Han, Yuefeng; Tsay, Ruey S.
High-dimensional linear regression for dependent data with applications to nowcasting.

Summary: Recent research has focused on $\ell_1$ penalized least squares (Lasso) estimators for high-dimensional linear regressions in which the number of covariates $p$ is considerably larger than the sample size $n$. However, few studies have examined the properties of the estimators when the errors and/or the covariates are serially dependent. In this study, we investigate the theoretical properties of the Lasso estimator for a linear regression with a random design and weak sparsity under serially dependent and/or nonsubGaussian errors and covariates. In contrast to the traditional case, in which the errors are independent and identically distributed and have finite exponential moments, we show that $p$ can be at most a power of $n$ if the errors have only finite polynomial moments. In addition, the rate of convergence becomes slower owing to the serial dependence in the errors and the covariates. We also consider the sign consistency of the model selection using the Lasso estimator when there are serial correlations in the errors or the covariates, or both. Adopting the framework of a functional dependence measure, we describe how the rates of convergence and the selection consistency of the estimators depend on the dependence measures and moment conditions of the errors and the covariates. Simulation results show that a Lasso regression can be significantly more powerful than a mixed-frequency data sampling regression (MIDAS) and a Dantzig selector in the presence of irrelevant variables. We apply the results obtained for the Lasso method to nowcasting with mixed-frequency data, in which serially correlated errors and a large number of covariates are common. The empirical results show that the Lasso procedure outperforms the MIDAS regression and the autoregressive model with exogenous variables in terms of both forecasting and nowcasting.

MSC:
62J05 Linear regression; mixed models
62J07 Ridge regression; shrinkage estimators (Lasso)
62M10 Time series, auto-correlation, regression, etc. in statistics (GARCH)

Keywords:
consistency; forecasting; high-dimensional time series; Lasso; mixed-frequency data; model selection; nowcasting

Full Text: DOI