Federal Reserve Bank of Dallas

Variable Selection and Forecasting in High Dimensional Linear Regressions with Structural Breaks

Alexander Chudik, M. Hashem Pesaran and Mahrad Sharifvaghefi

Globalization Institute Working Paper 394 **August 2020**

Research Department

https://doi.org/10.24149/gwp394

Working papers from the Federal Reserve Bank of Dallas are preliminary drafts circulated for professional comment. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

Variable Selection and Forecasting in High Dimensional Linear Regressions with Structural Breaks*

Alexander Chudik†, M. Hashem Pesaran‡ and Mahrad Sharifvaghefi§

July 23, 2020

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 breaks, including the use of rolling windows or exponential down-weighting. However, these studies start with a given model specification and do not consider the problem of variable selection. It is clear that, in the absence of breaks, researchers should weigh the observations equally at both the variable selection and forecasting stages. In this study, we investigate whether or not we should use weighted observations at the variable selection stage in the presence of structural breaks, 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 that allows a natural distinction between the selection and forecasting stages, and provide theoretical justification for using the full (not down-weighted) sample in 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 empirical applications to forecasting output growths and stock market returns.

Keywords: Time-varying parameters, structural breaks, high-dimensionality, multiple testing, variable selection, one covariate at a time multiple testing (OCMT), forecasting

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.

† Alexander Chudik, Federal Reserve Bank of Dallas.

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

§ Corresponding author: Mahrad Sharifvaghefi, University of Southern California, Kaprielian (KAP) Hall 359, 3620 South Vermont Ave, Los Angeles, CA 90089. Email: sharifva@usc.edu.

1 Introduction

"*When you have eliminated the impossible, whatever remains, however improbable, must be the truth*" Sir Arthur Conan Doyle, *The Sign of the Four* (1890)

There is mounting evidence that models fitted to many economic relationships are subject to structural breaks. In an extensive early study, Stock and Watson (1996) find that a large majority of time series regressions are subject to structural breaks. Clements and Hendry (1998) consider parameter instability to be one of the main sources of forecast failure. This problem has been addressed at the estimation stage given a set of selected regressors. However, the issue of variable selection in the presence of structural breaks is still largely underdeveloped. In this study, we investigate whether or not we should use weighted observations at the variable selection stage in the presence of structural breaks, particularly when the number of potential covariates is large. We provide theoretical arguments in favor of using the full (unweighted) sample at the selection stage and consider weighted observations post selection only at the estimation stage and forecasting stages. The benefits of the proposed method are illustrated by empirical applications to forecasting cross-country output growths and stock market returns.

Studies on structural breaks at the estimation stage usually assume a number of different model specifications that allow for parameter instability. Typical solutions are either to use rolling windows or exponential down-weighting. For instance, Pesaran and Timmermann (2007), Pesaran and Pick (2011) and Inoue et al. (2017) consider the choice of an observation window, and Hyndman et al. (2008) and Pesaran et al. (2013), respectively consider exponential and non-exponential down-weighting of observations. There are also Bayesian approaches to prediction that allow for a possibility of breaks over the forecast horizon, e.g. Chib (1998), Koop and Potter (2004), and Pesaran et al. (2006). Rossi (2013) provides a review of the literature on forecasting under instability. There are also related time varying parameter (TVP) and regime switching models that are used for forecasting. See, for example, Hamilton (1988) and Dangl and Halling (2012). All these studies take the model specification as given and then consider different ways of modeling and allowing for parameter instability. But, to the best of our knowledge, none of these studies considers the problem of variable selection in the presence of structural breaks.

In the absence of breaks, observations should be weighted equally for both variable selection and estimation purposes. Yet, in the presence of breaks, the literature does not discuss whether or not weighted observations should be used at the variable selection stage, particularly when the number of potential covariates is large. There are a number of recent studies that consider predicting stock returns using penalized regression, especially the Least

Absolute Shrinkage and Selection Operator (Lasso) initially proposed by Tibshirani (1996) - for example, Caner and Knight (2013), and Koo et al. (2020). But they do not allow for instability at the Lasso stage and suggest recursive application of Lasso using rolling windows. Lee et al. (2016) have proposed a Lasso procedure that allows for a threshold effect. Kapetanios and Zikes (2018) have proposed a time-varying Lasso procedure, where all the parameters of the model vary locally. These are interesting extensions of Lasso, but are likely to suffer from the over-fitting problem, and could be sensitive to how cross validation is carried out. Also recently, Yousuf and Ng (2019) propose an interesting boosting procedure for the estimation of high-dimensional models with locally time varying parameters. It is important to note that, in the case of both penalized regression and boosting procedures, variable selection and estimation are carried out in one stage.

Recently, Chudik et al. (2018) propose an alternative procedure called **o**ne **c**ovariate at a time **m**ultiple **t**esting (OCMT). In the absence of structural breaks, the authors establish that the suggested procedure asymptotically selects all the relevant covariates and none of the pure noise covariates under general assumptions. Moreover, they show that the estimation errors of coefficients and prediction loss converge to zero. Finally, their Monte Carlo studies indicate that OCMT tends to perform better than penalized regression or boosting procedures under various designs. Recently, Sharifvaghefi (2020) has generalized the OCMT procedure to allow the covariates under consideration to be highly correlated, while penalized regression methods require the covariates to be weakly correlated (see e.g. Zhao and Yu (2006)). One clear advantage of OCMT is its natural separation of the two problems of variable selection and estimation/forecasting. One can, therefore decides whether to use the weighted observations at the variable selection stage or not. In this paper, we provide a theoretical argument that the full (unweighted) sample should be used at the variable selection stage using the OCMT procedure. Existing theoretical results from the forecasting literature can then be applied to the post OCMT selected model to test for breaks and decide on the optimal choice of the estimation window or down-weighting.

We consider three applications related to forecasting output growth and stock market returns to illustrate the benefits of the OCMT procedure with full unweighted sample at the selection stage. Our results consistently suggest that using down-weighted observations at the selection stage of the OCMT procedure worsens forecast accuracy in terms of mean square forecast error and mean directional forecast accuracy. Moreover, our results suggest that overall OCMT with no down-weighting at the selection stage outperforms penalized regression methods, such as Lasso and/or Adaptive Lasso, which are prone to the over-fitting problem.

The rest of the paper is organized as follows: Section 2 sets out the model specification. Section 3 explains the basic idea behind using the OCMT procedure with no down-weighting

for variable selection in the presence structural breaks. Section 4 discusses the technical assumptions and the asymptotic properties of the OCMT procedure under parameter instability. Section 5 presents the empirical applications, and Section 6 concludes. Mathematical proofs and the related lemmas are provided in the Appendix.

Notations: Generic finite positive constants are denoted by C_i for $i = 1, 2, \cdots$. They can take different values in different instances. $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$ denote the spectral and Frobenius norms of matrix $\mathbf{A}$, respectively. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequence of positive real numbers, then $f_n = O(g_n)$, if there exists a positive constant C_0 such that $|f_n|/g_n \leq C_0$ for all n. $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $f_n = \ominus(g_n)$ if there exist $n_0 \geq 1$ and positive constants C_0 and C_1, such that $\inf_{n \geq n_0}(f_n/g_n) \geq C_0$ and $\sup_{n \geq n_0}(f_n/g_n) \leq C_1$.

2 Model Setting

Consider the following Data Generating Process (DGP) for the dependent variable, y_t,

$$y_t = \mathbf{z}_t' \mathbf{a}_t + \sum_{i=1}^{k} \beta_{it} x_{it} + u_t, \text{ for } t = 1, 2, \cdots, T \tag{1}$$

with time-varying parameters, $\mathbf{a}_t = (\mathrm{a}_{1t}, \mathrm{a}_{2t}, \cdots, \mathrm{a}_{mt})'$ and $\{\beta_{it}, i = 1, 2, ..., k\}$, where $\mathbf{z}_t$ is an $m \times 1$ vector of conditioning variables, x_{it} for $i = 1, 2, \cdots, k$ are the covariates, and u_t is an error term. Since the parameters are time-varying we refer to the covariate i as "*signal*" if the average value of its coefficient, $\bar{\beta}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(\beta_{it})$, does not tend to zero very fast, i.e. $\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})$ for some $0 \leq \vartheta_i < 1/2$. We also assume the coefficients ($\mathbf{a}_t$, and β_{it}, for $i = 1, 2, ..., k$) are independently distributed of the conditioning variables ($\mathbf{z}_t$) and the covariates (x_{it}).

Parameters can vary continuously following a stochastic process as in the standard random walk time-varying parameter model,

$$\beta_{it} = \beta_{i,t-1} + \sigma_{it} \epsilon_{it},$$

or could be fixed and change at discrete intervals as in the standard structural break model:

$$\beta_{it} = \begin{cases} \beta_i^{[1]}, & 1 \leq t < T_1, \\ \beta_i^{[2]}, & T_1 \leq t < T_2, \\ & \vdots \\ \beta_i^{[s]}, & T_{s-1} \leq t < T. \end{cases}$$

The vector $\mathbf{z_t}$ can contains deterministic components such as a constant, dummy variables,

and a deterministic time trend as well as stochastic variables including common factors and lagged values of y_t. The problem is that both the structure of the breaks and the identity of the k signals are unknown. The task facing the investigator is to select the signals from a set of covariates under consideration, $S_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$, known as the active set, with N, the number of covariates in the active set, possibly much larger than T, the number of data points available for estimation prior to forecasting.

The application of penalized regression techniques to variable selection is theoretically justified under two key parameter stability assumptions: the stability of β_{it} and the stability of the correlation matrix of the covariates in the active set. Under these assumptions, the application of the penalized regression to the active set can proceed using the full sample without down-weighting or separating the variable selection from the forecasting stage. However, in the presence parameter instability, it is not clear how the use of penalized regressions could be justified. The problem has been recognized in the empirical literature focusing on slowly varying parameters and/or the use of rolling windows without making a distinction between variable selection and forecasting. It is also worth highlighting that in this paper, we relax the assumption of fixed correlation among the covariates in the active set, which is very common in the penalized regression studies, and allow for time-varying correlations.

In this paper we follow Chudik et al. (2018) and consider the application of the OCMT procedure for variable selection stage using the full unweighted sample, and provide theoretical arguments to justify such an approach. We first recall that OCMT's variable selection is based on the net effect of x_{it} on y_t conditional $\mathbf{z}_t$. However, when the regression coefficients and/or the correlations across the covariates in the active set are time-varying, the net effects will also be time-varying and we need to base our selection on average net effects. Also, we need to filter out the effects of the conditioning variables, $\mathbf{z}_t$, from x_{it} and y_t, before defining average net effects. To this end consider the following auxiliary regressions of x_{it} and y_t on $\mathbf{z}_t$:

$$\tilde{y}_t = y_t - \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{y,T}, \text{ and } \tilde{x}_{it} = x_{it} - \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{i,T} \tag{2}$$

where $\bar{\boldsymbol{\psi}}_{y,T}$ and $\bar{\boldsymbol{\psi}}_{i,T}$ are the $m \times 1$ vectors of projection coefficients defined by

$$\bar{\boldsymbol{\psi}}_{y,T} \equiv \left(T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right)\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t y_t)\right),$$

and

$$\bar{\boldsymbol{\psi}}_{i,T} \equiv \left(T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right)\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t x_{it})\right).$$

Given the filtered series, $\tilde{x}_{it}$ and $\tilde{y}_t$, we now define the average net effect of covariate x_{it} on

y_t, conditional $\mathbf{z}_t$, as

$$\bar{\theta}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}\tilde{y}_t). \tag{3}$$

Substituting for $\tilde{y}_t = y_t - \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{y,T}$ in the above and noting that $\bar{\theta}_{i,T}$ is a given constant, then

$$\bar{\theta}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}y_t) - \bar{\boldsymbol{\psi}}_{y,T}' \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}\mathbf{z}_t) \right].$$

Also,

$$T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}\mathbf{z}_t) = 0, \tag{4}$$

which follows since

$$\begin{aligned} T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}\mathbf{z}_t) &= T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it}\mathbf{z}_t) - \bar{\boldsymbol{\psi}}_{i,T}' \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{z}_t\mathbf{z}_t') \right] \\ &= T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it}\mathbf{z}_t) - T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it}\mathbf{z}_t) = \mathbf{0}. \end{aligned}$$

Hence $\bar{\theta}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}y_t)$. Now by substituting y_t from (1) we can further write,

$$\begin{aligned} \bar{\theta}_{i,T} &= T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}y_t) = T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[\tilde{x}_{it} \left(\mathbf{z}_t'\mathbf{a}_t + \sum_{j=1}^{k} \beta_{jt}x_{jt} + u_t \right) \right] \\ &= T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{a}_t')\mathbb{E}(\tilde{x}_{it}\mathbf{z}_t) + T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{k} \mathbb{E}(\beta_{jt})\mathbb{E}(\tilde{x}_{it}x_{jt}) + T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}u_t) \\ &= \mathbf{a}'T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}\mathbf{z}_t) + T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{k} \mathbb{E}(\beta_{jt})\mathbb{E}(\tilde{x}_{it}x_{jt}) + T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}u_t) \\ &= \sum_{j=1}^{k} \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}(\beta_{jt})\mathbb{E}(\tilde{x}_{it}x_{jt}) \right] + T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}u_t). \end{aligned}$$

Finally, using (2) to replace $x_{jt} = \tilde{x}_{jt} + \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{j,T}$ in the first term of the above and noting from (4) that $T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}\mathbf{z}_t) = 0$, then the average net effect can be written simply as

$$\bar{\theta}_{i,T} = \sum_{j=1}^{k} \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}(\beta_{jt})\sigma_{ij,t}(\mathbf{z}) \right] + \bar{\sigma}_{iu,T}(\mathbf{z}), \tag{5}$$

where $\sigma_{ij,t}(\mathbf{z}) = \mathbb{E}(\tilde{x}_{it}\tilde{x}_{jt})$, and $\bar{\sigma}_{iu,T}(\mathbf{z}) = T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it}u_t)$. Also

$$\bar{\sigma}_{iu,T}(\mathbf{z}) = T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it}u_t) - \bar{\boldsymbol{\psi}}_{i,T}' \left(T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{z}_t u_t) \right),$$

which will be identically zero if the covariates and the conditioning variables are weakly exogenous with respect to u_t. In what follows we allow for a mild degree of correlation between $(x_{it}, \mathbf{z}_t)$ and u_t by assuming that $\bar{\sigma}_{iu,T}(\mathbf{z}) = \ominus(T^{-\epsilon_i})$, for some $\epsilon_i > 1/2$. It is also easily seen that when the parameters and the cross covariate covariances are time-invariant the above average net effect reduces to $\theta_i = \Sigma_{j=1}^{k} \mathbb{E}(\beta_j)\sigma_{ij}(\mathbf{z})$.

Given the average net effect of covariate i on y_t, the covariates in the active can be categorized into three groups: *signals*, *pseudo-signals* and *average noise variates*. As mentioned before, *signals* are those covariates with average value of their coefficient, namely $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{it})$, which does not approach zero quickly, i.e. $\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$. *Pseudo signals* are the covariates that do not enter the DGP but have average net effects, $\bar{\theta}_{i,T}$'s, that do not converge to zero sufficiently fast, i.e. $\bar{\theta}_{i,T} = \ominus(T^{-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$. Finally, *average noise variables* are those covariates that do not enter the DGP and at the same time have either zero or sufficiently small average net effects in the sense that $\bar{\theta}_{i,T} = \ominus(T^{-\epsilon_i})$, for some $\epsilon_i > 1/2$.

In what follows, we first present the OCMT procedure and then discuss the conditions under which the approximating model that include all the signals and none of the average noise variables can be selected by OCMT.

3 Parameter instability and OCMT

The OCMT procedure consider the following one at a time regression of y_t on each covariate x_{it}, $i = 1, 2, \cdots, N$, conditional on $\mathbf{z}_t$:

$$y_t = \boldsymbol{\lambda}_i'\mathbf{z}_t + \phi_{i,T}x_{it} + \eta_{it}, \text{ for } t = 1, 2, \cdots, T, \tag{6}$$

where

$$\phi_{i,T} = \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\tilde{x}_{it}^2\right)\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\tilde{x}_{it}\tilde{y}_t\right)\right] = \left[\bar{\sigma}_{ii,T}(\mathbf{z})\right]^{-1}\bar{\theta}_{i,T},$$

with $\bar{\sigma}_{ii,T}(\mathbf{z}) = T^{-1}\sum_{t=1}^{T}\sigma_{ii,t}(\mathbf{z})$. Chudik et al. (2018) assume parameter stability, and set $\beta_{it} = \beta_i$ for all t, where β_i is deterministic, and zero conditional correlation between the signals and the error term, $\sigma_{iu,t} = 0$ for all t. In this case the average net effect can be simplified to net effect $\theta_{i,T} = \sum_{j=1}^{k}\beta_j\bar{\sigma}_{ij,T}(\mathbf{z})$ where $\bar{\sigma}_{ij,T}(\mathbf{z}) = T^{-1}\sum_{t=1}^{T}\sigma_{ij,t}(\mathbf{z})$ and hence

$$\phi_{i,T} = \frac{\theta_{i,T}}{\bar{\sigma}_{ii,T}(\mathbf{z})} = \frac{\sum_{j=1}^{k}\beta_j\bar{\sigma}_{ij,T}(\mathbf{z})}{\bar{\sigma}_{ii,T}(\mathbf{z})}. \tag{7}$$

However, in the more general set up of DGP (1), the net effect of x_{it} on y_t is time varying. Therefore, by running one at a time regression of y_t on each covariate x_{it}, $i = 1, 2, \cdots, N$, we focus on average net effect of x_{it} on y_t over the full sample as presented by $\bar{\theta}_{i,T}$ in equation (5).

Due to correlation between the covariates, knowing whether $\bar{\theta}_{i,T}$ is zero or not does not

necessarily allow us to establish whether $\bar{\beta}_{i,T}$ is zero or not. There are four possibilities:

(I) *Signals*	$\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})$ and $\bar{\theta}_{i,T} = \ominus(T^{-\vartheta_i})$
(II) *Hidden Signals*	$\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})$ and $\bar{\theta}_{i,T} = \ominus(T^{-\epsilon_i})$
(III) *Pseudo-signals*	$\bar{\beta}_{i,T} = \ominus(T^{-\epsilon_i})$ and $\bar{\theta}_{i,T} = \ominus(T^{-\vartheta_i})$
(IV) *Average noise variables*	$\bar{\beta}_{i,T} = \ominus(T^{-\epsilon_i})$ and $\bar{\theta}_{i,T} = \ominus(T^{-\epsilon_i})$

for some $0 \leq \vartheta_i < 1/2$ and $\epsilon_i > 1/2$. Notice, if the covariate x_{it} is an average noise variable, then $\bar{\theta}_{i,T}$, the average net effect of x_{it} on y_t, converges to zero very fast. Therefore, down-weighting of observations at the variable selection stage is likely to be inefficient for eliminating the average noise variables. Moreover, for a signal to remain hidden, we need the terms of higher order, $\ominus(T^{-\vartheta_j})$ with $0 \leq \vartheta_i < 1/2$, to *exactly* cancel out such that $\theta_{i,T}$ becomes a lower order, i.e. $\ominus(T^{-\varepsilon_i})$, that tends to zero at a sufficiently fast rate (with $\epsilon_i > 1/2$). This combination of events seem quite unlikely, and to simplify the theoretical derivations in what follows we abstract from such a possibility and assume that there are no hidden signals and focus on the first stage of the OCMT procedure for variable selection.[1]

First stage of OCMT procedure

1. *For $i = 1, 2, \cdots, N$, regress $\mathbf{y} = (y_1, y_1, ..., y_T)'$ on $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T)'$ and $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$; $\mathbf{y} = \mathbf{Z}\boldsymbol{\lambda}_i + \phi_{i,T}\mathbf{x}_i + \boldsymbol{\eta}_i$; and compute the t-ratio of $\phi_{i,T}$, given by*

$$t_{i,T} = \frac{\hat{\phi}_{i,T}}{s.e.\left(\hat{\phi}_{i,T}\right)} = \frac{\mathbf{x}_i'\mathbf{M}_\mathbf{z}\mathbf{y}}{\hat{\sigma}_i\sqrt{\mathbf{x}_i'\mathbf{M}_\mathbf{z}\mathbf{x}_i}},$$

   where $\hat{\phi}_{i,T} = \left(\mathbf{x}_i'\mathbf{M}_\mathbf{z}\mathbf{x}_i\right)^{-1}\left(\mathbf{x}_i'\mathbf{M}_\mathbf{z}\mathbf{y}\right)$ is the Ordinary Least Square (OLS) estimator of $\phi_{i,T}$, $\hat{\sigma}_i^2 = \hat{\boldsymbol{\eta}}_i'\hat{\boldsymbol{\eta}}_i/T$, and $\hat{\boldsymbol{\eta}}_i$ is a $T \times 1$ vector of regression residuals.

2. *Consider the critical value function, $c_p(N, \delta)$, defined by*

$$c_p(N, \delta) = \Phi^{-1}\left(1 - \frac{p}{2N^\delta}\right), \tag{8}$$

   where $\Phi^{-1}(.)$ is the inverse of a standard normal distribution function; δ is a finite positive constant; and p $(0 < p < 1)$ is the nominal size of the individual tests to be set by the investigator.

3. *Given $c_p(N, \delta)$, the selection indicator is given by*

$$\hat{\mathcal{J}}_i = I\left[|t_{i,T}| > c_p(N, \delta)\right], \text{ for } i = 1, 2, \cdots, N. \tag{9}$$

[1] To deal with possibility of existence of hidden signals, Chudik et al. (2018) extend the OCMT method to have multiple stages.

The covariates x_{it} is selected if $\hat{\mathcal{J}}_i = 1$.

When hidden signals are present, additional stages of OCMT procedure will be required to uncover them.

Higher stages of OCMT procedure

1. *Stages $j = 2, 3, ...,$ regress y_t on the conditioning variables $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T)'$, the variables selected in previous stages, and the remaining variables x_{it} in the active set, one at the time. Compute the associated t-ratios, denoted as $t_{i,(j),T}$.*

2. *Regressors for which $|t_{i,(j),T}| > c_p(n, \delta^*)$, are then added to the set of selected variables, where in general a larger value of $\delta^* > \delta$ is required at the higher stages of the OCMT procedure.*

3. *If no additional variables are selected in stage j, then the multi-stage OCMT procedure stops, otherwise one proceeds to the next stage by increasing j by one.*

The main goal of OCMT is to use the t-ratio of the estimated $\phi_{i,T}$ to select all the signals and none of the average noise variables, the selected model is referred to as an *approximating model* since it can include pseudo-signals. To deal with the multiple testing nature of the problem, the critical value of the tests, $c_p(N, \delta)$, is chosen to be an appropriately increasing function of N, the number of covariates in the active set.

4 Asymptotic Properties of OCMT Procedure under Parameter instability

We now provide the theoretical justification for using the OCMT procedure for variable selection in models with time-varying parameters. It is assumed that k, the number of signals, and m, the number of conditioning variables, are unknown fixed numbers. But we allow that the number of pseudo-signals, which we denote by k^*, to grow at a sufficiently slow rate relative to N and T. Finally, we define an *approximating model* to be a model that contains all the signals, $\{x_{it} : i = 1, 2, \cdots, k\}$, and none of the average noise variables, $\{x_{it} : k + k^* + 1, k + k^* + 2, \cdots, N\}$. Clearly, such a model can contain one or more of the pseudo-signals, $\{x_{it} : k + 1, k + 2, \cdots, k + k^*\}$. We start with some technical assumptions in Section 4.1 and then provide the asymptotic properties of the OCMT procedure under structural breaks in Section 4.2.

4.1 Technical Assumptions

In what follows we make use of the following filtrations: $\mathcal{F}^z_{\ell t} = \sigma(z_{\ell t}, z_{\ell,t-1}, \cdots)$ for $\ell = 1, 2, \cdots, m$, $\mathcal{F}^{\mathrm{a}}_{\ell t} = \sigma(\mathrm{a}_{\ell t}, \mathrm{a}_{\ell,t-1}, \cdots)$ for $\ell = 1, 2, \cdots, m$, $\mathcal{F}^x_{it} = \sigma(x_{it}, x_{i,t-1}, \cdots)$ for $i = 1, 2, \cdots, N$, and $\mathcal{F}^u_t = \sigma(u_t, u_{t-1}, \cdots)$. Moreover, we set $\mathcal{F}^z_t = \cup_{\ell=1}^m \mathcal{F}^z_{\ell t}$, $\mathcal{F}^{\mathrm{a}}_t = \cup_{\ell=1}^m \mathcal{F}^{\mathrm{a}}_{\ell t}$, $\mathcal{F}^x_t = \cup_{i=1}^N \mathcal{F}^x_{it}$, and $\mathcal{F}_t = \mathcal{F}^z_t \cup \mathcal{F}^{\mathrm{a}}_t \cup \mathcal{F}^x_t \cup \mathcal{F}^u_t$.

Assumption 1 (Martingale Difference Processes)

(i) $\mathbb{E}\left[z_{\ell t} z_{\ell' t} - \mathbb{E}(z_{\ell t} z_{\ell' t}) | \mathcal{F}_{t-1}\right] = 0$ *for* $\ell, \ell' = 1, 2, \cdots, m$ *and* $t = 1, 2, \cdots, T$.

(ii) $\mathbb{E}\left[x_{it} x_{jt} - \mathbb{E}(x_{it} x_{jt}) | \mathcal{F}_{t-1}\right] = 0$ *for* $i, j = 1, 2, \cdots, N$ *and* $t = 1, 2, \cdots, T$.

(iii) $\mathbb{E}\left[u_t^2 - \mathbb{E}\left(u_t^2\right) | \mathcal{F}_{t-1}\right] = 0$ *for* $t = 1, 2, \cdots, T$.

(iv) $\mathbb{E}\left[z_{\ell t} x_{it} - \mathbb{E}(z_{\ell t} x_{it}) | \mathcal{F}_{t-1}\right] = 0$ *for* $\ell = 1, 2, \cdots, m$, $i = 1, 2, \cdots, N$ *and* $t = 1, 2, \cdots, T$.

(v) $\mathbb{E}\left[z_{\ell t} u_t - \mathbb{E}(z_{\ell t} u_t) | \mathcal{F}_{t-1}\right] = 0$ *for* $\ell = 1, 2, \cdots, m$ *and* $t = 1, 2, \cdots, T$.

(vi) $\mathbb{E}\left[x_{it} u_t - \mathbb{E}(x_{it} u_t) | \mathcal{F}_{t-1}\right] = 0$ *for* $i = 1, 2, \cdots, N$ *and* $t = 1, 2, \cdots, T$.

(vii) $\mathbb{E}\left[\mathrm{a}_{\ell t} - \mathbb{E}(\mathrm{a}_{\ell t}) | \mathcal{F}_{t-1}\right] = 0$ *for* $\ell = 1, 2, \cdots, m$ *and* $t = 1, 2, \cdots, T$.

(viii) $\mathbb{E}\left[\beta_{it} - \mathbb{E}(\beta_{it}) | \mathcal{F}_{t-1}\right] = 0$ *for* $i = 1, 2, \cdots, k$ *and* $t = 1, 2, \cdots, T$.

Assumption 2 (Exponential Decaying Probability Tail)
There exist sufficiently large positive constants C_0 *and* C_1, *and* $s > 0$ *such that*

(i) $\sup_{\ell,t} \Pr(|z_{\ell t}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, *for all* $\alpha > 0$.

(ii) $\sup_{\ell,t} \Pr(|\mathrm{a}_{\ell t}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, *for all* $\alpha > 0$.

(iii) $\sup_{i,t} \Pr(|x_{it}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, *for all* $\alpha > 0$.

(iv) $\sup_{i,t} \Pr(|\beta_{it}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, *for all* $\alpha > 0$.

(v) $\sup_t \Pr(|u_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, *for all* $\alpha > 0$.

Assumption 3 (Coefficients of Signals)

(i) The number of signals, k, *is finite.*

(ii) β_{it} *for* $i = 1, 2, \cdots, k$ *are independent of* $x_{jt'}$, $j = 1, 2, \cdots, N$, $z_{\ell t'}$, $\ell = 1, 2, \cdots, m$, *and* $u_{t'}$ *for all* t *and* t'.

(iii) $\bar{\beta}_{i,T} \equiv T^{-1} \sum_{t=1}^T \mathbb{E}(\beta_{it}) = \ominus(T^{-\vartheta_i})$, *for some* $0 \leq \vartheta_i < 1/2$.

Assumption 4 (Coefficients of Conditioning Variables)

(i) The number of conditioning variates, m, is finite.

(ii) $\mathrm{a}_{\ell t}$ *for* $\ell = 1, 2, \cdots, m$ *are independent of* $x_{it'}$,$i = 1, 2, \cdots, n$, $z_{\ell' t'}$, $\ell' = 1, 2, \cdots, m$, *and* $u_{t'}$ *for all* t *and* t'.

(iii) $\mathbb{E}(\mathrm{a}_{\ell t}) = \mathrm{a}_{\ell}$ *for* $\ell = 1, 2, \cdots, m$ *and all* t.

Before presenting our theoretical results, we briefly mention pros and cons of our assumptions and compare them with the assumptions typically made in the high-dimensional linear regression and the structural breaks literature.

Assumptions 1 allows the variables $z_{\ell t}$, $\mathrm{a}_{\ell t}$, x_{it}, β_{it} and u_t to follow martingale difference processes, which is weaker than the IID assumption typically made in the literature. Following a similar line of argument as in Section 4.2 of Chudik et al. (2018), we can relax this assumption to allow for weak serial correlation.

Assumption 2 imposes the variables $z_{\ell t}$, $\mathrm{a}_{\ell t}$, x_{it}, β_{it} and u_t to have exponentially decaying probability tails to ensure all moments exist. This assumption is stronger than those needed in the studies on structural breaks, but it is required to drive upper and lower probability bounds for the selection of the average noise variables and the signals, respectively. It is common in the high-dimensional linear literature to assume some form of exponentially decaying probability bound for the variables. For example, see Zheng et al. (2014), Fan et al. (2020) and Chudik et al. (2018).

Assumptions 3(i) and 4(i) are required to establish that the target variable, y_t, has the exponentially decaying probability tail of the same order as the other random variables. Assumptions 3(ii) and 4(ii) ensure the distribution of time-varying parameters $\mathrm{a}_{\ell t}$ and β_{it} to be independent of the observed covariates (x_{it} and $z_{\ell t}$) and u_t, which is a standard assumption in the literature on structural breaks. Assumptions 3(iii) ensures the average value of the coefficients of the signal variables does not approach zero too fast. It is an identification assumption that distinguishes signal from noise variables. Finally, 4(iii) constrains the expected values of coefficients of conditioning variables to be time-invariant.

4.2 Theoretical Findings

As mentioned in Section 1, the purpose of this paper is to provide the theoretical argument for applying the OCMT procedure with no down-weighting for variable selection in linear high-dimensional settings subject to structural breaks. We now show that under certain conditions discussed in Section 4.1, the OCMT procedure selects the approximating model that contains all the signals; $\{x_{it} : i = 1, 2, \cdots, k\}$; and none of the average noise variables;

$\{x_{it} : k+k^*+1, k+k^*+2, \cdots, N\}$. The event of choosing the approximating model is define by

$$\mathcal{A}_0 = \left\{\sum_{i=1}^{k} \hat{\mathcal{J}}_i = k\right\} \cap \left\{\sum_{i=k+k^*+1}^{N} \hat{\mathcal{J}}_i = 0\right\}. \tag{10}$$

In what follows, we show that $\Pr(\mathcal{A}_0) \to 1$ as $N, T \to \infty$.

Theorem 1 *Let y_t for $t = 1, 2, \cdots, T$ be generated by (1), and let $T = \ominus(N^{\kappa_1})$ with $\kappa_1 > 0$, and $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$ which contains k signals, k^* pseudo-signals, and $N - k - k^*$ average noise variables. Consider the single-stage OCMT procedure with the critical value function $c_p(N, \delta)$ given by (8), for some $\delta > 0$. Then under Assumptions 1-4, there exist finite positive constants C_0, and C_1 such that, the probability of selecting the approximating model, $\mathcal{A}_0$, defined by (10), is given by*

$$\Pr(\mathcal{A}_0) = 1 - O(N^{1-2C_0\delta}) - O[\exp(-N^{C_1\kappa_1})]. \tag{11}$$

See Appendix A.1 for a proof.

The above theorem can be readily extended to a multi-stage OCMT where the selection procedure is repeated after augmenting the conditioning variables with the selected variables from the earlier stages. The results turn out to be similar to the ones reported in part (b) of Theorem 1 in Chudik et al. (2018). We have focussed on the first stage of OCMT to simplify the proofs. It is interesting that the asymptotic results regarding the probability of selecting the approximating model are unaffected by parameter instability, so long as the average net effects of the true covariates are non-zero or tend to zero sufficiently slowly in T, as defined formally by Assumption 3. But, we do not consider the problems of inference on selected variables (post OCMT inference) which go beyond the forecasting problem, and have been covered in Theorem 2 of Chudik et al. (2018) under parameter stability.

Remark 1 *Based on the results from Theorem 1, to ensure that the probability of selecting the approximating model tends to one as $T, N \to \infty$, we need $\delta > \frac{1}{2C_0}$. The extensive Monte Carlo Studies in Chudik et al. (2018) suggest that setting $\delta = 1$ preforms well in practice.*

5 Empirical Applications

The rest of the paper considers a number of empirical applications whereby the forecast performance of the proposed OCMT approach with no down-weighting at the selection stage is compared with those of Lasso and Adaptive Lasso. In particular, we consider the following three applications:

- Forecasting Euro Area quarterly output growth using the European Central Bank (ECB) survey of professional forecasters.
- Forecasting monthly rate of price changes, $r_{lt} = 100 \ln(P_{lt}/P_{l,t-1})$, for $l = 1, 2, ..., 30$ of stocks in Dow Jones over the period 01/1980 and 12/2017.
- Forecasting quarterly output growth rates across 33 countries using macro and financial variables.

In each application, we first compare the performance of OCMT with and without down-weighted observations at the selection stage. We then consider the comparative performance of OCMT (with variable selection carried out without down-weighting) relative to Lasso and Adaptive Lasso, with and without down-weighting. For down-weighting we make use of exponentially down-weighted observations, namely

$$\hat{x}_{it}(\lambda) = \lambda^{T-t} x_{it}, \qquad \text{and} \qquad \hat{y}_t(\lambda) = \lambda^{T-t} y_t,$$

where y_t is the target variable to be forecasted, x_{it}, for $i = 1, 2, ..., N$ are the covariates in the active set, and λ is the exponential decay coefficient. We consider two sets of values for the degree of exponential decay, λ:

1. Light down-weighting with $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$,
2. Heavy down-weighting with $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$.

For each of the above two sets of exponential down-weighting schemes we focus on simple average forecasts computed over the individual forecasts obtained for each value of λ in the set under consideration.

For forecast evaluation we consider Mean Squared Forecasting Error (MSFE) and Mean Directional Forecast Accuracy (MDFA), together with related pooled versions of Diebold-Mariano (DM), and Pesaran-Timmermann (PT) test statistics. A panel version of Diebold and Mariano (2002) test is proposed by Pesaran et al. (2009). Let $q_{lt} \equiv e^2_{ltA} - e^2_{ltB}$ be the difference in the squared forecasting errors of procedures A and B, for the target variable y_{lt} ($l = 1, 2, ..., L$) and $t = 1, 2, ..., T^f_l$, where T^f_l is the number of forecasts for target variable l (could be one or multiple step ahead) under consideration. Suppose $q_{lt} = \alpha_l + \varepsilon_{lt}$ with $\varepsilon_{lt} \sim \mathcal{N}(0, \sigma^2_l)$. Then under the null hypothesis of $H_0 : \alpha_l = 0$ for all l we have

$$\overline{DM} = \frac{\bar{q}}{\sqrt{V(\bar{q})}} \overset{a}{\sim} \mathcal{N}(0, 1), \text{ for } T_{Lf} \to \infty,$$

where $T_{Lf} = \sum_{l=1}^{L} T_l^f$, $\bar{q} = T_{Lf}^{-1} \sum_{l=1}^{L} \sum_{t=1}^{T_l^f} q_{lt}$ and

$$V(\bar{q}) = \frac{1}{T_{Lf}^2} \sum_{l=1}^{L} T_l^f \hat{\sigma}_l^2, \text{ with } \hat{\sigma}_l^2 = \frac{1}{T_l^f} \sum_{t=1}^{T_l^f} (q_{lt} - \bar{q}_l)^2 \text{ and } \bar{q}_l = \frac{1}{T_l^f} \sum_{t=1}^{T_l^f} q_{lt}.$$

Note that $V(\bar{q})$ needs to be modified in the case of multiple-step ahead forecast errors, due to the serial correlation that results in the forecast errors from the use of over-lapping observations. There is no adjustment needed for one-step ahead forecasting, since it is reasonable to assume that in this case the loss differentials are serially uncorrelated. However, to handle possible serial correlation for h-step ahead forecasting with $h > 1$, we can modify the panel DM test by using the Newey-West type estimator of σ_l^2.

The $MDFA$ statistic compares the accuracy of forecasts in predicting the direction (sign) of the target variable, and is computed as

$$MDFA = 100 \left\{ \frac{1}{T_{Lf}} \sum_{l=l}^{L} \sum_{t=1}^{T_l^f} \mathbf{1}[\text{sgn}(y_{lt} y_{lt}^f) > 0] \right\},$$

where $\mathbf{1}(w > 0)$ is the indicator function which takes the value of 1 when $w > 0$ and zero otherwise, $\text{sgn}(w)$ is the sign function, y_{lt} is the actual value of dependent variable at time t and y_{lt}^f is its corresponding predicted value. To evaluate statistical significant of the directional forecasts for each method, we also report a pooled version of the test suggested by Pesaran and Timmermann (1992):

$$PT = \frac{\hat{P} - \hat{P}^*}{\sqrt{\hat{V}(\hat{P}) - \hat{V}(\hat{P}^*)}}$$

where $\hat{P}$ is the estimator of the probability of correctly predicting the sign of y_{lt}, computed by

$$\hat{P} = \frac{1}{T_{Lf}} \sum_{l=1}^{L} \sum_{t=1}^{T_l^f} \mathbf{1}[\text{sgn}(y_{lt} y_{lt}^f) > 0], \text{ and } \hat{P}^* = \bar{d}_y \bar{d}_{y^f} + (1 - \bar{d}_y)(1 - \bar{d}_{y^f}),$$

where

$$\bar{d}_y = \frac{1}{T_{Lf}} \sum_{l=1}^{L} \sum_{t=1}^{T_l^f} \mathbf{1}[\text{sgn}(y_{lt}) > 0], \text{ and } \bar{d}_{y^f} = \frac{1}{T_{Lf}} \sum_{l=1}^{L} \sum_{t=1}^{T_l^f} \mathbf{1}[\text{sgn}(y_{lt}^f) > 0].$$

Finally, $\hat{V}(\hat{P}) = T_{Lf}^{-1}\hat{P}^*(1-\hat{P}*)$, and

$$\hat{V}(\hat{P}^*) = \frac{1}{T_{Lf}}(2\bar{d}_y - 1)^2\bar{d}_{y^f}(1-\bar{d}_{y^f}) + \frac{1}{T_{Lf}}(2\bar{d}_y^f - 1)^2\bar{d}_y(1-\bar{d}_y) + \frac{4}{T_{Lf}^2}\bar{d}_y\bar{d}_{y^f}(1-\bar{d}_y)(1-\bar{d}_{y^f}).$$

The last term of $\hat{V}(\hat{P}^*)$ is negligible and can be ignored. Under the null hypothesis, that prediction and realization are independently distributed, PT is asymptotically distributed as a standard normal distribution.

5.1 Forecasting Euro Area output growth with the ECB survey of professional forecasters

This application considers forecasting one-year ahead Euro Area real output growth using the ECB survey of professional forecasters, recently analyzed by Diebold and Shin (2019). The dataset consists of quarterly predictions of 25 professional forecasters over the period 1999Q3 to 2014Q1.[2] The predictions of these forecasters are highly correlated suggesting the presence of a common factor across these forecasts. To deal with this issue at the variable selection stage following Sharifvaghefi (2020) we also include the simple average of the 25 forecasts in the conditioning set, $\mathbf{z}_t$, as a proxy for the common factor in addition to the intercept. We consider 39 quarterly forecasts (from 2004Q3 and 2014Q1) for forecast evaluation, using expanding samples (weighted and unweighted) from 1999Q3. We also consider two simple baseline forecasts: a simple cross sectional (CS) average of the professional forecasts, and forecasts computed using a regression of output growths on an intercept and the CS average of the professional forecasts.

Table 1 compares the forecast performance of OCMT with and without down-weighting at the selection and forecasting stages, in terms of MSFE. The results suggest that down-weighting at the selection stage leaves us with larger forecasting errors. The MSFE goes from 3.765 (3.995) to 3.874 (4.672) in case of light (heavy) down-weighting. However, the panel DM tests indicate that the MSFE among different scenarios are not statistically significant, possibly due to the short samples being considered. In Table 2, we compare OCMT (with no down-weighting at the selection stage) with Lasso and Adaptive Lasso. The results indicate that the OCMT procedure outperforms Lasso and Adaptive Lasso in terms of MSFE when using no down-weighting, light down-weighting, and heavy down-weighting at the forecasting stage. It is worth mentioning that OCMT selects 3 forecasters (Forecaster #21 for 2004Q4-2005Q1, Forecaster #7 for 2007Q2-2008Q3, and Forecaster #18 for 2011Q2-2011Q3). This means that over the full evaluating sample, only 0.3 variables are selected by OCMT from the active set on average. In contrast, Lasso selects 12.6 forecasters on average. Each individual

[2] We are grateful to Frank Diebold for providing us with the data set.

Table 1: Mean square forecast error (MSFE) and panel DM test of OCMT of one-year ahead Euro Area annual real output growth forecasts between 2004Q3 and 2014Q1 (39 forecasts)

	Down-weighting at†				
	Selection stage	Forecasting stage		MSFE	
(M1)	no	no		3.507	
Light down-weighting, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$					
(M2)	no	yes		3.765	
(M3)	yes	yes		3.874	
Heavy down-weighting, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$					
(M4)	no	yes		3.995	
(M5)	yes	yes		4.672	
	Pair-wise panel DM tests				
	Light down-weighting			Heavy down-weighting	
	(M2)	(M3)		(M4)	(M5)
(M1)	-0.737	-0.474	(M1)	-0.656	-0.741
(M2)	-	-0.187	(M5)	-	-0.645

Notes: The active set consists of 25 individual forecasts. The conditioning set consists of an intercept and the cross sectional average of 25 forecasts.

†For each of the two sets of exponential down-weighting (light/heavy) forecasts of the target variable are computed as the simple average of the forecasts obtained using the down-weighting coefficient, λ, in the "light" or the "heavy" down-weighting set under consideration.

Table 2: Mean square forecast error (MSFE) and panel DM test of OCMT versus Lasso and Adaptive Lasso of one-year ahead Euro Area annual real output growth forecasts between 2004Q3 and 2014Q1 (39 forecasts)

	MSFE under different down-weighting scenarios					
	No down-weighting		Light down-weighting†		Heavy down-weighting‡	
OCMT	3.507		3.765		3.995	
Lasso	5.242		5.116		5.385	
A-Lasso	7.559		6.475		6.539	
	Selected pair-wise panel DM tests					
	No down-weighting		Light down-weighting		Heavy down-weighting	
	Lasso	A-Lasso	Lasso	A-Lasso	Lasso	A-Lasso
OCMT	-1.413	-1.544	-0.990	-1.265	-1.070	-1.267
Lasso	-	-1.484	-	-1.589	-	-1.527

Notes: The active set consists of forecasts by 25 individual forecasters. The conditioning set contains an intercept and the cross sectional average of the 25 forecasts.

† Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.

‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

forecaster is selected for at least part of the evaluation period. As to be expected, Adaptive Lasso selects a fewer number of forecasters (9.8 on average) as compared to Lasso (12.6 on average), and performs slightly worse.

To summarize, we find that down-weighting at the selection stage of OCMT leads to forecast deterioration (in terms of MSFE). OCMT outperforms Lasso and Adaptive Lasso, but the panel DM tests are not statistically significant. Moreover, none of the considered big data methods can beat the simple baseline models.

5.2 Forecasting monthly rates of price changes of stocks in Dow Jones

In this application the focus is on forecasting one-month ahead stock returns, defined as monthly change in natural logarithm of stock prices. We consider stocks that were part of the Dow Jones index in 2017m12, and have non-zero prices for at least 120 consecutive data points (10 years) over the period 1980m1 and 2017m12. We ended up forecasting 28 blue chip stocks.[3] Daily close prices for all the stocks are obtained from Data Stream. For stock i, the price at the last trading day of each month is used to construct the corresponding monthly stock prices, P_{it}. Finally, monthly returns are computed by $r_{i,t+1} = 100 \ln(P_{i,t+1}/P_{it})$, for $i = 1, 2, ..., 28$. For all 28 stocks we use an expanding window starting with the observations for the first 10 years. The active set for predicting $r_{i,t+1}$ consists of 40 financial, economic, and technical variables.[4] Financial covariates include lagged monthly returns of stock i^{th} as well as the average lagged stock return of the group to which stock i belongs. We consider 19 groups according to the Industry Classification Benchmark.[5] Each group includes all the S&P 500 stocks with the same group classification. The list of the covariates considered include:[6]

- r_{it}: The monthly stock return for the target stock, i.
- $\bar{r}_{it}^{g}$: Average return of stocks in group g to which the ith stock belongs.
- RV_{it}: Realized return volatility of stock i.
- RV_{it}^{g}: Realized volatility of group g's return to which the ith stock belongs.

[3] Visa and DwoDuPont are excluded since they have less than 10 years of historical price data.

[4] All regressions include the intercept as the only conditioning (pre-selected) variable.

[5] The 19 groups are as follows: Oil & Gas, Chemicals, Basic Resources, Construction & Materials, Industrial Goods & Services, Automobiles & Parts, Food & Beverage, Personal & Household Goods, Health Care, Retail, Media, Travel & Leisure, Telecommunications, Utilities, Banks, Insurance, Real Estate, Financial Services, and Technology.

[6] All stock returns are computed excluding dividend payments.

- $r_{sp,t}$: Return on S&P 500.
- $RV_{sp,t}$: Realized volatility of S&P 500.
- Small Minus Big (SMB), and High Minus Low (HML) factors available at Fama and French database.[7]
- Δpo_t: The change in natural logarithm of the Brent Crude oil price index.
- First order difference in Interest Rate Spreads: First order difference of the gap between long, medium and short term interest rates. We consider annualized percentage yield on 10-year, 2-year and 3 month government bonds for long, medium and short term interest rates, respectively.[8]
- We consider 28 technical indicators, constructed based on past prices using familiar algorithms (see e.g. Colby and Meyers, 1988). They include price and volatility gaps, price oscillator and relative strength indicators. The full list and the description of technical indicators considered can be found in Appendix B.

Overall we computed 8,659 monthly forecasts for the 28 target stocks. The results are summarized as average forecast performances across the different variable selection procedures. Table 3 reports the effects of down-weighting at the selection stage of the OCMT procedure. It is clear that down-weighting worsens the predictive accuracy of OCMT. From the Panel DM tests, we can also see that down-weighting at the selection stage worsens the forecasts significantly. Panel DM test statistics is -5.606 (-11.352) for light (heavy) versus no down-weighing at the selection stage. Moreover, Table 4 shows that the OCMT procedure with no down-weighting at the selection stage dominates Lasso and Adaptive Lasso in terms of MSFE and the differences are statistically highly significant.

Further, OCMT outperforms Lasso and Adaptive Lasso in terms of Mean Directional Forecast Accuracy (MDFA), measured as the percent number of correctly signed one-month ahead forecasts across all the 28 stocks over the period 1990m2-2017m12. See Table 5. As can be seen from this table, OCMT with no down-weighting performs the best; correctly predicting the direction of 56.057% of 8,659 forecasts, as compared to 55.33%, which we obtain for Lasso and Adaptive Lasso forecast, at best. This difference is highly significant considering the very large number of forecasts involved. It is also of interest that the better of performance of OCMT is achieved with a much fewer number of selected covariates as compared to Lasso and Adaptive Lasso. As can be seen from the last column of Table 5,

[7] Fama and French database can be found at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[8] The data fro these variables were harvested from the Federal Reserve Bank of St. Louis database, available at `https://fred.stlouisfed.org`

Table 3: Mean square forecast error (MSFE) and panel DM test of OCMT of one-month ahead monthly return forecasts across the 28 stocks in Dow Jones index between 1990m2 and 2017m12 (8659 forecasts)

	Down-weighting at†				
	Selection stage	Forecasting stage			MSFE
(M1)	no	no			61.231
Light Down-weighting, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$					
(M2)	no	yes			61.641
(M3)	yes	yes			68.131
Heavy Down-weighting, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$					
(M4)	no	yes			62.163
(M5)	yes	yes			86.073
	Pair-wise panel DM tests				
	Light down-weighting			Heavy down-weighting	
	(M2)	(M3)		(M4)	(M5)
(M1)	-1.528	-5.643	(M1)	-2.459	-11.381
(M2)	-	-5.606	(M4)	-	-11.352

Notes: The active set consists of 40 covariates. The conditioning set only contains an intercept.
†For each of the two sets of exponential down-weighting (light/heavy) forecasts of the target variable are computed as the simple average of the forecasts obtained using the down-weighting coefficient, λ, in the "light" or the "heavy" down-weighting set under consideration. See footnote to Table 1.

Lasso and Adaptive Lasso on average select many more covariates than OCMT (1-3 variables as compared to 0.072 for OCMT).

So far we have focussed on average performance across all the 28 stocks. Table 6 provides the summary results for individual stocks, showing the relative performance of OCMT in terms of the number of stocks, using MSFE and MDFA criteria. The results show that OCMT performs better than Lasso and Adaptive Lasso in the majority of the stocks in terms of MSFE and MDFA. OCMT outperforms Lasso in 23 out of 28 stocks in terms of MSFE, under no down-weighting, and almost universally when Lasso or Adaptive Lasso are implemented with down-weighting. Similar results are obtained when we consider MDFA criteria, although the differences in performance are somewhat less pronounced. Overall, we can conclude that the better average performance of OCMT (documented in Tables 4 and 5) is not driven by a few stocks and holds more generally.

5.3 Forecasting quarterly output growth rates across 33 countries using macro and financial variables

In this application we consider one and two years ahead predictions of output growth for 33 countries (20 advanced and 13 emerging). We use quarterly data from 1979Q2 to 2016Q4

Table 4: Mean square forecast error (MSFE) and panel DM test of OCMT versus Lasso and Adaptive Lasso of one-month ahead monthly return forecasts across the 28 stocks in Dow Jones index between 1990m2 and 2017m12 (8659 forecasts)

	MSFE under different down-weighting scenarios					
	No down-weighting		Light down-weighting†		Heavy down-weighting‡	
OCMT	61.231		61.641		62.163	
Lasso	61.849		63.201		69.145	
A-Lasso	63.069		65.017		72.038	
	Selected pair-wise panel DM tests					
	No down-weighting		Light down-weighting		Heavy down-weighting	
	Lasso	A-Lasso	Lasso	A-Lasso	Lasso	A-Lasso
OCMT	-1.533	-4.934	-2.956	-6.025	-7.676	-10.261
Lasso	-	-4.661	-	-6.885	-	-9.569

Notes: The active set consists of 40 covariates. The conditioning set contains only the intercept.
† Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

Table 5: Mean directional forecast accuracy (MDFA) and the average number of selected variables ($\hat{k}$) of OCMT, Lasso and Adaptive Lasso of one-month ahead monthly return forecasts across the 28 stocks in Dow Jones index between 1990m2 and 2017m12 (8659 forecasts).

	Down-weighting	MDFA	$\hat{k}$
OCMT	No	56.057	0.072
	Light†	55.330	0.072
	Heavy‡	54.302	0.072
Lasso	No	55.364	1.659
	Light	54.221	2.133
	Heavy	53.205	3.794
Adaptive Lasso	No	54.648	1.312
	Light	53.840	1.623
	Heavy	52.951	2.855

Notes: The active set consists of 40 variables. The conditioning set contains an intercept.
† Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

Table 6: The number of stocks out of the 28 stocks in Dow Jones index where OCMT outperforms/underperforms Lasso, and Adaptive Lasso in terms of mean square forecast error (MSFE), panel DM test and mean directional forecast accuracy (MDFA) between 1990m2 and 2017m12 (8659 forecasts).

		MSFE			
	Down-weighting	OCMT outperforms	OCMT significantly outperforms	OCMT underperforms	OCMT significantly underperforms
Lasso	No	23	4	5	2
	Light†	25	5	3	0
	Heavy‡	26	14	2	0
A-Lasso	No	24	9	4	2
	Light	27	10	1	0
	Heavy	28	24	0	0

		MDFA	
	Down-weighting	OCMT outperforms	OCMT underperforms
Lasso	No	14	6
	Light	24	4
	Heavy	17	10
A-Lasso	No	18	4
	Light	21	3
	Heavy	19	7

Notes: The active set consists of 40 variables. The conditioning set only contains an intercept.
† Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

Table 7: List of variables in the conditioning and active sets for forecasting quarterly output growths across 33 countries

Conditioning set	
c, $\Delta_1 y_{it}$	
Active Set	
(a) Domestic variables, $\ell = 0, 1$.	(b) Foreign counterparts, $\ell = 0, 1$.
$\Delta_1 y_{i,t-1}$	$\Delta_1 y^*_{i,t-\ell}$
$\Delta_1 r_{i,t-\ell} - \Delta_1 \pi_{i,t-\ell}$	$\Delta_1 r^*_{i,t-\ell} - \Delta_1 \pi^*_{i,t-\ell}$
$\Delta_1 r^L_{i,t-\ell} - \Delta_1 r_{i,t-\ell}$	$\Delta_1 r^{L*}_{i,t-\ell} - \Delta_1 r^*_{i,t-\ell}$
$\Delta_1 q_{i,t-\ell} - \Delta_1 \pi_{i,t-\ell}$	$\Delta_1 q^*_{i,t-\ell} - \Delta_1 \pi^*_{i,t-\ell}$
Total number of variables in the active set $\mathbf{x}_t$: $n = 15$ (max)	

taken from the GVAR dataset.[9] We predict $\Delta_4 y_{it} = y_{it} - y_{i,t-4}$ and $\Delta_8 y_{it} = (y_{it} - y_{i,t-8})$, where y_{it}, is the log of real output for country i in terms of lagged values of a number of commonly used variables in the literature. We adopt the following direct forecasting equations:

$$\Delta_h y_{i,t+h} = y_{i,t+h} - y_{it} = \alpha_{ih} + \lambda_{ih} \Delta_1 y_{it} + \boldsymbol{\beta}'_{ih} \mathbf{x}_{it} + u_{iht},$$

where we consider $h = 4$ (one-year-ahead forecasts) and $h = 8$ (two-years-ahead forecasts). Given the known persistence in output growth, in addition to the intercept in the present application we also condition on the most recent lagged output growth, denoted by $\Delta_1 y_{it} = y_{it} - y_{i,t-1}$, and confine the variable selection to list of variables set out at the bottom panel of Table 7. Overall, we consider a maximum of 15 covariates in the active set covering quarterly changes in domestic variables such as real output growth, real short term interest rate, and long-short interest rate spread and quarterly change in the corresponding foreign variables.

We use expanding samples, starting with the observations on the first 15 years (60 data points), and evaluate the forecasting performance of the three methods over the period 1997Q2 to 2016Q4.

Tables 8 and 9, respectively, report the MSFE of OCMT for one-year and two-year ahead forecasts of output growth, with and without down-weighting at the selection stage. Consistent with the previous two applications, down-weighting at the selection stage worsens the forecasting accuracy. Moreover, in Tables 10 and 11, we can see that OCMT (without down-weighting at the selection stage) outperforms Lasso and Adaptive Lasso in two-year ahead forecasting. In the case of one-year ahead forecasts, OCMT and Lasso are very close to each other and both outperform Adaptive Lasso. Table 12 summarizes country-specific

[9] The GVAR dataset is available at `https://sites.google.com/site/gvarmodelling/data`.

Table 8: Mean square forecast error (MSFE) and panel DM test of OCMT of one-year ahead output growth forecasts across 33 countries over the period 1997Q2-2016Q4 (2607 forecasts)

	Down-weighting at†		MSFE ($\times 10^4$)		
	Selection stage	Forecasting stage	All	Advanced	Emerging
(M1)	no	no	11.246	7.277	17.354
Light down-weighting, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$					
(M2)	no	yes	10.836	6.913	16.871
(M3)	yes	yes	10.919	6.787	17.275
Heavy down-weighting, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$					
(M4)	no	yes	11.064	7.187	17.028
(M5)	yes	yes	11.314	6.906	18.094

Pair-wise panel DM tests (all countries)

	Light down-weighting			Heavy down-weighting	
	(M2)	(M3)		(M4)	(M5)
(M1)	2.394	1.662	(M1)	0.668	-0.204
(M2)	-	-0.780	(M4)	-	-1.320

Notes: There are up to 15 macro and financial variables in the active set.
†For each of the two sets of exponential down-weighting (light/heavy) forecats of the target variable are computed as the simple average of the forecasts obtained using the down-weighting coefficient, λ, in the "light" or the "heavy" down-weighting set under consideration.

MSFE and DM findings for OCMT relative to Lasso and Adaptive Lasso. The results show OCMT under-performs Lasso in more than half of the countries for one-year ahead horizon, but outperforms Lasso and Adaptive Lasso in more than 70 percent of the countries in the case of two-year ahead forecasts. It is worth noting that while Lasso generally outperforms OCMT in the case of one-year ahead forecasts, overall its performance is not significantly better than OCMT. See Panel DM test of Table 10. On the other hand we can see from Table 11 that overall OCMT significantly outperforms Lasso in the case of the two-year ahead forecasts.

Finally in Tables 13 and 14 we reports MDFA and PT test statistics for OCMT, Lasso and Adaptive Lasso. Overall, OCMT has a slightly higher MDFA and hence predicts the direction of real output growth better than Lasso and Adaptive Lasso in most cases. The PT test statistics suggest that while all the methods perform well in forecasting the direction of one-year ahead real output growth, none of the methods considered are successful at predicting the direction of two-year ahead output growth.

It is also worth noting that as with the previous applications, OCMT selects very few variables from the active set (0.1 on average for both horizons, with the maximum number

Table 9: Mean square forecast error (MSFE) and panel DM test of OCMT of two-year ahead output growth forecasts across 33 countries over the period 1997Q2-2016Q4 (2343 forecasts)

	Down-weighting at†		MSFE ($\times 10^4$)		
	Selection stage	Forecasting stage	All	Advanced	Emerging
(M1)	no	no	9.921	7.355	13.867
Light down-weighting, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$					
(M2)	no	yes	9.487	6.874	13.505
(M3)	yes	yes	9.549	6.848	13.704
Heavy down-weighting, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$					
(M4)	no	yes	9.734	7.027	13.898
(M5)	yes	yes	10.389	7.277	15.177

Pair-wise panel DM test (all countries)					
	Light down-weighting			Heavy down-weighting	
	(M2)	(M3)		(M4)	(M5)
(M1)	3.667	2.827	(M1)	0.943	-1.664
(M2)	-	-1.009	(M4)	-	-3.498

Notes: There are up to 15 macro and financial variables in the active set.
†For each of the two sets of exponential down-weighting (light/heavy) forecasts of the target variable are computed as the simple average of the forecasts obtained using the down-weighting coefficient, λ, in the "light" or the "heavy" down-weighting set under consideration..

Table 10: Mean square forecast error (MSFE) and panel DM test of OCMT versus Lasso, and Adaptive Lasso for one-year ahead output growth forecasts across 33 countries over the period1997Q2-2016Q4 (2607 forecasts)

	MSFE under different down-weighting scenarios								
	No down-weighting			Light down-weighting†			Heavy down-weighting‡		
	All	Adv.*	Emer.**	All	Adv.	Emer.	All	Adv.	Emer.
OCMT	11.246	7.277	17.354	10.836	6.913	16.871	11.064	7.187	17.028
Lasso	11.205	6.975	17.714	10.729	6.427	17.347	11.749	7.186	18.769
A-Lasso	11.579	7.128	18.426	11.153	6.548	18.236	12.254	7.482	19.595

Pair-wise panel DM tests (All countries)						
	No down-weighting		Light down-weighting		Heavy down-weighting	
	Lasso	A-Lasso	Lasso	A-Lasso	Lasso	A-Lasso
OCMT	0.220	-1.079	0.486	-1.007	-1.799	-2.441
Lasso	-	-2.625	-	-3.626	-	-3.157

Notes: There are up to 15 macro and financial covariates in the active set.
† Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.
* Adv. stands for advanced economies.
** Emer. stands for emerging economies.

Table 11: Mean square forecast error (MSFE) and panel DM test of OCMT versus Lasso, and Adaptive Lasso of two-year ahead output growth forecasts across 33 countries over the period1997Q2-2016Q4 (2343 forecasts)

	MSFE under different down-weighting scenarios								
	No down-weighting			Light down-weighting†			Heavy down-weighting‡		
	All	Adv.*	Emer.**	All	Adv.	Emer.	All	Adv.	Emer.
OCMT	9.921	7.355	13.867	9.487	6.874	13.505	9.734	7.027	13.898
Lasso	10.151	7.583	14.103	9.662	7.099	13.605	10.202	7.428	14.469
A-Lasso	10.580	7.899	14.705	10.090	7.493	14.087	11.008	8.195	15.336

	Pair-wise panel DM tests (All countries)					
	No down-weighting		Light down-weighting		Heavy down-weighting	
	Lasso	A-Lasso	Lasso	A-Lasso	Lasso	A-Lasso
OCMT	-2.684	-4.200	-2.137	-4.015	-3.606	-4.789
Lasso	-	-5.000	-	-4.950	-	-4.969

Notes: There are up to 15 macro and financial covariates in the active set.
† Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.
* Adv. stands for advanced economies.
** Emer. stands for emerging economies.

Table 12: The number of countries out of the 33 countries where OCMT outperforms/underperforms Lasso, and Adaptive Lasso in terms of mean square forecast error (MSFE) and panel DM test over the period 1997Q2 -2016Q4

	Down-weighting	OCMT outperforms	OCMT significantly outperforms	OCMT underperforms	OCMT significantly underperforms
		One-year-ahead horizon ($h = 4$ quarters)			
Lasso	No	13	0	20	3
	Light†	12	1	21	3
	Heavy‡	17	1	16	3
Adaptive Lasso	No	16	1	17	2
	Light	14	2	19	2
	Heavy	19	1	14	0
		Two-years-ahead horizon ($h = 8$ quarters)			
Lasso	No	24	1	9	0
	Light	25	1	8	1
	Heavy	25	1	8	0
Adaptive Lasso	No	25	2	8	0
	Light	28	3	5	1
	Heavy	30	3	3	0

Notes: There are up to 15 macro and financial covariates in the active set.
†Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

of selected variables being 2 for $h = 4$ and 8). On the other hand, Lasso on average selects 2.7 variables from the active set for $h = 4$, and 1 variable on average for $h = 8$. Maximum number of variables selected by Lasso is 9 and 13 for $h = 4$, 8, respectively (out of possible 15). Again as to be expected, Adaptive Lasso selects a fewer number of variables as compared to Lasso (2.3 and 0.8 on average for $h = 4, 8$, respectively), but this does not lead to a better forecast performance in comparison with Lasso.

In conclusion, down-weighting at both selection and forecasting stages deteriorates OCMT's MSFE for both one–year and two-years ahead forecast horizons, as compared to down-weighting only at the forecasting stage. Moreover, light down-weighting at the forecasting stage improves forecasting performance for both horizons. Statistically significant evidence of forecasting skill is found for OCMT relative to Lasso only in the case of two-years ahead forecasts. However, it is interesting that none of the big data methods can significantly beat the simple (light down-weighted) AR(1) baseline model.

Table 13: Mean directional forecast accuracy (MDFA) and PT test of OCMT, Lasso and Adaptive Lasso for one-year ahead output growth forecasts over the period 1997Q2-2016Q4 (2607 forecasts)

	Down-	MDFA			PT tests		
	weighting	All	Advanced	Emerging	All	Advanced	Emerging
OCMT	No	87.6	87.4	88.0	8.12	7.40	3.48
	Light†	87.4	87.1	87.8	7.36	6.95	2.53
	Heavy‡	86.8	86.3	87.5	6.25	5.93	1.95
Lasso	No	87.0	86.9	87.2	9.64	9.15	3.80
	Light	87.1	87.1	87.1	8.12	8.22	2.26
	Heavy	86.0	85.8	86.4	6.24	6.43	1.40
Adaptive Lasso	No	87.3	87.3	87.2	10.80	9.91	4.75
	Light	86.5	86.6	86.4	8.25	8.36	2.48
	Heavy	85.5	85.3	85.7	6.84	6.92	1.88

Notes: There are up to 15 macro and financial variables in the active set.
† Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$..

Table 14: Mean directional forecast accuracy (MDFA) and PT test of OCMT, Lasso and Adaptive Lasso for two-year ahead output growth forecasts over the period 1997Q2-2016Q4 (2343 forecasts)

	Down-	MDFA			PT tests		
	weighting	All	Advanced	Emerging	All	Advanced	Emerging
OCMT	No	88.0	86.7	89.9	0.52	0.00	0.47
	Light†	87.7	86.6	89.3	1.11	0.39	0.94
	Heavy‡	87.0	85.8	88.8	0.50	0.89	0.34
Lasso	No	87.6	86.6	89.2	0.77	0.60	0.66
	Light	87.5	86.3	89.4	0.07	0.79	0.88
	Heavy	86.8	85.5	88.8	1.54	1.87	0.34
Adaptive Lasso	No	87.0	85.6	89.2	0.33	0.13	1.00
	Light	87.1	85.9	88.9	1.03	1.82	1.10
	Heavy	86.2	84.8	88.4	1.53	1.92	0.62

Notes: There are up to 15 macro and financial variables in the active set.
†Light down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$.
‡ Heavy down-weighted forecasts are computed as simple averages of forecasts obtained using the down-weighting coefficient, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

6 Conclusion

The penalized regression approach has become the de facto benchmark in the literature in the context of linear regression models without breaks. These studies (with few exceptions, including Kapetanios and Zikes, 2018) do not consider the problem of variable selection when breaks are present. Recently, Chudik et al. (2018) proposed OCMT as an alternative procedure to penalized regression. One clear advantage of the OCMT procedure is the fact that the problem of variable selection is separated from the forecasting stage, which is in contrast to the penalized regression techniques where the variable selection and estimation are carried out simultaneously. Using OCMT one can decide whether to use the weighted observations at the variable selection stage or not, without preempting a different down-weighting procedure at the forecasting stage.

We have provided theoretical arguments for using the full (not down-weighted) sample at the selection stage of OCMT, and down-weighted observations (if needed) at the forecasting stage of OCMT. The benefits of the proposed method are illustrated by a number of empirical applications to forecasting output growth and stock market returns. Our results consistently suggests that using down-weighted observations at the selection stage of OCMT deteriorate the forecasting accuracy in terms of mean square forecast error and mean directional forecast accuracy. Moreover, our results suggest that overall OCMT with no down-weighting at the selection stage outperforms penalized regression methods, i.e. Lasso and Adaptive Lasso, which are prone to over-fitting.

A Appendix A: Mathematical Derivations

The proof of Theorem 1 is provided in Section A.1. This proof is based on a number of Lemmas presented in Section A.2. Section A.2 presents and establishes the main lemmas mostly needed to proof the key lemma of the paper, namely Lemma A.6. For each covariate $i = 1, 2, \cdots, N$, this lemma establishes exponential probability inequalities for the t-ratio multiple tests conditional on the average net effect, $\bar{\theta}_{i,T}$, being either of the order $\ominus(T^{-\varepsilon_i})$ for some $\varepsilon_i > 1/2$, or of the order $\ominus(T^{-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$. Section B.1 of the online supplement provides the complementary lemmas needed for the proofs of the main lemma in Section A.2.

Additional Notations and Definitions: Throughout this Appendix we consider the following events:

$$\mathcal{A}_0 = \mathcal{H} \cap \mathcal{G}, \text{ where } \mathcal{H} = \left\{ \sum_{i=1}^{k} \hat{\mathcal{J}}_i = k \right\} \text{ and } \mathcal{G} = \left\{ \sum_{i=k+k^*+1}^{N} \hat{\mathcal{J}}_i = 0 \right\}, \tag{A.1}$$

where $\{\hat{\mathcal{J}}_i \text{ for } i = 1, 2, \cdots, N\}$ are the selection indicators defined by (9). $\mathcal{A}_0$ is the event of selecting the approximating model, defined by $\mathcal{H}$, is the event that all signals are selected, and $\mathcal{G}$ is the event that no (average) noise variable is selected. To simplify the exposition, with slight abuse of notation, we denote the probability of an event $\mathcal{E}$ conditional on $\bar{\theta}_{i,T}$ being of order $\ominus(T^{-a})$ by $\Pr[\mathcal{E}|\bar{\theta}_{i,T} = \ominus(T^{-a})]$,where a is a nonnegative constant.

A.1 Proof of Theorem 1

To establish result (11), first note that $\mathcal{A}_0^c = \mathcal{H}^c \cup \mathcal{G}^c$ and hence ($\mathcal{H}^c$ denotes the complement of $\mathcal{H}$)

$$\Pr(\mathcal{A}_0^c) = \Pr(\mathcal{H}^c) + \Pr(\mathcal{G}^c) - \Pr(\mathcal{H}^c \cap \mathcal{G}^c) \leq \Pr(\mathcal{H}^c) + \Pr(\mathcal{G}^c), \tag{A.2}$$

where $\mathcal{H}$ and $\mathcal{G}$ are given by (A.1). We also have $\mathcal{H}^c = \{\sum_{i=1}^{k} \hat{\mathcal{J}}_i < k\}$ and $\mathcal{G}^c = \{\sum_{i=k+k^*+1}^{N} \hat{\mathcal{J}}_i > 0\}$. Let's consider $\Pr(\mathcal{H}^c)$ and $\Pr(\mathcal{G}^c)$ in turn:

$$\Pr(\mathcal{H}^c) \leq \sum_{i=1}^{k} \Pr(\hat{\mathcal{J}}_i = 0).$$

But for any signal

$$\Pr(\hat{\mathcal{J}}_i = 0) = \Pr\left[|t_{i,T}| < c_p(N, \delta)|\bar{\theta}_{i,T} = \ominus(T^{-\vartheta_i})\right] = 1 - \Pr\left[|t_{i,T}| > c_p(N, \delta)|\bar{\theta}_{i,T} = \ominus(T^{-\vartheta_i})\right],$$

where $0 \leq \vartheta_i < 1/2$ and hence by Lemma A.6, we can conclude that there exist sufficiently large positive constants C_0 and C_1 such that

$$\Pr(\hat{\mathcal{J}}_i = 0) = O\left[\exp(-C_0 T^{C_1})\right].$$

Since by Assumption 3, the number of signals is finite we can further conclude that

$$\Pr(\mathcal{H}^c) = O\left[\exp(-C_0 T^{C_1})\right] \tag{A.3}$$

for some finite positive constants C_0 and C_1. In the next step note that

$$\Pr(\mathcal{G}^c) = \Pr\left(\sum_{i=k+k^*+1}^{N} \hat{\mathcal{J}}_i > 0\right) \leq \sum_{i=k+k^*+1}^{N} \Pr\left(\hat{\mathcal{J}}_i = 1\right).$$

But for any average noise variable

$$\Pr(\hat{\mathcal{J}}_i = 0) = \Pr\left[|t_{i,T}| > c_p(N,\delta) | \bar{\theta}_{i,T} = \ominus(T^{-\epsilon_i})\right],$$

where $\epsilon_i > 1/2$ and hence by Lemma A.6, we can conclude that there exist sufficiently large positive constants C_0, C_1 and C_2 such that

$$\Pr(\hat{\mathcal{J}}_i = 0) \leq \exp\left[-C_0 c_p^2(N,\delta)\right] + \exp(-C_1 T^{C_2}).$$

Therefore,

$$\Pr(\mathcal{G}^c) \leq N \exp\left[-C_0 c_p^2(N,\delta)\right] + N \exp(-C_1 T^{C_2}),$$

and by result (II) of Lemma S.2 in online supplement we can further write

$$\Pr(\mathcal{G}^c) = O(N^{1-2C_0\delta}) + O\left[N \exp(-C_1 T^{C_2})\right]. \tag{A.4}$$

Using (A.3) and (A.4) in (A.2), we obtain

$$\Pr(\mathcal{A}_0^c) = O(N^{1-2C_0\delta}) + O\left[N \exp(-C_1 T^{C_2})\right],$$

and

$$\Pr(\mathcal{A}_0) = 1 - O(N^{1-2C_0\delta}) - O\left[N \exp(-C_1 T^{C_2})\right],$$

which completes the proof.

A.2 Main Lemmas

Lemma A.1 *Let y_t be a target variable generated by equation (1), $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{mt})'$ be the $m \times 1$ vector of conditioning covariates in DGP(1) and x_{it} be a covariate in the active*

set $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. *Under Assumptions 1, 3, and 4 we have*

$$\mathbb{E}\left[y_t x_{it} - \mathbb{E}(y_t x_{it})|\mathcal{F}_{t-1}\right] = 0,$$

for $i = 1, 2, \cdots, N$,

$$\mathbb{E}\left[y_t z_{\ell t} - \mathbb{E}(y_t z_{\ell t})|\mathcal{F}_{t-1}\right] = 0,$$

for $\ell = 1, 2, \cdots, m$, *and*

$$\mathbb{E}\left[y_t^2 - \mathbb{E}(y_t^2)|\mathcal{F}_{t-1}\right] = 0.$$

Proof. Note that y_t can be written as

$$y_t = \mathbf{z}_t'\mathbf{a}_t + \mathbf{x}_{k,t}'\boldsymbol{\beta}_t + u_t = \textstyle\sum_{\ell=1}^m \mathrm{a}_{\ell t} z_{\ell t} + \sum_{j=1}^k \beta_{jt} x_{jt} + u_t,$$

where $\mathbf{x}_{k,t} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, and $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \cdots, \beta_{kt})'$. Moreover, By Assumption 4, $\mathrm{a}_{\ell t}$ is independent of $x_{it'}$ and $z_{\ell' t'}$ for all i, ℓ', and t'. Hence, for $i = 1, 2, \cdots, N$, we have

$$\mathbb{E}(y_t x_{it}|\mathcal{F}_{t-1}) = \textstyle\sum_{\ell=1}^m \mathbb{E}(\mathrm{a}_{\ell t}|\mathcal{F}_{t-1})\mathbb{E}(z_{\ell t} x_{it}|\mathcal{F}_{t-1}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt}|\mathcal{F}_{t-1})\mathbb{E}(x_{jt} x_{it}|\mathcal{F}_{t-1}) + \mathbb{E}(u_t x_{it}|\mathcal{F}_{t-1}).$$

By Assumption 1, we have $\mathbb{E}(\mathrm{a}_{\ell t}|\mathcal{F}_{t-1}) = \mathbb{E}(\mathrm{a}_{\ell t})$, $\mathbb{E}(z_{\ell t} x_{it}|\mathcal{F}_{t-1}) = \mathbb{E}(z_{\ell t} x_{it})$, $\mathbb{E}(\beta_{jt}|\mathcal{F}_{t-1}) = \mathbb{E}(\beta_{jt})$, $\mathbb{E}(x_{jt} x_{it}|\mathcal{F}_{t-1}) = \mathbb{E}(x_{jt} x_{it})$, and $\mathbb{E}(u_t x_{it}|\mathcal{F}_{t-1}) = \mathbb{E}(u_t x_{it})$. Therefore,

$$\mathbb{E}(y_t x_{it}|\mathcal{F}_{t-1}) = \textstyle\sum_{\ell=1}^m \mathbb{E}(\mathrm{a}_{\ell t})\mathbb{E}(z_{\ell t} x_{it}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt})\mathbb{E}(x_{jt} x_{it}) + \mathbb{E}(u_t x_{it}) = \mathbb{E}(y_t x_{it}).$$

Similarly, we can show that for $\ell = 1, 2, \cdots, m$,

$$\begin{aligned}\mathbb{E}(y_t z_{\ell t}|\mathcal{F}_{t-1}) &= \textstyle\sum_{\ell'=1}^m \mathbb{E}(\mathrm{a}_{\ell' t}|\mathcal{F}_{t-1})\mathbb{E}(z_{\ell' t} z_{\ell t}|\mathcal{F}_{t-1}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt}|\mathcal{F}_{t-1})\mathbb{E}(x_{jt} z_{\ell t}|\mathcal{F}_{t-1}) + \mathbb{E}(u_t z_{\ell t}|\mathcal{F}_{t-1}) \\ &= \textstyle\sum_{\ell'=1}^m \mathbb{E}(\mathrm{a}_{\ell' t})\mathbb{E}(z_{\ell' t} z_{\ell t}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt})\mathbb{E}(x_{jt} z_{\ell t}) + \mathbb{E}(u_t z_{\ell t}) = \mathbb{E}(y_t z_{\ell t}).\end{aligned}$$

Also to establish the last result, we can write y_t as $y_t = \mathbf{w}_t'\boldsymbol{\delta}_t + u_t$, where $\mathbf{w}_t = (\mathbf{z}_t', \mathbf{x}_{k,t}')'$ and $\boldsymbol{\delta}_t = (\mathbf{a}_t', \boldsymbol{\beta}_t')'$. We have,

$$\begin{aligned}\mathbb{E}(y_t^2|\mathcal{F}_{t-1}) &= \mathbb{E}(\boldsymbol{\delta}_t'|\mathcal{F}_{t-1})\mathbb{E}(\mathbf{w}_t\mathbf{w}_t'|\mathcal{F}_{t-1})\mathbb{E}(\boldsymbol{\delta}_t|\mathcal{F}_{t-1}) + \mathbb{E}(u_t^2|\mathcal{F}_{t-1}) + 2\mathbb{E}(\boldsymbol{\delta}_t'|\mathcal{F}_{t-1})\mathbb{E}(\mathbf{w}_t u_t|\mathcal{F}_{t-1}) \\ &= \mathbb{E}(\boldsymbol{\delta}_t')\mathbb{E}(\mathbf{w}_t\mathbf{w}_t')\mathbb{E}(\boldsymbol{\delta}_t) + \mathbb{E}(u_t^2) + 2\mathbb{E}(\boldsymbol{\delta}_t')\mathbb{E}(\mathbf{w}_t u_t) = \mathbb{E}(y_t^2).\end{aligned}$$

■

Lemma A.2 *Let* y_t *be a target variable generated by equation (1). Under Assumptions 2-4, for any value of* $\alpha > 0$, *there exist some positive constants* C_0 *and* C_1 *such that*

$$\sup_t \Pr(|y_t| > \alpha) \leq C_0 \exp(C_1 \alpha^{s/2})$$

Proof. Note that

$$|y_t| \leq \sum_{\ell=1}^{m} |\mathrm{a}_{\ell t} z_{\ell t}| + \sum_{j=1}^{k} |\beta_{jt} x_{jt}| + |u_t|.$$

Therefore,

$$\Pr(|y_t| > \alpha) \leq \Pr(\sum_{\ell=1}^{m} |\mathrm{a}_{\ell t} z_{\ell t}| + \sum_{j=1}^{k} |\beta_{jt} x_{jt}| + |u_t| > \alpha),$$

and by Lemma S.3 for any $0 < \pi_i < 1$, $i = 1, 2, \cdots, k+m+1$, with $\sum_{i=1}^{k+m+1} \pi_j = 1$, we can further write

$$\Pr(|y_t| > \alpha) \leq \sum_{\ell=1}^{m} \Pr(|\mathrm{a}_{\ell t} z_{\ell t}| > \pi_\ell \alpha) + \sum_{j=1}^{k} \Pr(|\beta_{jt} x_{jt}| > \pi_j \alpha) + \Pr(|u_t| > \pi_{k+m+1} \alpha).$$

Moreover, by Lemma S.4, we have

$$\Pr(|\mathrm{a}_{\ell t} z_{\ell t}| > \pi_\ell \alpha) \leq \Pr[|z_{\ell t}| > (\pi_\ell \alpha)^{1/2}] + \Pr[|\mathrm{a}_{\ell t}| > (\pi_\ell \alpha)^{1/2}],$$
$$\Pr(|\beta_{jt} x_{jt}| > \pi_j \alpha) \leq \Pr[|x_{jt}| > (\pi_j \alpha)^{1/2}] + \Pr[|\beta_{jt}| > (\pi_i \alpha)^{1/2}],$$

and hence

$$\Pr(|y_t| > \alpha) \leq \sum_{\ell=1}^{m} \Pr[|z_{\ell t}| > (\pi_\ell \alpha)^{1/2}] + \sum_{\ell=1}^{m} \Pr[|\mathrm{a}_{\ell t}| > (\pi_\ell \alpha)^{1/2}] + \sum_{j=1}^{k} \Pr[|x_{jt}| > (\pi_j \alpha)^{1/2}] + \sum_{j=1}^{k} \Pr[|\beta_{jt}| > (\pi_j \alpha)^{1/2}] + \Pr(|u_t| > \pi_{k+1} \alpha),$$

Therefore, under Assumptions 2-4, we can conclude that for any value of $\alpha > 0$, there exist some positive constants C_0 and C_1 such that

$$\sup_t \Pr(|y_t| > \alpha) \leq C_0 \exp(C_1 \alpha^{s/2}).$$

■

Lemma A.3 *Let x_{it} be a covariate in the active set, $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$ and $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{mt})'$ be the $m \times 1$ vector of conditioning covariates in the DGP, given by (1). Define the projection regression of x_{it} on $\mathbf{z}_t$ as*

$$x_{it} = \bar{\boldsymbol{\psi}}_i' \mathbf{z}_t + \tilde{x}_{it},$$

where $\bar{\boldsymbol{\psi}}_i = (\psi_1, \cdots, \psi_m)'$ is the $m \times 1$ vector of projection coefficients which is equal to $[T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{z}_t \mathbf{z}_t')^{-1}][T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{z}_t x_{it})]$. Under Assumptions 1, 2, and 4, there exist some finite positive constants C_0, C_1 and C_2 such that if $0 < \lambda \leq (s+2)/(s+4)$, then

$$\Pr(|\mathbf{x}_i' \mathbf{M}_z \mathbf{x}_j - \mathbb{E}(\tilde{\mathbf{x}}_i' \tilde{\mathbf{x}}_j)| > \zeta_T) \leq \exp(-C_0 T^{-1} \zeta_T^2) + \exp(-C_1 T^{C_2})$$

and if $\lambda > (s+2)/(s+4)$*, then*

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j)| > \zeta_T) \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2})$$

for all i *and* j*, where* $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$*,* $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$*, and* $\mathbf{M}_z = \mathbf{I} - T^{-1}\mathbf{Z}\hat{\mathbf{\Sigma}}_{zz}^{-1}\mathbf{Z}'$ *with* $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ *and* $\hat{\mathbf{\Sigma}}_{zz} = T^{-1}\sum_{t=1}^{T}(\mathbf{z}_t\mathbf{z}_t')$*.*

Proof. By Assumption 1 we have

$$\mathbb{E}\left[z_{\ell t}z_{\ell' t} - \mathbb{E}(z_{\ell t}z_{\ell' t})|\mathcal{F}_{t-1}\right] = 0.$$

for $\ell, \ell' = 1, 2, \cdots, m$,

$$\mathbb{E}\left[x_{it}x_{jt} - \mathbb{E}(x_{it}x_{jt})|\mathcal{F}_{t-1}\right] = 0,$$

for $i, j = 1, 2, \cdots, N$, and

$$\mathbb{E}\left[z_{\ell t}x_{it} - \mathbb{E}(z_{\ell t}x_{it})|\mathcal{F}_{t-1}\right] = 0,$$

for $\ell = 1, 2, \cdots, m$, $i = 1, 2, \cdots, N$. Moreover, by Assumption 2, for all i, ℓ, and t, x_{it}, and $z_{\ell t}$ have exponential decaying probability tails. Additionally, by Assumption 4 the number of conditioning variables m is finite. Therefore by Lemma S.16, we can conclude that there exist sufficiently large positive constants C_0, C_1, and C_2 such that if $0 < \lambda \leq (s+2)/(s+4)$,

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j)| > \zeta_T) \leq \exp(-C_0T^{-1}\zeta_T^2) + \exp(-C_1T^{C_2})$$

and if $\lambda > (s+2)/(s+4)$

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j)| > \zeta_T) \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2})$$

for all i and j. ■

Lemma A.4 *Let* y_t *be a target variable generated by the DGP given by (1),* $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{mt})'$ *be the* $m \times 1$ *vector of conditioning covariates in DGP(1) and* x_{it} *be a covariate in the active set,* $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$*. Define the projection regression of* x_{it} *on* $\mathbf{z}_t$ *as*

$$x_{it} = \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{i,T} + \tilde{x}_{it},$$

where $\bar{\boldsymbol{\psi}}_{i,T} = (\psi_{1i,T}, \cdots, \psi_{mi,T})'$ *is the* $m \times 1$ *vector of projection coefficients which is equal to* $\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t x_{it})\right]$*. Additionally define the projection regression of* y_t *on* $\mathbf{z}_t$ *as*

$$y_t = \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{y,T} + \tilde{y}_t,$$

where $\bar{\boldsymbol{\psi}}_{y,T} = (\psi_{1y,T}, \cdots, \psi_{my,T})'$ *is equal to* $\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t y_t)]$. *Under Assumptions 1-4, if* $0 < \lambda \leq (s+2)/(s+4)$,

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp(-C_0 T^{-1}\zeta_T^2) + \exp(-C_1 T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp(-C_0 \zeta_T^{s/(s+1)}) + \exp(-C_1 T^{C_2}),$$

for all $i = 1, 2, \cdots, N$; *where* $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$, $\theta_{i,T} = T\bar{\theta}_{i,T} = \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{y}})$, $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$, $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_T)'$, $\mathbf{M}_z = \mathbf{I} - T^{-1}\mathbf{Z}\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\mathbf{Z}'$, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ *and* $\hat{\boldsymbol{\Sigma}}_{zz} = T^{-1}\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'$.

Proof. Note that by Assumption 1 and Lemma A.1, for all i and ℓ, cross products of x_{it}, $z_{\ell t}$ and y_t minus their expected values are martingale difference processes with respect to filtration $\mathcal{F}_{t-1}$. Moreover, by Assumption 2 and Lemma A.2, for all i, ℓ, and t, x_{it}, $z_{\ell t}$ and y_t have exponential decaying probability tails. Additionally, by Assumption 4 the number of conditioning variables m is finite. Therefore by Lemma S.16, we can conclude that there exist sufficiently large positive constants C_0, C_1, and C_2 such that if $0 < \lambda \leq (s+2)/(s+4)$, then

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp(-C_0 T^{-1}\zeta_T^2) + \exp(-C_1 T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$, then

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp(-C_0 \zeta_T^{s/(s+1)}) + \exp(-C_1 T^{C_2}),$$

for all $i = 1, 2, \cdots, N$. ■

Lemma A.5 *Let* y_t *be a target variable generated by equation (1),* $\mathbf{z}_t$ *be the* $m \times 1$ *vector of conditioning covariates in DGP(1) and* x_{it} *be a covariate in the active set,* $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. *Define the projection regression of* y_t *on* $\mathbf{q}_t \equiv (\mathbf{z}_t', x_{it})'$ *as*

$$y_t = \bar{\boldsymbol{\phi}}_{i,T}'\mathbf{q}_t + \eta_{it},$$

where $\bar{\boldsymbol{\phi}}_{i,T} \equiv \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_t\mathbf{q}_t')\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_t y_t)]$ *is the projection coefficients. Under Assumptions 1-4, there exist sufficiently large positive constants* C_0, C_1 *and* C_2 *such that if* $0 < \lambda \leq (s+2)/(s+4)$, *then*

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0 T^{-1}\zeta_T^2) + \exp(-C_1 T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$*, then*

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2}),$$

for all $i = 1, 2, \cdots, N$*; where* $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \cdots, \eta_{iT})'$*,* $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$*, and* $\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_T)'$*.*

Proof. Note that $\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i = \mathbf{y}'\mathbf{M}_q\mathbf{y}$ where $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$. By Lemma A.1 we have

$$\mathbb{E}\left[y_t x_{it} - \mathbb{E}(y_t x_{it})|\mathcal{F}_{t-1}\right] = 0,$$

for $i = 1, 2, \cdots, N$,

$$\mathbb{E}\left[y_t z_{\ell t} - \mathbb{E}(y_t z_{\ell t})|\mathcal{F}_{t-1}\right] = 0,$$

for $\ell = 1, 2, \cdots, m$, and

$$\mathbb{E}\left[y_t^2 - \mathbb{E}(y_t^2)|\mathcal{F}_{t-1}\right] = 0.$$

Moreover, by Assumption 2 and Lemma A.2, for all i, ℓ, and t, x_{it}, $z_{\ell t}$ and y_t have exponential decaying probability tails. Additionally, by Assumption 4 the number of conditioning variables m is finite. Therefore by Lemma S.16, we can conclude that there exist sufficiently large positive constants C_0, C_1, and C_2 such that if $0 < \lambda \leq (s+2)/(s+4)$, then

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0T^{-1}\zeta_T^2) + \exp(-C_1T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$, then

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2}),$$

for all $i = 1, 2, \cdots, N$. ■

Lemma A.6 *Let* y_t *be a target variable generated by equation (1),* $\mathbf{z}_t$ *be the* $m \times 1$ *vector of conditioning covariates in DGP(1) and* x_{it} *be a covariate in the active set* $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$*. Define the projection regression of* x_{it} *on* $\mathbf{z}_t$ *as*

$$x_{it} = \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{i,T} + \tilde{x}_{it},$$

where $\bar{\boldsymbol{\psi}}_{i,T} = (\psi_{1i,T}, \cdots, \psi_{mi,T})'$ *is the* $m \times 1$ *vector of projection coefficients which is equal to* $[T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t\mathbf{z}_t')^{-1}][T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t x_{it})]$*. Additionally define the projection regression of* y_t *on* $\mathbf{z}_t$ *as*

$$y_t = \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{y,T} + \tilde{y}_t,$$

where $\bar{\boldsymbol{\psi}}_{y,T} = (\psi_{1y,T}, \cdots, \psi_{my,T})'$ *is equal to* $\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t y_t)]$. *Lastly, define the projection regression of* y_t *on* $\mathbf{q}_t \equiv (\mathbf{z}_t', x_{it})'$ *as*

$$y_t = \bar{\boldsymbol{\phi}}_{i,T}'\mathbf{q}_t + \eta_{it},$$

where $\bar{\boldsymbol{\phi}}_{i,T} \equiv \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_t\mathbf{q}_t')\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_t y_t)]$ *is the vector of projection coefficients. Consider*

$$t_{i,T} = \frac{T^{-1/2}\mathbf{x}_i'\mathbf{M}_z\mathbf{y}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}},$$

for all $i = 1, 2, \cdots, N$*; where* $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$, $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \cdots, \eta_{iT})'$, $\mathbf{M}_z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$, $\mathbf{M}_q = \mathbf{I} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, $\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_T)'$. *Under Assumptions 1-4, there exist sufficiently large positive constants* C_0, C_1 *and* C_2 *such that*

$$\Pr\left[|t_{i,T}| > c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-C_0 c_p^2(N,\delta)\right] + \exp(-C_1 T^{C_2}), \text{ for } \epsilon_i > \frac{1}{2}$$

where $c_p(N,\delta)$ *is defined by (8),* $\theta_{i,T} = T\bar{\theta}_{i,T} = \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{y}})$, $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$, *and* $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_T)'$. *Moreover, if* $c_p(N,\delta) = o(T^{1/2-\vartheta-c})$ *for any* $0 \leq \vartheta < 1/2$ *and a finite positive constant* c*, then there exist some finite positive constants* C_0 *and* C_1 *such that,*

$$\Pr\left[|t_{i,T}| > c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \geq 1 - \exp(-C_0 T^{C_1}), \text{ for } 0 \leq \vartheta_i < \tfrac{1}{2}.$$

Proof. Let $\sigma_{\eta_i}^2 = \mathbb{E}(T^{-1}\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)$, and $\sigma_{\tilde{x}_i}^2 = \mathbb{E}(T^{-1}\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i)$. We have $|t_{i,T}| = \mathcal{A}_{iT}\mathcal{B}_{iT}$, where,

$$\mathcal{A}_{iT} = \frac{|T^{-1/2}\mathbf{x}_i\mathbf{M}_z\mathbf{y}|}{\sigma_{\eta_i}\sigma_{\tilde{x}_i}},$$

and

$$\mathcal{B}_{iT} = \frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}}.$$

In the first case where $\theta_{i,T} = \ominus(T^{1-\epsilon_i})$ for some $\epsilon_i > 1/2$, by using Lemma S.4 we have

$$\begin{aligned}\Pr\left[|t_{i,T}| > c_p(n,\delta)|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq{}& \Pr\left[\mathcal{A}_{iT} > c_p(N,\delta)/(1+d_T)|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] + \\ & \Pr\left[\mathcal{B}_{iT} > 1 + d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right],\end{aligned}$$

where $d_T \to 0$ as $T \to \infty$. By using Lemma S.6,

$$\begin{aligned}
&\Pr\left[\mathcal{B}_{iT} > 1 + d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \\
&\quad = \Pr\left(\left| \frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}} - 1 \right| > d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i}) \right) \\
&\quad \leq \Pr\left(\left| \frac{(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)}{\sigma_{\eta_i}^2\sigma_{\tilde{x}_i}^2} - 1 \right| > d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i}) \right) \\
&\quad = \Pr\left[\mathcal{M}_{iT} + \mathcal{R}_{iT} + \mathcal{M}_{iT}\mathcal{R}_{iT} > d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right]
\end{aligned}$$

where $\mathcal{R}_{iT} = |(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)/\sigma_{\eta_i}^2 - 1|$ and $\mathcal{M}_{iT} = |(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)/\sigma_{\tilde{x}_i}^2 - 1|$. By using Lemmas S.3 and S.4, for any values of $0 < \pi_i < 1$ with $\sum_{i=1}^3 \pi_i = 1$ and a strictly positive constant, c, we have

$$\begin{aligned}
&\Pr\left[\mathcal{B}_{iT} > 1 + d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \\
&\quad \leq \Pr\left[\mathcal{M}_{iT} > \pi_1 d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] + \Pr\left[\mathcal{R}_{iT} > \pi_2 d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] + \\
&\qquad \Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] + \Pr\left[\mathcal{R}_{iT} > c | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right].
\end{aligned}$$

First, consider $\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right]$, and note that

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] = \Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i - \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i)| > \pi_1\sigma_{\tilde{x}_i}^2 T d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right].$$

Therefore, by Lemma A.3, there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Also note that

$$\Pr\left[\mathcal{R}_{iT} > \pi_2 d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] = \Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \pi_2\sigma_{\eta_i}^2 T d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right].$$

Therefore, by Lemma A.5, there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{R}_{iT} > \pi_2 d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr\left[\mathcal{R}_{iT} > c | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Therefore, we can conclude that there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{B}_{iT} > 1 + d_T | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1})$$

Now consider $\Pr\left[\mathcal{A}_{iT} > c_p(N,\delta)/(1+d_T) | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right]$, which is equal to

$$\begin{aligned}
&\Pr\left(\frac{|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T} + \theta_{i,T}|}{\sigma_{\eta_i}\sigma_{\tilde{x}_i}} > T^{1/2}\frac{c_p(N,\delta)}{1+d_T} | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right) \\
&\quad \leq \Pr\left(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{1+d_T} T^{1/2} c_p(N,\delta) - |\theta_{i,T}| | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right).
\end{aligned}$$

Note that since $\epsilon_i > 1/2$ the first term on the right hand side of the inequality dominate the second one. Moreover, Since $c_p(N,\delta) = o(T^\lambda)$ for all values of $\lambda > 0$, by Lemma A.4, there exists a finite positive constant C_0 such that

$$\Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{y}| > k_1 T^{1/2} c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-C_0 c_p^2(N,\delta)\right],$$

where $k_1 = \frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{1+d_T}$.

Given the probability upper bound for $\mathcal{A}_{iT}$ and $\mathcal{B}_{iT}$, we can conclude that there exist some finite positive constants C_0, C_1 and C_2 such that

$$\Pr\left[|t_{i,T}| > c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-C_0 c_p^2(N,\delta)\right] + \exp(-C_1 T^{C_2}).$$

Let's consider the next case where $\theta_{i,T} = \ominus(T^{1-\vartheta_i})$ for some $0 \leq \vartheta_i < 1/2$. We know that

$$\Pr\left[|t_{i,T}| > c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] = 1 - \Pr\left[t_{i,T} < c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].$$

By Lemma S.8,

$$\begin{aligned}
\Pr\left[|t_{i,T}| < c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \Pr\left[\mathcal{A}_{iT} < \sqrt{1+d_T} c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \\
\Pr\left[\mathcal{B}_{iT} < 1/\sqrt{1+d_T} | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].
\end{aligned}$$

Since $\theta_{i,T} = \ominus(T^{1-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$ and $c_p(N,\delta) = o(T^{1/2-\vartheta-c})$, for any $0 \leq \vartheta < 1/2$, $|\theta_{i,T}| - \sigma_{\eta_i}\sigma_{\tilde{x}_i}[(1+d_T)T]^{1/2} c_p(N,\delta) = \ominus(T^{1-\vartheta_i}) > 0$ and by Lemma S.5, we have

$$\begin{aligned}
&\Pr\left[\mathcal{A}_{iT} < \sqrt{1+d_T} c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \\
&\quad = \Pr\left[\frac{|T^{-1/2}\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - T^{-1/2}\theta_{i,T} + T^{-1/2}\theta_{i,T}|}{\sigma_{\eta_i}\sigma_{\tilde{x}_i}} < \sqrt{1+d_T} c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \\
&\quad \leq \Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > |\theta_{i,T}| - \sigma_{\eta_i}\sigma_{\tilde{x}_i}[(1+d_T)T]^{1/2} c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].
\end{aligned}$$

Therefore, by Lemma A.4, there exist some finite positive constants C_0 and C_1 such that,

$$\Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > |\theta_{i,T}| - \sigma_{\eta_i}\sigma_{\tilde{x}_i}[(1+d_T)T]^{1/2} c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}),$$

and therefore

$$\Pr\left[\mathcal{A}_{iT} < \sqrt{1+d_T}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Now let consider the probability of $\mathcal{B}_{iT}$,

$$\begin{aligned}
&\Pr\left(\mathcal{B}_{iT} < 1/\sqrt{1+d_T}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right) \\
&= \Pr\left(\frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}} < \frac{1}{\sqrt{1+d_T}}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right) \\
&\quad = \Pr\left(\frac{(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)}{\sigma_{\eta_i}^2\sigma_{\tilde{x}_i}^2} > 1+d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right) \\
&\quad \leq \Pr(\mathcal{M}_{iT} + \mathcal{R}_{iT} + \mathcal{M}_{iT}\mathcal{R}_{iT} > d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})),
\end{aligned}$$

where $\mathcal{R}_{iT} = |(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)/\sigma_{\eta_i}^2 - 1|$ and $\mathcal{M}_{iT} = |(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)/\sigma_{\tilde{x}_i}^2 - 1|$. By using Lemmas S.3 and S.4, for any values of $0 < \pi_i < 1$ with $\sum_{i=1}^3 \pi_i = 1$ and a positive constant, c, we have

$$\begin{aligned}
&\Pr\left[\mathcal{B}_{iT} < 1/\sqrt{1+d_T}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \\
&\quad \leq \Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \Pr\left[\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \\
&\qquad \Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \Pr\left[\mathcal{R}_{iT} > c|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].
\end{aligned}$$

Let's first consider the $\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right]$. Note that

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] = \Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_i)| > \pi_1\sigma_{\tilde{x}_i}^2 T d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].$$

So, by Lemma A.3, we know that there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Also note that

$$\Pr\left[\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] = \Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \pi_2\sigma_{\eta_i}^2 T d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].$$

Therefore, by Lemma A.5, there exist some constants C_0 and C_1 such that,

$$\Pr(\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} \neq 0) \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr(\mathcal{R}_{iT} > c|\theta_{i,T} \neq 0) \leq \exp(-C_0 T^{C_1}).$$

Therefore, we can conclude that there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{B}_{iT} < 1/\sqrt{1+d_T}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}).$$

So, overall we conclude that

$$\begin{aligned}\Pr&\left[|t_{i,T}| > c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \\ &= 1 - \Pr\left[t_{i,T} < c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \geq 1 - \exp(-C_0 T^{C_1}).\end{aligned}$$

■

B Appendix B: Technical and Financial Indicators

Our choice of the technical trading indicators is based on the extensive literature on system trading, reviewed by Wilder (1978) and Kaufman (2020). Most of the technical indicators are based on historical daily high, low and adjusted close prices, which we denote by $H_{it}(\tau)$, $L_{it}(\tau)$, and $P_{it}(\tau)$, respectively. These prices refer to stock i in month t, for day τ. Moreover, let D_t^i be the number of trading days, and denote by $D_{l_t}^i$ the last trading day of stock i in month t. For each stock i, monthly high, low and close prices are set to the last trading day of the month, namely $H_{it}(D_{l_t}^i)$, $L_{it}(D_{l_t}^i)$ and $P_{it}(D_{l_t}^i)$, or H_{it}, L_{it}, and P_{it}, for simplicity. The logarithms of these are denoted by h_{it}, l_{it}, and p_{it}, respectively.

The 28 stocks considered in our study are allocated to 19 sectoral groups according to Industry Classification Benchmark. The group membership of stock i is denoted by the set $\mathfrak{g}_i$, which includes all S&P 500 stocks in stock i^{th} group, and $|\mathfrak{g}_i|$ is the number of stocks in the group.

The technical and financial indicators considered are:

1. Return of Stock i (r_{it}):

$$r_{it} = 100(p_{it} - p_{i,t-1}).$$

2. The Group Average Return of Stock i ($\bar{r}_{it}^g$):

$$\bar{r}_{it}^g = |\mathfrak{g}_i|^{-1} \sum_{j \in \mathfrak{g}_i} r_{jt}.$$

3. Moving Average Stock Return of order s ($mar_{it}(s)$): This indicator, which is also known as s-day momentum (see, for example, Kaufman, 2020), is defined as

$$mar_{it}(s) = \mathrm{MA}(r_{it}, s),$$

where $\mathrm{MA}(x_{it}, s)$ is Moving Average of a time-series process x_{it} with degree of smoothness s which can be written as

$$\mathrm{MA}(x_{it}, s) = s^{-1} \sum_{\ell=1}^{s} x_{i,t-\ell}.$$

4. Return Gap ($gr_{it}(s)$): This indicator represents a belief in mean reversion that prices will eventually return to their means (for further details see Kaufman, 2020).

$$gr_{it}(s) = r_{it} - \mathrm{MA}(r_{it}, s).$$

5. Price Gap ($gp_{it}(s)$):

$$gp_{it}(s) = 100\left[p_{it} - \mathrm{MA}(p_{it}, s)\right].$$

6. Realized Volatility (RV_{it}): Let

$$RV_{it} = \sqrt{\sum_{\tau=1}^{D_t^i} \left(R_{it}(\tau) - \bar{R}_{it}\right)^2},$$

where $R_{it}(\tau) = 100\left[P_{it}(\tau)/P_{it}(\tau-1) - 1\right]$, and $\bar{R}_{it} = \sum_{\tau=1}^{D_t^i} R_{it}(\tau)/D_t^i$.

7. Group Realized Volatility (RV_{it}^g)

$$RV_{it}^g = \sqrt{|\mathfrak{g}|^{-1} \sum_{i \in \mathfrak{g}} RV_{it}^2}.$$

8. Moving Average Realized Volatility ($mav_{it}(s)$): "Signals are generated when a price change is accompanied by an unusually large move relative to average volatility" (Kaufman, 2020). The following two indicators are constructed to capture such signals

$$mav_{it}(s) = \mathrm{MA}(RV_{it}, s)$$

9. Realized Volatility Gap ($RVG_{it}(s)$):

$$RVG_{it}(s) = RV_{it} - \mathrm{MA}(RV_{it}, s)$$

10. Percent Price Oscillator($PPO_{it}(s_1, s_2)$):

$$PPO_{it}(s_1, s_2) = 100\left(\frac{\mathrm{MA}(P_{it}, s_1) - \mathrm{MA}(P_{it}, s_2)}{\mathrm{MA}(P_{it}, s_2)}\right)$$

where $s_1 < s_2$.

11. Relative Strength Indicator (RSI_{it}^s): This is a price momentum indicator developed by Wilder (1978) to capture overbought and oversold conditions. Let

$$\Delta P_{it}^+ = \Delta P_{it}\mathrm{I}_{\Delta P_{it}>0}(\Delta P_{it}), \text{ and } \Delta P_{it}^- = \Delta P_{it}\mathrm{I}_{\Delta P_{it}\leq 0}(\Delta P_{it}),$$

where $\Delta P_{it} = P_{it} - P_{i,t-1}$ and $\mathrm{I}_A(x_{it})$ is an indicator function that take a value of one if $x_{it} \in A$ and zero otherwise. Then

$$RS_{it}^s = -\frac{\mathrm{MA}(\Delta P_{it}^+, s)}{\mathrm{MA}(\Delta P_{it}^-, s)},$$

and

$$RSI_{it}^s = 100\left(1 - \frac{1}{1 + RS_{it}^s}\right).$$

12. Williams R ($WILLR_{it}(s)$): This indicator proposed by Williams (1979) to measure buying and selling pressure.

$$WILLR_{it}(s) = -100\left(\frac{\max\limits_{j\in\{1,\cdots,s\}}(h_{i,t-s+j}) - p_{it}}{\max\limits_{j\in\{1,\cdots,s\}}(h_{i,t-s+j}) - \min\limits_{j\in\{1,\cdots,s\}}(l_{i,t-s+i})}\right).$$

13. Average Directional Movement Index ($ADX_{it}(s)$): This is a filtered momentum indicator by Wilder (1978). To compute $ADX_{it}(s)$, we first calculate up-ward directional movement (DM_{it}^+), down-ward directional movement (DM_{it}^-), and true range (TR_{it}) as:

$$DM_{it}^+ = \begin{cases} h_{it} - h_{i,t-1}, & \text{if } h_{it} - h_{i,t-1} > 0 \text{ and } h_{it} - h_{i,t-1} > l_{i,t-1} - l_{it}, \\ 0, & \text{otherwise.} \end{cases}$$

$$DM_{it}^- = \begin{cases} l_{i,t-1} - l_{it}, & if\ l_{i,t-1} - l_{it} > 0\ and\ l_{i,t-1} - l_{it} > h_{it} - h_{i,t-1}, \\ 0, & otherwise. \end{cases}$$

$$TR_{it} = \max\{h_{it} - l_{it}, |h_{it} - p_{i,t-1}|, |p_{i,t-1} - l_{it}|\}.$$

Then, positive and negative directional indexes denoted by $ID_{it}^+(s)$ and $ID_{it}^-(s)$ re-

spectively, are computed by

$$ID_{it}^{+}(s) = 100\left(\frac{\mathrm{MA}(DM_{it}^{+}, s)}{\mathrm{MA}(TR_{it}, s)}\right),$$

$$ID_{it}^{-}(s) = 100\left(\frac{\mathrm{MA}(DM_{it}^{-}, s)}{\mathrm{MA}(TR_{it}, s)}\right),$$

Finally, directional index $DX_{it}(s)$ and $ADX_{it}(s)$ are computed as

$$DX_{it}(s) = 100\left(\frac{|ID_{it}^{+}(s) - ID_{it}^{-}(s)|}{ID_{it}^{+}(s) + ID_{it}^{-}(s)}\right),$$

$$ADX_{it}(s) = \mathrm{MA}(DX_{it}(s), s).$$

14. Percentage Change in Kaufman's Adaptive Moving Average ($\Delta KAMA_{it}(s_1, s_2, m)$): Kaufman's Adaptive Moving Average accounts for market noise or volatility. To compute $\Delta KAMA_{it}(s_1, s_2, m)$, we first need to calculate the Efficiency Ratio (ER_{it}) defined by

$$ER_{it} = 100\left(\frac{|p_{it} - p_{i,t-m}|}{\sum_{j=1}^{m} |\Delta P_{i,t-m+j}|}\right),$$

where $\Delta P_{it} = P_{it} - P_{i,t-1}$, and then calculate the Smoothing Constant (SC_{it}) which is

$$SC_{it} = \left[ER_{it}\left(\frac{2}{s_1+1} - \frac{2}{s_2+1}\right) + \frac{2}{s_2+1}\right]^2,$$

where $s_1 < m < s_2$. Then, Kaufman's Adaptive Moving Average is computed as

$$\mathrm{KAMA}(P_{it}, s_1, s_2, m) = SC_{it}P_{it} + (1 - SC_{it})\mathrm{KAMA}(P_{i,t-1}, s_1, s_2, m)$$

where

$$\mathrm{KAMA}(P_{is_2}, s_1, s_2, m) = s_2^{-1}\sum_{\kappa=1}^{s_2} P_{i\kappa}.$$

The Percentage Change in Kaufman's Adaptive Moving Average is then computed as

$$\Delta KAMA_{it}(s_1, s_2, m) = 100\left(\frac{\mathrm{KAMA}(P_{it}, s_1, s_2, m) - \mathrm{KAMA}(P_{i,t-1}, s_1, s_2, m)}{\mathrm{KAMA}(P_{i,t-1}, s_1, s_2, m)}\right).$$

For further details see Kaufman (2020).

B.1 Other financial indicators

In addition to the above technical indicators, we also make use of daily prices of Brent Crude Oil, S&P 500 index, monthly series on Fama and French market factors, and annualized percentage yield on 3-month, 2-year and 10-year US government bonds. Based on this data, we have constructed the following variables. These series are denoted by PO_t and $P_{sp,t}$ respectively, and their logs by po_t and $p_{sp,t}$. The list of additional variables are:

1. Return of S&P 500 index ($r_{sp,t}$):

$$r_{sp,t} = 100(p_{sp,t} - p_{sp,t-1}),$$

   where $p_{sp,t}$ is the log of $S\&P$ 500 index at the end of month t.

2. Realized Volatility of S&P 500 index ($RV_{sp,t}$):

$$RV_{sp,t} = \sqrt{\sum_{\tau=1}^{D_t^{sp}} \left(R_{sp,t}(\tau) - \bar{R}_{sp,t}\right)^2},$$

   where $\bar{R}_{sp,t} = \sum_{\tau=1}^{D_t^{sp}} R_{it}(\tau)/D_t^{sp}$, $R_{sp,t}(\tau) = 100([P_{sp,t}(\tau)/P_{sp,t}(\tau-1) - 1]$, $P_{sp,t}(\tau)$ is the $S\&P$ 500 price index at close of day τ of month t, and D_t^{sp} is the number of days in month t.

3. Percent Rate of Change in Oil Prices (Δpo_t):

$$\Delta po_t = 100(po_t - po_{t-1}),$$

   where po_t is the log of oil princes at the close of month t.

4. Long Term Interest Rate Spread ($LIRS_t$): The difference between annualized percentage yield on 10-year and 3-month US government bonds.

5. Medium Term Interest Rate Spread ($MIRS_t$): The difference between annualized percentage yield on 10-year and 2-year US government bonds.

6. Short Term Interest Rate Spread ($SIRS_t$): The difference between annualized percentage yield on 2-year and 3-month US government bonds.

7. Small Minus Big Factor (SMB_t): Fama and French Small Minus Big market factor.

8. High Minus Low Factor (HML_t): Fama and French High Minus Low market factor.

A summary of the covariates in the active set used for prediction of monthly stock returns is given in Table B.1.

Table B.1: Active set for percentage change in equity price forecasting

Target Variable:	r_{it+1}
	(one-month ahead percentage change in equity price of stock i)
A. Financial Variables:	r_{it}
	$\bar{r}_{it}^{g}$
	$r_{sp,t}$
	RV_{it}
	RV_{it}^{g}
	$RV_{sp,t}$
	SMB_t
	HML_t
B. Economic Variables:	Δpo_t
	$LIRS_t - LIRS_{t-1}$
	$MIRS_t - MIRS_{t-1}$
	$SIRS_t - SIRS_{t-1}$
C. Technical Indicators:	mar_{it}^{s} for $s = \{3, 6, 12\}$
	mav_{it}^{s} for $s = \{3, 6, 12\}$
	gr_{it}^{s} for $s = \{3, 6, 12\}$
	gp_{it}^{s} for $s = \{3, 6, 12\}$
	RVG_{it}^{s} for $s = \{3, 6, 12\}$
	ADX_{it}^{s} for $s = \{3, 6, 12\}$
	RSI_{it}^{s} for $s = \{3, 6, 12\}$
	$WILLR_{it}^{s}$ for $s = \{3, 6, 12\}$
	$PPO_{it}(s_1, s_2)$ for $(s_1, s_2) = \{(3, 6), (6, 12), (3, 12)\}$
	$\Delta KAMA_{it}(s_1, s_2, m)$ for $(s_1, s_2, m) = (2, 12, 6)$

References

Caner, M. and Knight, K. (2013). An alternative to unit root tests: Bridge estimators differentiate between nonstationary versus stationary models and select optimal lag. *Journal of Statistical Planning and Inference*, 143(4):691–715. https://doi.org/10.1016/j.jspi.2012.10.009.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2):221–241. https://doi.org/10.1016/S0304-4076(97)00115-2.

Chudik, A., Kapetanios, G., and Pesaran, M. H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica*, 86(4):1479–1512. https://doi.org/10.3982/ECTA14176.

Clements, M. and Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge University Press, Cambridge, England. ISBN-10: 0521634806.

Colby, R. W. and Meyers, T. A. (1988). *The Encyclopedia of Technical Market Indicators*. Richard D. Irwin, Illinois, US. ISBN-10: 0070120579.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181. https://doi.org/10.1016/j.jfineco.2012.04.003.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144. https://doi.org/10.1198/073500102753410444.

Diebold, F. X. and Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian Lasso and its derivatives. *International Journal of Forecasting*, 35:1679–1691. https://doi.org/10.1016/j.ijforecast.2018.09.006.

Fan, Y., Lv, J., Sharifvaghefi, M., and Uematsu, Y. (2020). IPAD: Stable interpretable forecasting with knockoffs inference. *Journal of American Statistical Association, forthcoming*. https://doi.org/10.1080/01621459.2019.1654878.

Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12(2-3):385–423. https://doi.org/10.1016/0165-1889(88)90047-4.

Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing : The State Space Approach*. Springer Series in Statistics, Berlin, Germany. ISBN-10: 3540719164.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67. https://doi.org/10.1016/j.jeconom.2016.03.006.

Kapetanios, G. and Zikes, F. (2018). Time-varying Lasso. *Economics Letters*, 169:1–6. https://doi.org/10.1016/j.econlet.2018.04.029.

Kaufman, P. (2020). *Trading Systems and Methods*. John Wiley & Sons, New Jersey, US. ISBN-10: 9781118043561.

Koo, B., Anderson, H. M., Seo, M. H., and Yao, W. (2020). High-dimensional predictive regression in the presence of cointegration. *Journal of Econometrics, forthcoming*. https://doi.org/10.1016/j.jeconom.2020.03.011.

Koop, G. and Potter, S. (2004). Forecasting in dynamic factor models using Bayesian model averaging. *The Econometrics Journal*, 7(2):550–565. https://doi.org/10.1111/j.1368-423x.2004.00143.x.

Lee, S., Seo, M. H., and Shin, Y. (2016). The Lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):193–210. https://doi.org/10.1111/rssb.12108.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73(4):1057–1084. https://doi.org/10.1111/j.1467-937x.2006.00408.x.

Pesaran, M. H. and Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, 29(2):307–318. https://doi.org/10.1198/jbes.2010.09018.

Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152. https://doi.org/10.1016/j.jeconom.2013.04.002.

Pesaran, M. H., Schuermann, T., and Smith, L. V. (2009). Forecasting economic and financial variables with global VARs. *International journal of forecasting*, 25(4):642–675. https://doi.org/10.1016/j.ijforecast.2009.08.007.

Pesaran, M. H. and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4):461–465. https://doi.org/10.2307/1391822.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161. https://doi.org/10.1016/j.jeconom.2006.03.010.

Rossi, B. (2013). Advances in forecasting under instability. In *Handbook of Economic Forecasting*, volume 2B, chapter 21, pages 1203–1324. Elsevier. https://doi.org/10.1016/b978-0-444-62731-5.00021-x.

Sharifvaghefi, M. (2020). Variable selection in high dimensional linear regression setting with high multicollinearity. *unpublished manuscript, available at: https://sites.google.com/view/mahrad/research.*

Stock, J. and Watson, M. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14:11–30. https://doi.org/10.1080/07350015.1996.10524626.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Wilder, J. W. (1978). *New Concepts in Technical Trading Systems.* Trend Research, North Carolina, US. ISBN-10: 0894590278.

Williams, L. R. (1979). *How I Made One Million Dollars ... Last Year ... Trading Commodities.* Windsor Books, Place of publication not identified. ISBN-10: 0930233107.

Yousuf, K. and Ng, S. (2019). Boosting high dimensional predictive regressions with time varying parameters. *arXiv preprint arXiv:1910.03109.* https://arxiv.org/abs/1910.03109.

Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine learning research*, 7(Nov):2541–2563. http://www.jmlr.org/papers/v7/zhao06a.html.

Zheng, Z., Fan, Y., and Lv, J. (2014). High dimensional thresholded regression and shrinkage effect. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):627–649. https://doi.org/10.1111/rssb.12037.