosen never to publish in their academic careers. Including these individuals in the analysis may impair the ability to focus on the problem of interest, which is to identify those factors that help to improve the publication rate of academics in the Faculty of Science and Agriculture at UKZN who have chosen to publish. The purpose of this study was to identify a specific set of factors that affect research productivity in the Faculty of Science and Agriculture at UKZN. Because the data set available does not explicitly identify someone as being a publishing or non-publishing academic, a method had to be developed to appropriately separate this pool of academics into those that publish and those that have elected to never publish. A set of covariates were then included in the model to establish their effect on research productivity.

In particular, having a PhD and working in a large school were found to have a significant impact on improving the research output of a publishing academic. If UKZN wants to become a research-focused university, non-publishing academics need therefore to be encouraged to complete a PhD degree.

## References

1. Melck AP. Methods of financing universities with special reference to formula funding in South Africa [DCom thesis]. Stellenbosch: Stellenbosch University; 1982.
2. Venter RH. An Investigation of government financing of universities. 2nd ed. Report: SAPSE-110. Pretoria: Department of National Education; 1985.
3. Xie Y, Shauman K. Sex differences in research productivity: New evidence about an old puzzle. *Am Sociol Rev.* 1998;63(6):847–870. <http://dx.doi.org/10.2307/2657505>

4. Aksnes DW, Rorstad K, Piro F, Sivertsen G. Are female researchers less cited? A large-scale study of Norwegian scientists. *J Am Soc Inform Sci Technol.* 2011;62(4):628–636. <http://dx.doi.org/10.1002/asi.21486>
5. Kyvik S, Teigen M. Child care, research collaboration, and gender differences in scientific productivity. *Sci Technol Hum Val.* 1996;21(1):54–71. <http://dx.doi.org/10.1177/016224399602100103>
6. Barjak F. Research productivity in the Internet era. *Scientometrics.* 2006;68(3):343–360. <http://dx.doi.org/10.1007/s11192-006-0116-y>
7. Gonzalez-Brambila C, Veloso FM. The determinants of research output and impact: A study of Mexican researchers. *Res Policy.* 2007;36(7):1035–1051. <http://dx.doi.org/10.1016/j.respol.2007.03.005>
8. Kyvik S. Age and scientific productivity: Differences between fields of learning. *Higher Education.* 1990;19:37–55. <http://dx.doi.org/10.1007/BF00142022>
9. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics.* 1992;34:1–14. <http://dx.doi.org/10.2307/1269547>

