Analysis

In this section we will discuss the processes used to perform each of the five regression methods: Ordinary Least Squares, Ridge regression, LASSO, PCR, and PLS. Each regression model was fitted on the training data and we identified the best model. Using this best model we fit it again on the testing data and calculated the MSE value. Finally, we fit the best model again on the entire dataset for comparison.

Ridge Regression

```
# Fit Model

ridge_mod <- cv.glmnet(x, y, intercept = FALSE, standardize = FALSE, lambda = grid, alpha = 0)

# Find best model

ridge_best_labmda <- ridge_mod$lambda.min

# Find MSE on test set

ridge_pred <- predict(ridge_mod,
                      s= ridge_best_labmda,
                      newx= model.matrix(Balance ~ ., data = data.frame(credit_test)))

ridge_mse <- mean((ridge_pred - credit_test[,12])^2)

# Refit on full dataset

ridge_fit_full <- cv.glmnet(model.matrix(Balance ~ .,data = data.frame(scaled_credit)),
                            data.frame(scaled_credit)$Balance,
                            intercept = FALSE,
                            standardize = FALSE,
                            alpha = 0)
```

The MSE for the Ridge model was found to be 4