Variable Selection and Forecasting in High Dimensional Linear Regressions with Parameter Instability*

Alexander Chudik
Federal Reserve Bank of Dallas

M. Hashem Pesaran
University of Southern California, USA and Trinity College, Cambridge, UK

Mahrad Sharifvaghefi
University of Pittsburgh

September 24, 2022

Abstract

This paper is concerned with the problem of variable selection and forecasting in the presence of parameter instability. There are a number of approaches proposed for forecasting in the presence of time-varying parameters, including the use of rolling windows and exponential down-weighting. However, these studies start with a given model specification and do not consider the problem of variable selection, which is complicated by time variations in the effects of signals on target variables. In this study we investigate whether or not we should use weighted observations at the variable selection stage in the presence of parameter instability, particularly when the number of potential covariates is large. Amongst the extant variable selection approaches we focus on the recently developed One Covariate at a time Multiple Testing (OCMT) method. This procedure allows a natural distinction between the selection and forecasting stages. We establish three main theorems on selection, estimation post selection, and in-sample fit. These theorems provide justification for using the full (not down-weighted) sample at the selection stage of OCMT and down-weighting of observations only at the forecasting stage (if needed). The benefits of the proposed method are illustrated by Monte Carlo studies and empirical applications to forecasting monthly stock market returns and quarterly output growths.

Keywords: Time-varying parameters, high-dimensional, multiple testing, variable selection, Lasso, one covariate at a time multiple testing (OCMT), forecasting, monthly returns, Dow Jones
JEL Classifications: C22, C52, C53, C55

*We are grateful to George Kapetanios and Ron Smith for constructive comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System. This research was supported in part through computational resources provided by the Big-Tex High Performance Computing Group at the Federal Reserve Bank of Dallas. This paper in part was written when Sharifvaghefi was a doctoral student at the University of Southern California (USC). Sharifvaghefi gradefully acknowledges financial support from the Center for Applied Financial Economics at USC.