Matrix Uncertainty Selector with random noise

Alexandre B. Tsybakov (CREST-ENSAE)

We consider the linear regression model with observation error in the design:

$$
\begin{aligned}
y &= X\theta^* + \xi, \\
Z &= X + \Xi.
\end{aligned}
$$

Here the random vector $y \in \mathbf{R}^n$ and the random $n \times p$ matrix Z are observed, the $n \times p$ matrix X is unknown, Ξ is an $n \times p$ random noise matrix, $\xi \in \mathbf{R}^n$ is a random noise vector, and θ^* is a vector of unknown parameters to be estimated. We consider the setting where the dimension p can be much larger than the sample size n and θ^* is sparse. Because of the presence of the noise matrix Ξ, the commonly used Lasso and Dantzig selector are unstable. An alternative procedure called the Matrix Uncertainty (MU) selector has been proposed in Rosenbaum and Tsybakov (2010) in order to account for the noise. The properties of the MU selector have been studied in Rosenbaum and Tsybakov (2010) for sparse θ^* under the assumption that the noise matrix Ξ is deterministic and its values are small. In this paper, we propose a modification of the MU selector when Ξ is a random matrix with zero-mean entries having the variances that can be estimated. This is, for example, the case when the entries of X are missing at random. Under these conditions, we suggest new estimators and show, both theoretically and numerically, that they achieve better accuracy than the original MU selector. This is a joint work with Mathieu Rosenbaum.