Regularized Panel Regression

Karl Shutes

January 30, 2013

Abstract

This paper considers the impact of using the regularisation techniques for the analysis of panel data. This gives rise to a number of useful analyses and approaches not considered elsewhere in the literature.

1 Introduction & Motivation

The use of regularisation in econometrics is far from widespread. The problem of variable selection is commonly side-stepped with legitimate appeals to theoretical frameworks. This paper extends this to consider situations where theory is not prescriptive and into situations where one might be tempted into using hypothesis tests to determine the independent variables in one's analyses. The use of these machine learning techniques is far from a carte blanche for mindless data mining. The use and selection of relevant data is still driven by theoretical foundations. However it sometimes informative to ascertain which variables are driving the underlying relationships and thus the problem of variable selection continues to exist.

Further the growth of data availability means that it is not uncommon to face a situation where $p << n$. This means that the standard approach of ordinary least squares is not feasible without some form of variable selection. The literature on the use and abuse of stepwise regression is significant. The situation of 'excessive data' can be dealt with by the regularised regressions, such as the LASSO and elastic net, for example Zou and Hastie [2005] & Zou [2005] where it is possible to have more independent variables than observations, unlike the situation in standard OLS. It is not uncommon that approaches such as the Aikake or Schwartz Information criteria are used in the variable selection problem (Akaike [1974] amongst others). These are not without their problems, though perhaps not as many as the frequently used stepwise regression techniques. These can further be contrasted with subset regressions, which take the various permutations of individual variables leading to the empty boast of 'I have run a million regressions'.

2 Literature Review & Definitions

Within the econometric literature, regularisation has a limited history. Ridge regression is perhaps the best known example (for example Hoerl and Kennard [1970]), where the problem of multicollinearity is dealt with by the imposition of a constraint on the coefficients of the regressions. This estimator is known to be biased however it is the case that the approach gives estimators with lower standard errors. The penalised function for the estimation is given by:

$$\begin{aligned} \beta_R &= \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum \beta_j^2 \\ &= \arg\min_{\beta} \sum_{i} \left(Y_i - X_i \beta^T\right)^2 + \lambda \beta^T \beta \\ &= \left(X^T X + \lambda I\right)^{-1} X^T y \end{aligned} \tag{1}$$

This approach does not perform any form of variable selection as, although it does shrink coefficients, it does not shrink them to 0. the λ parameter acts as the shrinkage control with $\lambda = 0$ being no shrinkage and therefore ordinary least squares. This can be compared to the Least Absolute Shrinkage & Selection Operator (LASSO). In this case the penalty is based on the ℓ_1 norm rather than the ℓ_2 norm of the ridge approach. Hence the problem becomes:

$$\begin{aligned} \beta_R &= \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum \mid \beta_j \mid \\ &= \arg\min_{\beta} \sum_{i} \left(Y_i - X_i^T \beta\right)^2 + \lambda \mid \beta \mid \end{aligned} \tag{2}$$

This is clearly shown graphically when considering two parameter estimates, with the LASSO (black) and ridge (red). The estimator loss functions are shown as ellipses. The point of tangency are the estimates for each technique. The LASSO shrinks β_1 to 0, whereas the ridge regression approaches it. The OLS estimator is given as $\hat{\beta}$. The parameter λ controls the amount of penalty applied to the parameters for the LASSO. Fu and Knight [2000] show that under certain regularity conditions,

There is a generalisation such that the γ-th norm is used. This is the bridge estimator. This has There are a number of similarities between the bridge estimator with $1 < \gamma < 2$[1] however the elastic net approach has non-differentiable corners at the axes (Hastie et al. [2008]). This therefore implies that the bridge regression, despite first impressions will not select variables unless $\gamma < 1$ in which case the penalty function is non-concave and the estimates may not be unique, though they may be set at zero.

[1] The limits here are LASSO and ridge regressions.

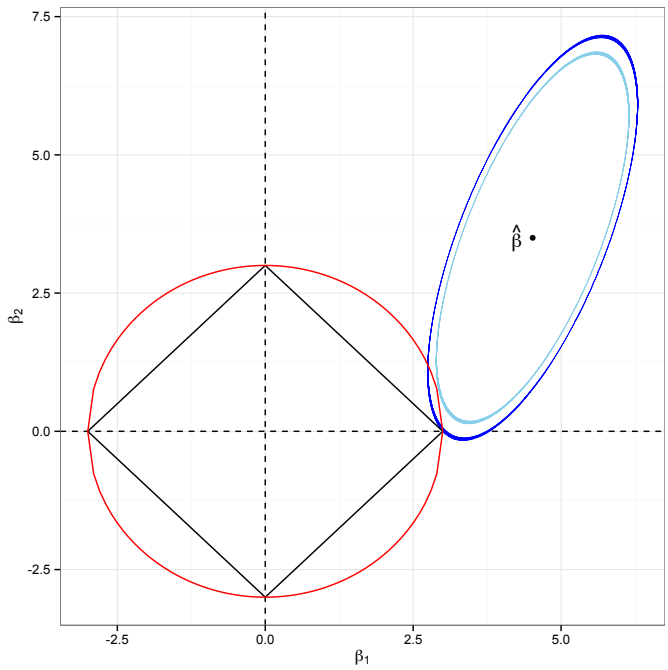

Figure 1: Differences Between LASSO & Ridge Regressions

3 Example

The LASSO type penalty approach to estimation and variable selection is given in Equation 2.

4 Conclusions

References

H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716 – 723, dec 1974. ISSN 0018-9286. doi: 10.1109/TAC.1974.1100705.

W. Fu and K. Knight. Asymptotics for lasso-type estimators. *Annals of Statistics*, 28 (5):1356–1378, 2000.

T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning; Data Mining, Inference & Prediction*. Springer Verlag, 2008.

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634. URL `http://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634`.

H. Zou. *Some Perspectives of Sparse Statistical Modeling.* PhD thesis, Standford University, 2005.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2): 301–320, 2005. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2005.00503.x. URL `http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x`.