Testing regression models based on sample survey data

Muhammad Ishaq Bhatti

This thesis considers several aspects of hypothesis testing problems associated with the linear regression model applied to multi-stage sample-survey data. Sample-surveys are increasingly being used as a tool by economists, businessmen and statisticians who need to gather data before making an important decision or giving much needed advice. There is an increasing tendency to perform regression analysis using sample-survey data in Market Research, Econometrics, Social and Biological Sciences, Geology and Geography. The statistical universe in these fields, as indeed in many others, while conceived as a whole is often comprehended as constituted of a number of natural blocks. In such cases, it is more useful as well as procedurally more practical to collect data in terms of blocks. For example, the population of the country falls naturally into the states, cities, districts, subdistricts, or urban and rural blocks. Similarly a single industry is often perceived as a complex of blocks and subblocks, each are concerned with well defined process of manufacture.

When regression analysis is used to analyse data obtained from such blocks, sub-blocks or groups, the residuals are often found to be correlated within ultimate blocks. This correlation is sometimes constant across blocks and its coefficient is called the intrablock or equicorrelation coefficient. This equicorrelation in the error term is also called a block effect. The main aim of this thesis is to develop and investigate tests for these effects.

This study is divided into three parts. In part I, the literature survey of Chapter 2 reveals our choice of using a model-based approach in the analysis of survey data. It suggests a unified regression model which could be equally useful for survey statisticians and sometimes for panel-data econometricians. It also summarises the theory of point optimal testing and suggests how we can apply optimal tests to our unified regression model in the subsequent chapters. In Chapters 3 and 4 we develop and investigate the small-sample power properties of some tests for equicorrelation in Standard Symmetric Multivariate Normal (SSMN) and Semmetric Multivariate Normal (SMN) distributions,

respectively. The SMN distribution is a multivariate distribution in which all components of a k dimensional random vector, y, have equal means, equal variances and all covariances between components take the same value. These common covariances give rise to a common equicorrelation coefficient, ρ. In this part of the thesis, we develop some optimal tests for two- and three-stage SSMN and SMN distributions for detecting intrablock and intrasubblock equicorrelation among blocked or subblocked observed random vectors. Some of the findings reported in Chapter 3 have appeared in [1].

Part II extends Part I to linear regression models in which the disturbances follow two-, three- and higher-stage SSMN and SMN distributions. In this part of the thesis, an attempt is made to find an "optimal" exact test for the intrablock equicorrelation for the two-stage linear regression model (see [2]) and then generalised to a multi-stage linear regression model by using King's Point-Optimal Invariant (POI) approach [6]. We also explore the possibility of the Locally Most Mean Powerful Invariant (LMMPI) tests by using King and Wu's approach from [7].

In part III of the thesis we consider Cox and Solomon's model (see [4, 5]) and construct a locally best invariant test, POI test and LMMPI test (see [3]). Throughout this thesis we have used the principle of invariance to eliminate the nuisance parameters where possible thus reducing the dimension of the testing problem.

REFERENCES

[1] M.I. Bhatti and M.L. King, 'A beta-optimal test for equicorrelation coefficient', *Austral. J. Statist.* **32** (1990), 87–97.

[2] M.I. Bhatti, 'Optimal testing for block effects in regression models', The 3rd Pacific Area Statistical Conference, Tokyo, Japan (1991), 401–404.

[3] M.I. Bhatti, 'Optimal testing for serial correlation in a large number of small samples', *Biometrical J.* **34** (1992), 57–67.

[4] D.R. Cox and P.J. Solomon, 'Analysis of variability with large numbers of small samples', *Biometrika* **73** (1988), 543–545.

[5] D.R. Cox and P.J. Solomon, 'On testing for serial correlation in large numbers of small samples', *Biometrika* **75** (1988), 145–148.

[6] M.L. King, 'Towards a theory of point optimal testing', *Econometric Rev.* **6** (1987), 169–218.

[7] M.L. King and X.P. Wu, 'Locally optimal one-sided tests for multiparameter hypothesis', Presented at the Australian meeting of the Econometric Society, University of New England, Armidale NSW, Australia.

Faculty of Asian and International Studies
Griffith University
Nathan Qld 4111
Australia