Multiple Dependent Hypothesis Tests in Geographically Weighted Regression

Graeme Byrne[1], Martin Charlton[2], and Stewart Fotheringham[3]

[1]La Trobe University, Bendigo, Victoria Austrlaia
Telephone: +61 3 5444 7263
Fax: +61 3 54447998
Email: g.byrne@latrobe.edu.au

[2]National Centre for Geocomputation ,National University of Ireland, Maynooth, Co. Kildare, Ireland
Telephone: +353 (0) 1 708 6455
Fax: +353 (0) 1 708 6456
Email: martin.charlton@nuim.ie

[3]National Centre for Geocomputation ,National University of Ireland, Maynooth, Co. Kildare, Ireland
Telephone: +353 (0) 1 708 6455
Fax: +353 (0) 1 708 6456
Email: stewart.fotheringham@nuim.ie

1. Introduction

Geographically weighted regression (Fotheringham et al., 2002) is a method of modelling spatial variability in regression coefficients. The procedure yields a separate model for each spatial location in the study area with all models generated from the same data set using a differential weighting scheme. The weighting scheme, which allows for spatial variation in the model parameters, involves a bandwidth parameter which is usually determined from the data using a cross-validation procedure. Part of the main output is a set of location-specific parameter estimates and associated t statistics which can be used to test hypotheses about individual model parameters. If there are n spatial locations and p parameters in each model, there will be up to np hypotheses to be tested which in most applications defines a very high order multiple inference problem. Solutions to problems of this type usually involve an adjustment to the decision rule for individual tests designed to contain the overall risk of mistaking chance variation for a genuine effect. An undesirable by-product of achieving this control is a reduction in statistical power for individual tests, which may result in genuine effects going undetected. These two competing aspects of multiple inference have become known as the multiplicity problem. In this paper we develop a simple Bonferroni style adjustment for testing multiple hypotheses about GWR model coefficients. The adjustment takes advantage of the intrinsic dependency between local GWR models to contain the overall risk mentioned above, without the large sacrifice in power associated with the traditional Bonferroni correction.

We illustrate this adjustment and a range of other corrective procedures on two data sets. The first models the determinants of educational attainment in the counties of Georgia USA. Using area based census data we examine the links between levels of educational attainment and four potential predictors: the proportion of elderly, the proportion who are foreign born, the proportion living below the poverty line and the proportion of ethnic blacks. The second model is a geographically weighted hedonic house price model based on individual mortgage records in Greater London in 1990. In both models we show how the various corrections can be used to guide the interpretation of the spatial variations in the parameter estimates. Finally we compare the statistical power of the proposed method with Bonferroni/Sidak corrections and those based on *false discovery rate* control.

2. Controlling the family–wise error rate

The traditional approach to the multiplicity problem is to control the probability of a type I error over a set (family) of m related hypothesis tests. The probability of rejecting one or more true null hypotheses is called the family–wise error rate which we denote ξ_m and controlling this quantity is a standard goal of most traditional multiple inference procedures.

If, in a set of m hypothesis tests, E_i is the event – "a type I errors occurs in the ith test", the family–wise error rate is given by

$$\xi_m = P\left(\bigcup_{i=1}^{m} E_i\right) \leq \sum_{i=1}^{m} P(E_i), \tag{1}$$

where the last term follows from Boole's inequality. If $P(E_i)$ has the same value (say α) for all tests, the Bonforroni (1935) correction follows from equation 1 since then $\xi_m \leq m\alpha$ and so setting $\alpha = \xi_m/m$ controls the family-wise error rate to be ξ_m or less. This result does not require the tests to be independent however if they are independent, we can write

$$\xi_m = P\left(\bigcup_{i=1}^{m} E_i\right) = 1 - P\left(\bigcap_{i=1}^{m} \overline{E}_i\right) = 1 - \prod_{i=1}^{m} P(\overline{E}_i) = 1 - \prod_{i=1}^{m}(1 - P(E_i)).$$

Thus, for m independent tests with $P(E_i) = \alpha$, the family–wise error rate is $\xi_m = 1 - (1-\alpha)^m$ and the Šidák (1967) correction follows by choosing $\alpha = 1 - (1-\xi_m)^{1/m}$ which controls the family–wise error rate at exactly ξ_m.

Both of the above approaches become very conservative when the number of hypotheses to be tested is large and when the tests are not independent (Abdi, 2007). For dependent tests Moyè (2003, p. 417–428) uses conditional probability and induction arguments to derive the following generalisation of Šidák's result,

$$\xi_m = 1 - (1-\alpha)\left(1 - \alpha\left(1 - D^2\right)\right)^{m-1}, \tag{2}$$

where the family–wise error rate now depends on $D \in [0,1]$ which is called the degree of dependency (assumed constant) across all tests. For a full discussion of the dependency parameter see Moyè (2003, Ch. 5). If $D = 1$ we need only test any one of the hypotheses in the set using the desired family–wise error rate, since then the decision for this hypothesis completely determines the outcome of all remaining hypotheses in the set. If $D = 0$ the hypotheses are independent and equation 2 reduces to the Šidák result.

Of course it still remains to choose a value for D and unfortunately there are few general guidelines. Moyè (2003, pp. 188–190) discusses the issue in the context of multiple endpoints in clinical trials and later in this paper we propose a method in the GWR context. It appears that the choice of D will always depend on the context of the study and in some cases the analyst will have no clear direction and more powerful alternatives such as the

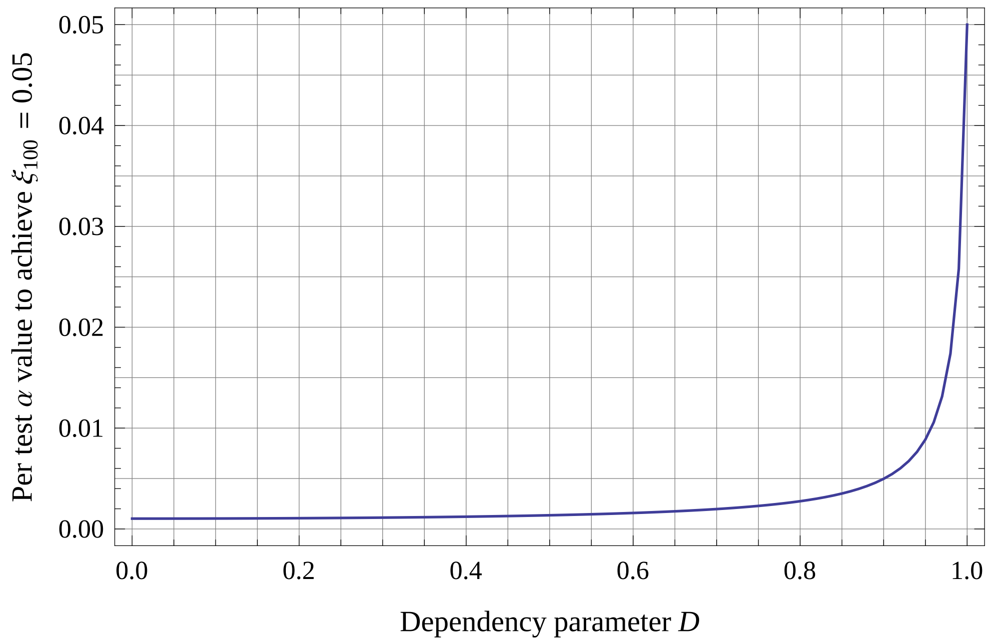

Figure 1: Plot of the per test α value against the dependency parameter D to achieve a family-wise error rate of $\xi_{100} = 0.05$.

Holm–Bonferonni procedure (Holm, 1979) can be employed. If one is willing to redefine the problem in terms of the false discovery rate (the expected fraction of incorrectly rejected hypotheses), the methods of Benjamini and Hochberg (1995) or, for dependent tests, Benjamini and Yekutieli (2001) may be appropriate.

Applying the binomial theorem to equation 2 it is straight forward to show that

$$\xi_m \leq \alpha(1+(m-1)(1-D^2))$$

a result that can also be established using Boole's inequality as demonstrated by Moyè (2003). Therefore, assuming we have a value for D, the family–wise error rate can be controlled at ξ_m or less by choosing

$$\alpha = \frac{\xi_m}{1+(m-1)(1-D^2)} \tag{3}$$

which is a Bonferroni style correction for dependent tests.

Moyè (2003, p. 188) warns against overestimation of D since then the family-wise error rate will not be preserved. The importance of this advice is illustrated by the steepness of the curve in fig. 1 for $D > 0.95$. In this region even a small overestimate of D can lead to large overestimate of the per test α value resulting in the family-wise error rate being considerably larger than required. With this in mind it would be prudent to adopt a value of D at the lower end of its expected range.

3. Dependency in Geographically Weighted Regression

Since the calibration procedure in geographically weighted regression uses the same data set to calibrate models at each spatial location, the method necessarily results in a certain degree of dependency between the models, which must flow over to the t statistics and associated hypothesis tests. A measure of this dependency can be calculated from the GWR hat matrix $\mathbf{S}$ defined in Brunsdon et al. (1999). The effective number of parameters is defined by $p_e = 2\mathrm{tr}(\mathbf{S}) - \mathrm{tr}(\mathbf{S}'\mathbf{S})$ where $\mathrm{tr}(\cdot)$ is the matrix trace operator. Now suppose that there are p parameters in each model to be calibrated and there are n spatial locations. A sensible definition for the dependency between models is

$$D = \sqrt{1 - \frac{p_e}{np}}, \tag{4}$$

which is justified as follows. First consider the (admittedly impossible) case of complete independence among the models for which $p_e = np$ giving $D = 0$. In the case of complete dependence among the models (i.e. the global model is appropriate) we would have $p_e = p$ giving $D = \sqrt{1 - 1/n}$ which will be close to unity in most GWR applications since n is usually large. Therefore equation 4 satisfies the requirement $0 \leq D \leq 1$ and on substituting into equation 3, the family–wise error rate for testing hypotheses about GWR model coefficients is controlled at ξ_m or less by selecting

$$\alpha = \frac{\xi_m}{1 + p_e - \frac{p_e}{np}}. \tag{5}$$

In most applications p_e will be much less than the maximum number of parameters (np) and so a significant gain in statistical power is expected when using this approach.

4. Summary

The main result in this paper equation 5, provides a very simple method of preserving the family–wise error rate when testing multiple hypotheses about GWR model coefficients. It avoids the large sacrifice of statistical power associated with the ordinary Bonferroni correction by incorporating the intrinsic dependency between the local models into the procedure. We note that the dependency parameter is a function of the effective number of parameters which is calculated from the trace of the hat matrix which itself is a function of the estimated bandwidth. Since the bandwidth is calculated from the data it is really a random variable with an unknown distribution. Therefore D and consequently α are also random variables with unknown distributions. If this distribution can be estimated (via bootstrap or Monte–Carlo methods for example), confidence bounds could be developed for D and α which would enhance the procedure and provide some protection against an inflated family-wise error resulting from overestimating D. We leave this for future work.

5. Acknowledgements

The first author thanks the National Centre for Geocomputation for providing resources during his time there working on this research and La Trobe University for funding study leave and travel costs.

References

Abdi, H. (2007). The Bonferonni and Šidàk corrections for multiple comparisons. In: N. Salkind, ed., *Encyclopedia of Measurement and Statistics*, Sage.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.

Bonforroni, C.E. (1935). Il calcolo delle assicurazioni su gruppi di teste. In: *Studi in Onore del Professore Salvatore Ortu Carboni*, Rome, Italy. pp. 13–60.

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1999). Some notes on parametric soignificance tests for geographically weighted regression. *Journal of Regional Science*, 39(3):497–524.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, Chichester.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.

Moyè, L.A. (2003). *Multiple Analyses in Clinical Trials: Fundamentals for Investigators*. Springer.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633.