

HIGH-DIMENSIONAL PREDICTIVE REGRESSION WITH MIXED INTEGRATION ORDERS AND ENDOGENEITY

Alien Shreshta

Preprint no. 2026-07

Abstract

Predictive regression models are a standard tool for forecasting asset returns using lagged state variables. Existing approaches face three limitations: (1) they typically examine predictors individually rather than jointly, (2) they depend upon the order of integration of the state variables, which is often unknown, and (3) they do not adequately address endogeneity in high-dimensional settings. This dissertation proposes a unified error projection-based framework that handles multiple predictors with mixed integration orders without requiring prior knowledge of their stationarity or endogeneity status. Our methodology extends the weighted estimating equations (WEE) framework developed by Hong et al. (2024) for univariate regression and combines it with the error projection model of Amihud and Hurvich (2004) to handle endogeneity and mixed integration orders. We then apply the smoothly clipped absolute deviation (SCAD) penalty to this model, which handles high-dimensional settings as well as the dimension inflation caused by the error projection, creating the WEE-SCAD framework. We establish the oracle property of WEE-SCAD and examine its finite-sample performance through low- and high-dimensional simulations. Finally, we demonstrate the WEE-SCAD framework in practice by applying it to monthly excess returns on the S& P 500 from 1969 to 2021.