

## **ABSTRACT**

Evangelia Mitsaki

### **Ridge regression analysis of collinear data**

Multicollinearity is a commonly occurring problem in regression analysis. It is the situation in which two or more explanatory variables are strongly (but not perfectly) correlated to one other, making it difficult to interpret the strength of the effect of each variable. This thesis deals with the theory of multicollinearity as well as with ways that have been proposed to detect and correct it. In order to cope with collinear data we present several remedial measures such as principal components, variable selection and biased estimation.

The focus of the thesis is on ridge regression. Since the seminal work of Hoerl and Kennard ridge regression has proven to be a useful technique to tackle the multicollinearity problem in the linear regression model. The thesis presents the ridge estimator (ordinary and generalized) and its properties and also ways for selecting the ridge constant. Different interpretations of ridge regression are also discussed as well as applications of ridge regression to cases other than multiple linear regression. A recent advance concerning influence analysis is also presented. Illustrative examples are given where necessary.

In order to demonstrate the application of the ridge regression model to data, Monte Carlo simulations will be primarily used. The simulations are intended to give some insight to the behaviour of the ridge estimators, i.e. to compare their characteristics and performance.