**Appendix. A**

1. **The Shrinkage Methods**

The family of shrinkage methods take the form of penalized regression in order to improve ordinary least squares (OLS) regression. They are similar to an OLS regression in that the objective is to minimize the residual sum of squares (RSS), but they add a term that imposes a size constraint on the coefficient estimates.

The different models among these methods are distinguished by the penalty function , which regularizes the coefficient estimates and shrinks the coefficients of variables with less explanatory power. The shrinkage penalty term depends on the tuning parameter , which regulates the amount of shrinkage imposed on the coefficients; a higher results in a stronger shrinkage of the regression coefficients, while would reduce the model to an OLS with no shrinkage.

**1.1 Least Absolute Shrinkage and Selection Operator (LASSO)**

LASSO was proposed by Tibshirani (1996), in which the penalty function is given as

Compared to the primitive shrinkage method, ridge regression of Hoerl & Kennard (1970), the penalty of LASSO is able to shrink the less relevant variables to exactly zero via soft thresholding and thus presents the feature of variable selection. Moreover, due to the absolute value operator in the penalty term, LASSO does not have a closed form solution and is computed through algorithmic methods.

* 1. **Adaptive LASSO**

Consistency of variable selection by LASSO is achieved only under strict conditions, and the adaptive LASSO was proposed by Zou (2006) to overcome this issue. The penalty term includes a weighting parameter that is derived from a first-step estimation. The penalty function is given as

where adaptive weights are used for penalizing different coefficients in the LASSO penalty. Adaptive LASSO can be solved through the same efficient algorithm for solving LASSO.

* 1. **Elastic Net**

The elastic net is a compromise between ridge regression and LASSO. While it retains the variable selection feature of LASSO, it also shrinks the coefficients of correlated variables towards each other like the ridge regression. The penalty function takes the form of a weighted mean of ridge and LASSO penalties.

The elastic net includes the special cases of LASSO() and ridge regression (). In this paper the parameter is set to be 0.5. Moreover, an adaptive version of elastic net is also considered which includes adaptive weights as in adaptive LASSO.

1. **Complete Subset Regression (Elliott, 2011)**

Another possible approach to handling a high-dimensional dataset is subset selection for linear regression. While there are a number of strategies for subset selection, testing all possible combinations of predictor variables is computationally demanding and becomes infeasible when the number of candidate variables is very large.

Complete subset regression (CSR) proposed by Elliott et al. (2013, 2015) takes an ensemble approach. For a given set of potential regressors, CSR combines forecasts from all possible linear regression models while keeping the number of predictors fixed. For a dataset with K possible regressors, the number of k-variate models () is . The set of models for a fixed value k is referred to as a complete subset, and the final forecast by CSR is the equal-weighted average of forecasts from all models within the complete subset indexed by k.

1. **Target Factors (Bai & Ng, 2008)**

Numerous forecasting methodologies using factor augmented models has been developed recently. The idea of these factor models is to first estimate the factors from a large number of predictors using the method of principal components, and to augment these factors to a linear forecasting equation. In order to refine the factor augmented forecasting methodology, Bai & Ng (2008) proposed targeting the predictors using hard and soft thresholding rules. The underlying rationale is that computing the principal components from all predictors may result in noisy factors, and that only the predictors with high forecasting power should be used.

In this paper the hard thresholding method as suggested in Bai & Ng (2008) and as implemented by Medeiros et al (2019) is used.

Let be the dependent variable or the logarithm of VIX, (i=1,…,q) the candidate predictors and a set of controls. Following Bai & Ng (2008), the lagged values of and a constant are used as .

1. For i=1,…, q, perform a regression of on and and compute the t-statistics corresponding to the coefficient of .
2. Choose a significance level and find the set of significant variables based on the computed t-statistics.
3. Estimate the Factors from .
4. Regress on and where and the number of factors in is decided using BIC.