

CONSTRUCTION OF ORDINARY REGRESSION TEST WITH PRINCIPAL COMPONENT REGRESSION UNDER COLLINEARITY¹

Yifan Xia, Dejin Liu and Yingsheng Su

Statistics School, Southwestern University of Finance and Economics
Chengdu 611130, P. R. China

xiayifan@swufe.edu.cn, liudejinldj@163.com, suys@swufe.edu.cn

Abstract. The multi-collinearity is one of the main problems we have to face in multi-variate analysis such as ordinary least square(OLS) regression. One solution to this problem is principal component regression(PCR). However, there are still some defects with PCR, the uneasy explanations of the PCR coefficient estimations as well as the derivation of significance tests for the ordinary regression coefficient estimators. This paper studies the relations between ordinary regression and PCR. In different conditions, the ordinary coefficient estimator t -tests are formulated by the PCR estimator t -test, and the statistical properties of the estimator based on PCR are discussed. Some numerical examples are also provided to support our theoretical results.

key words:Multi-collinearity; Principal component regression(PCR); Ordinary least square(OLS) estimator; Significance test of regression coefficient

1 Introduction

Consider a classic regression problem, the regression coefficients are often estimated by ordinary least square(OLS) method. However, multi-collinearity among data usually leads to sick coefficient estimators. Since the variances of the estimators are too large, the accuracy and significance of the estimation are greatly influenced, which has been discussed in many papers such as [1]-[4]. Because of correlation among predictors, multi-collinearity is so often in regression that it has been one of the problems must be settled in regression. (see Belsley [1])

Among the existing methods in literature, many proposals were raised to solve multi-collinearity (e.g. R.X. Liu[5], Foucart[6] and Mason[2]). These methods either mainly eliminated collinearity by reducing the dimensions of the predictors (for examples, to delete some predictors which led to collinearity, or to take fewer variable combinations); or did by finding an alternative biased estimator which was

¹This work was supported in part by Sichuan Science and Technology Project (No. 2014JY0210) and Sichuan Social Science Program Project (No. SC14B091). Corresponding author: Yifan Xia, suys@swufe.edu.cn