GMM-Lasso Solution Path

Zhentao Shi

How to efficiently computing the solution path of GMM Lasso? It is well known that LARS provides a solution path. Is there a counterpart for GMM-Lasso with a linear structural model? This is critical for data-driven tuning parameter determination and bootstrap.

For the criterion function

$$\sum_{i=1}^{n} g_i(\beta)' W \sum_{i=1}^{n} g_i(\beta)$$

with a positive definite matrix W. We can decompose it into

$$\sum_{i=1}^{n} h_i(\beta)' \sum_{i=1}^{n} h_i(\beta) = \sum_{l=1}^{L_n} \left(\sum_{i=1}^{n} h_{il}(\beta) \right)^2 = \sum_{l=1}^{L_n} \left[\tilde{h}_l(\beta) \right]^2$$

where $h_i(\beta) = W^{1/2} g_i(\beta)$ and $\tilde{h}_l(\beta) = \sum_{i=1}^{n} h_{il}(\beta)$. The farthermost part of the above equation resembles the criterion function of OLS.

To give a concrete example, we consider GMM-Lasso in the form of 2SLS. In this case $W = (Z'Z)^{-1}$ so that

$$\sum_{i=1}^{n} h_i(\beta) = W^{1/2} Z'(y - X\beta) = \left(Z'Z\right)^{-1/2} Z'(y - X\beta) = P(y - X\beta)$$

where $P = (Z'Z)^{-1/2} Z'$. This is as if we OLS with Py against PX.

In the asymptotics, the dimension $dim(Py) = L_n \to \infty$, therefore $\lambda_n \to 0$ as $n \to \infty$ has no problem. In reality, we need to use the information criterion or the cross-validation to decide.

Note that the dimension mismatch with OLS. This is as if we have L_n observations. But if we do bootstrap or cross-validation, the sample must be randomized or grouped over the dimension of n, not the dimension L_n. That means we cannot direct use the LARS cross-validation function in the R package LARS. Fortunately, randomize the sample and then apply the transformation is still straightforward.

There are several alternative ways to choose the tuning parameters. A thorough comparison with an additional empirical application will be interesting.