A ridge to homogeneity

Stanislav Anatolyev
CERGE-EI, Czech Republic and New Economic School, Russia

The ridge regression estimator was originally devised as an anti-collinearity tool. It also has a property of trading off variance for non-zero bias and achieves lower estimation mean squared error than that of OLS estimation. The ridge regression estimator results from a penalized least squares problem with the L^2 penalization. Although the typical aim of ridge regression and L^2 shrinkage in general is to produce an estimator with a lower MSE, this ability carries over to the predictive MSE as well.

Some econometric models are heavily parameterized, with a subset of the parameter set representing the same feature which may (or may not) be common throughout this subset. In that case, researchers sometimes exploit either totally unrestricted or, conversely, tightly constrained versions of the model. For example, the unrestricted BEKK(1,1) model for multivariate volatility contains $2m^2$ parameters in their dynamic matrix coefficients, 2m in the diagonal BEKK(1,1) version, and only 2 in the scalar BEKK(1,1), where m is the number of assets under consideration. Similar models of this sort where tight exclusion restrictions are imposed is conditional autoregressive Wishart model and spatial multivariate GARCH. One can also recall the DECO model where the conditional correlation matrix is set to be equicorrelated, i.e. all its (1/2)m(m-1) distinct non-diagonal elements are set to an equal value. One more example is represented by a heterogeneous coefficients panel data model for which, despite the fact that the coefficient homogeneity across the units is often rejected, the homogeneous specification is most often adopted in practice, even though a variety of shrinkage methods for heterogeneous panel data models have been developed.

Usually, the ridge penalization is performed towards a zero target; a rarer case is a constant or random target. In this paper, we analyze choosing, as a target, the homogeneity restriction when it is relevant. The idea is to balance between unrestricted estimation (i.e., allowing full heterogeneity) and estimation under the commonality restriction (i.e. imposing full homogeneity). The L^2 penalized mean squared error criterion results in a ridge regression type estimator that has a closed form in linear models. In other contexts such as volatility models mentioned above, one potentially may use another L^2 penalized loss function. The penalty parameter of the ridge regression estimator is tuned to minimize the out-of-sample mean squared error, which in practice can be implemented using the leave-one-out cross-validation. Apart from the cross-validation stage, the proposed estimator is one-step, in contrast to an alternative shrinkage scheme where the shrinkage estimator is a weighted average between a fully unrestricted and fully restricted estimators. As our toy and prototype models indicate, there is a reduction in the predictive mean squared error which tends to increase with the dimensionality of the parameter set that is subject to shrinkage. We also work out, both theoretically and empirically, a heterogenous linear panel data setup and compare several estimators and corresponding confidence intervals.

Keywords: shrinkage, ridge regression, predictive mean squared error, cross-validation, heterogeneous panel