Chudik, Alexander; Pesaran, M. Hashem; Sharifvaghefi, Mahrad

Working Paper

Variable Selection in High Dimensional Linear Regressions with Parameter Instability

CESifo Working Paper, No. 10223

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Chudik, Alexander; Pesaran, M. Hashem; Sharifvaghefi, Mahrad (2023) : Variable Selection in High Dimensional Linear Regressions with Parameter Instability, CESifo Working Paper, No. 10223, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/271867

CESifo WORKING PAPERS

10223
2023

January 2023

Variable Selection in High Dimensional Linear Regressions with Parameter Instability

Alexander Chudik, M. Hashem Pesaran, Mahrad Sharifvaghefi

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10223*

Variable Selection in High Dimensional Linear Regressions with Parameter Instability

Abstract

This paper is concerned with the problem of variable selection when the marginal effects of signals on the target variable as well as the correlation of the covariates in the active set are allowed to vary over time, without committing to any particular model of parameter instabilities. It poses the issue of whether weighted or unweighted observations should be used at the variable selection stage in the presence of parameter instability, particularly when the number of potential covariates is large. Amongst the extant variable selection approaches, we focus on the One Covariate at a time Multiple Testing (OCMT) method. This procedure allows a natural distinction between the selection and forecasting stages. We establish three main theorems on selection, estimation post selection, and in-sample fit. These theorems provide justification for using unweighted observations at the selection stage of OCMT and down-weighting of observations only at the forecasting stage. The benefits of the proposed method as compared to Lasso, Adaptive Lasso and Boosting are illustrated by Monte Carlo studies and empirical applications to forecasting monthly stock market returns and quarterly output growths.

JEL-Codes: C220, C520, C530, C550.

Keywords: parameter instability, high-dimensionality, variable selection, One Covariate at a time Multiple Testing (OCMT).

Alexander Chudik
Federal Reserve Bank of Dallas / TX / USA
alexander.chudik@gmail.com

M. Hashem Pesaran
University of Southern California / USA
pesaran@usc.edu

Mahrad Sharifvaghefi
University of Pittsburgh / PA / USA
sharifvaghefi@pitt.edu

January 10, 2023

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

This paper was previously entitled "Variable Selection and Forecasting in High Dimensional Linear Regressions with Structural Breaks" CESifo WP no. 8475 of July 2020.

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 statistical relationships are subject to parameter instabilities. In an extensive early study, Stock and Watson (1996) find that a large majority of time series regressions in economics are subject to 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/forecasting stage for a given set of selected regressors. However, the theory of variable selection in the presence of time-varying parameters is still largely underdeveloped. In this study, we investigate whether unweighted or weighted observations should be used at the variable selection stage when it is known or suspected that the parameters are subject to change. To address this issue we allow the marginal effects of signals on the target variable as well as the correlation of the covariates in the active set to vary over time, without committing to any particular model of parameter instabilities. We provide theoretical arguments in favor of using unweighted observations at the selection stage, and recommend that one should only consider weighting the observations post selection, at the estimation and forecasting stages.

For a given model specification, many different approaches have been proposed in the literature, basically trading off bias and efficiency of forecasts by considering differing estimation windows or down-weighting schemes. 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 the observations. There are also Bayesian approaches to prediction that allow for the 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 parameter instability.

In the case of models with stable parameters it is optimal to weigh the observations equally for both variable selection and estimation purposes. However, there is little discussion in the literature about 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 use machine learning techniques to allow for parameter instability, in particular 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) suggest recursive application of Lasso using rolling windows. Qian and Su (2016) consider a linear regression model with a finite number of covariates but allow for an unknown number of breaks and use group fused Lasso due to Alaíz et al. (2013) to consistently estimate the number of breaks and their locations. Lee et al. (2016) have proposed a Lasso procedure that allows for threshold effects. Kapetanios and Zikes (2018) have proposed a time-varying Lasso procedure, where all the parameters of the model vary locally. Fan et al. (2014) suggest an extension of the screening procedure initially proposed by Fan and Lv (2008) to the case where the regression coefficients vary smoothly with an observable exposure variable. Also recently, Yousuf and Ng (2021) 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 simultaneously in a single step.

Chudik et al. (2018) propose an alternative procedure called **o**ne **c**ovariate at a time **m**ultiple **t**esting (OCMT). The procedure focuses on the statistical significance of the net effects of the covariates under consideration on the target variable of interest, one-at-a-time, rather than simultaneous consideration of the partial effects of all the covariates, while taking full account of the multiple testing nature of the inferential problem involved. The idea of using one-at-a-time regressions is not unique to OCMT and has been used in boosting as well as in screening approaches. See, for example, Buhlmann (2006) and Fan and Lv (2018) as prominent examples of these approaches. What is unique about the OCMT procedure is its inferentially motivated stopping rule without resorting to the use of information criteria, or penalized regression after the initial stage. In the case of models with stable parameters, Chudik et al. (2018) establish that OCMT asymptotically selects all the relevant covariates and none of the pure noise covariates under a fairly general set of assumptions. Moreover, they provide rates for consistency of the regression error and coefficient norms of the selected model. Finally, using Monte Carlo studies, they show that OCMT tends to perform better than penalized regression or boosting procedures under various designs. Sharifvaghefi (2022) has recently generalized the OCMT procedure to allow the covariates under consideration to be strongly 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 that it allows for a natural separation of the two problems of variable selection and estimation/forecasting. As noted above, the focus of the present paper is the application of OCMT for variable selection in the presence of parameter instability, defined in a broad sense. The paper does not make any contribution to the existing literature on estimation and forecasting once the variable selection stage is completed. Existing theoretical results from the forecasting literature can 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 among the remaining true covariates.

To take account of the time variations in the coefficients of the signals, we consider their time averages and distinguish between strong signals whose average marginal effects tend to a non-zero value, semi-strong signals whose average marginal effects tend to zero, but sufficiently slow, and weak signals whose average marginal effects tend to zero quite fast. In this way we allow for variety of time variations that could arise in practice. Strong signals tend to have non-zero effects at all times, semi-strong signals could have zero effects during some periods, with weak signals enter the model relatively rarely. Weak signals are often indistinguishable from noise variables. In our theoretical analysis we will focus on selection of strong and semi-strong signals.

We provide three main theorems in support of our proposed variable selection method. Under certain fairly general regularity conditions we show that the probability of OCMT selecting the true approximating model that contains all the signals (strong and semi-strong) and none of the noise variables tends to unity as the number of time series observations (T) tends to infinity. Our results apply both when N (the number of covariates in the active set) is fixed as well as when N tends to infinity jointly with T, covering the case where $N >> T$. We also establish conditions under which (a) least squares estimates of the coefficients of selected covariates will tend to zero unless they are signals, and (b) the average square of residuals of the selected model achieves the oracle rate for regression models with time-varying coefficients. These theoretical findings provide a formal justification for application of statistical techniques from the time-varying parameters literature to the post OCMT selected model. Our Monte Carlo experiments indicate that the OCMT procedure with weighted observations only at the estimation stage has appealing finite-sample performance relative to Lasso and Adaptive Lasso (A-Lasso by Zou (2006)), as well as Boosting by Buhlmann (2006), under many different settings. We also found that Lasso and A-Lasso consistently outperform Boosting.

Finally, we consider two empirical applications: forecasting monthly returns of stocks

in Dow Jones and output growths across 33 countries, using OCMT, Lasso and A-Lasso. We did not include Boosting in the empirical applications. The empirical results are in line with our theoretical and MC findings and 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. Overall, based on the empirical results we also find that OCMT with no down-weighting at the selection stage outperforms penalized regression methods, such as Lasso and/or A-Lasso.

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 for variable selection without down-weighting in the presence of parameter instability. Section 4 discusses the technical assumptions and the asymptotic properties of the OCMT procedure under parameter instability. Section 5 gives the details of the Monte Carlo experiments and a summary of the main results. Section 6 presents the empirical applications, and Section 7 concludes. Mathematical proofs are provided in the appendix. The paper is also accompanied with three online supplements: a theory supplement that contains lemmas required to establish the theorems in this paper; an empirical supplement that provides further details on the two empirical applications presented in the paper; and finally a Monte Carlo supplement that provides additional summary tables, the full set of Monte Carlo results, as well as the description of the algorithms used for Lasso, A-Lasso and Boosting.

Notations: Generic finite positive constants are denoted by C_i for $i = 1, 2, \cdots$. $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$ denote the spectral and Frobenius norms of matrix $\mathbf{A}$, respectively. $\text{tr}(\mathbf{A})$ and $\lambda_i(\mathbf{A})$ denote the trace and the i^{th} eigenvalue of a square matrix $\mathbf{A}$, respectively. $\|\mathbf{x}\|$ denotes the ℓ_2 norm of vector $\mathbf{x}$. 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 specification under parameter instability

Consider the following data generating process (DGP) for the target variable, y_t, in terms of the signal variables (x_{it}, for $i = 1, 2, ..., k$)

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

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 pre-selected 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 expected value of its coefficient, $\bar{\beta}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(\beta_{it})$, is not equal to zero. The strength of the signal can be captured by the exponent coefficient α_i in $\bar{\beta}_{i,T} = \ominus(T^{\alpha_i - 1})$. For $\alpha_i = 1$, the signal is strong and $\bar{\beta}_{i,T}$ does not converge to zero. For $1/2 < \alpha_i < 1$, the signal is semi-strong and $\bar{\beta}_{i,T}$ converges to zero, but not too fast. For $0 \leq \alpha_i \leq 1/2$, the signal is weak and $\bar{\beta}_{i,T}$ tend to zero at a fast rate. To simply the exposition, from now on we assume that there exist no weak signals, namely $1/2 < \alpha_i \leq 1$, or setting $\alpha_i = 1 - \vartheta_i$ we have $\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$, and we shall refer to strong and semi-strong signals as signals.

Parameters can vary continuously following a stochastic process as in the standard random coefficient model, $\beta_{it} = \beta_i + \sigma_{it}\xi_{it}$, or could change at discrete time intervals, for example $\beta_{it} = \beta_i^{[s]}$, if $t \in [T_{s-1}, T_s)$ for $s = 1, 2, \cdots, S$, where $T_0 = 1$ and $T_S = T$. The vector $\mathbf{z_t}$ can contain deterministic components such as a constant, dummy variables, and a deterministic time trend as well as stochastic variables including observed common factors. It is assumed that both the structure of the parameter instabilities 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, $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$, 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. We assume the coefficients ($\mathbf{a}_t$, and β_{it}, for $i = 1, 2, ..., k$) are independently distributed of the pre-selected covariates ($\mathbf{z}_t$) and all the covariates in the

active set $\mathcal{S}_{Nt}$.

The application of penalized regression techniques to variable selection is often 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 of presence parameter instability Lasso must be adapted to simultaneously deal with selection and parameter change. We are not aware of any machine learning technique that simultaneously addresses both issues. As noted in the introduction, 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 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 pre-selected covariates, $\mathbf{z}_t$, from x_{it} and y_t, before defining average net effects. To this end consider the auxiliary regressions of x_{it} and y_t on $\mathbf{z}_t$, as defined by

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

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

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

and

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

Using the the filtered series, $\tilde{x}_{it}$ and $\tilde{y}_t$, the average net effect of the covariate x_{it} on y_t, conditional on $\mathbf{z}_t$, can be defined as

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

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

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

Also,

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

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

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

Since the excepted values of coefficients of pre-selected covariates are time-invariant, $\mathbb{E}(\mathbf{a}_t) = \mathbf{a}$ for all t, we can further write $T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{a}_t') \mathbb{E}(\tilde{x}_{it} \mathbf{z}_t) = \mathbf{a}' T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it} \mathbf{z}_t) = 0$. Therefore, the average net effect can be written simply as

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

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

$\bar{\boldsymbol{\psi}}'_{i,T}\left(T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t u_t)\right)$, 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 $\bar{\sigma}_{iu,T}(\mathbf{z}) = O(T^{-\epsilon_i})$, for some $\epsilon_i > 1/2$. In this case the average net effects simplifies to

$$\bar{\theta}_{i,T} = \sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{jt})\sigma_{ij,t}(\mathbf{z})\right) + O(T^{-\epsilon_i})$$

In line with our assumption about the time averages of the marginal effects, namely that $\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$, we distinguish between covariates with strong and semi-strong net effects, and the noise variables whose net effects, averaged over time, tend to zero sufficiently fast. Specifically, for covariates with strong or semi-strong net effects we set $\bar{\theta}_{i,T} = \ominus(T^{-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$, and for the noise variables we shall assume that $\bar{\theta}_{i,T} = \ominus(T^{-\epsilon_i})$, for some $\epsilon_i > 1/2$.

In what follows, we first describe the OCMT procedure and then discuss the conditions under which the approximating model (that includes all the signals and none of the noise variables) is selected with probability one by OCMT.

3 Parameter instability and OCMT

The OCMT procedure begins with N separate regressions, conditional on $\mathbf{z}_t$, for each of the N covariates in the active set $\mathcal{S}_{Nt}$. Specifically, the focus is on the statistical significance of $\phi_{i,T}$ in the following simple regressions:

$$y_t = \boldsymbol{\varrho}'_{i,T}\mathbf{z}_t + \phi_{i,T}x_{it} + \eta_{it}, \text{ for } t = 1, 2, \cdots, T;\ i = 1, 2, ..., N,$$

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 $\bar{\sigma}_{ii,T}(\mathbf{z}) = T^{-1}\sum_{t=1}^{T}\sigma_{ii,t}(\mathbf{z})$.[1]

[1]Under parameter stability, as assumed by Chudik et al. (2018), $\beta_{it} = \beta_i$ for all t, and the average net effects can be simplified to the net effects defined by $\theta_{i,T} = \sum_{j=1}^{k}\beta_j\bar{\sigma}_{ij,T}(z)$, where $\bar{\sigma}_{ij,T}(z) =$

Due to non-zero cross-covariate correlations, knowing whether $\phi_{i,T}$ (or equivalently $\bar{\theta}_{i,T}$) is zero does not necessarily allow us to establish whether $\bar{\beta}_{i,T}$ is sufficiently close to 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*	$\beta_{it} = 0$ for all t and $\bar{\theta}_{i,T} = \ominus(T^{-\vartheta_i})$
(IV) *Noise variables*	$\beta_{it} = 0$ for all t 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 a noise variable, then $\bar{\theta}_{i,T}$ converges to zero very fast. Therefore, down-weighting of observations at the variable selection stage is likely to be inefficient for eliminating the 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 consider a single stage version of the OCMT procedure for variable selection.[2]

The 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 $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$; $\mathbf{y} = \mathbf{Z}\boldsymbol{\varrho}_{i,T} + \phi_{i,T}\mathbf{x}_i + \boldsymbol{\eta}_i$; 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

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

$T^{-1}\sum_{t=1}^{T}\sigma_{ij,t}(z)$.

[2]To allow for hidden signals, Chudik et al. (2018) extend the OCMT method to have multiple stages.

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

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

$$\hat{\mathcal{J}}_i = I\left[|t_{i,T}| > c_p(N,\delta)\right], \text{ for } i = 1,2,\cdots,N. \tag{3}$$

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

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 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 $c_p(N,\delta)$ used for the separate-induced tests is chosen to be an appropriately increasing function of N, by setting $\delta > 0$. The choice of δ is guided by our theoretical derivations, and will be discussed below. See Remark 6.

4 Asymptotic properties of OCMT 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 $m = \dim(\mathbf{z}_t)$ and k, the number of signals, are fixed integers. But we allow the number of pseudo-signals, which we denote by k_T^* , to grow at a sufficiently slow rate relative to N and T. Finally, we define the *approximating model* to be a model that contains all the signals (strong as well as semi-strong), $\{x_{it} : i = 1,2,\cdots,k\}$, and none of the noise variables, $\{x_{it} : k + k_T^* + 1, k + k_T^* + 2, \cdots, N\}$. Clearly, such a model can contain one or more of the pseudo-signals, $\{x_{it} : k+1, k+2, \cdots, k+k_T^*\}$. We start with some technical assumptions in Section 4.1 and then provide the asymptotic properties of the OCMT procedure under parameter instability in Section 4.2.

4.1 Technical assumptions

Let $\mathbf{q}_t = (\mathbf{z}_t', \mathbf{x}_t')'$, be the $(m+N)\times 1$ vector containing the pre-selected variables, $\mathbf{z}_t$, and the set of covariates, $\mathbf{x}_t = (x_{1t}, x_{2t}, \cdots, x_{Nt})'$ under consideration, and define the filtrations: $\mathcal{F}_t^q = \sigma(\mathbf{q}_t, \mathbf{q}_{t-1}, \cdots)$, $\mathcal{F}_t^{\mathrm{a}} = \sigma(\mathbf{a}_t, \mathbf{a}_{t-1}, \cdots)$, $\mathcal{F}_{jt}^{\beta} = \sigma(\beta_{jt}, \beta_{j,t-1}, \cdots)$, for $j = 1, 2, \cdots, k$, and $\mathcal{F}_t^u = \sigma(u_t, u_{t-1}, \cdots)$, and set $\mathcal{F}_t^{\beta} = \cup_{j=1}^{k} \mathcal{F}_{jt}^{\beta}$ and $\mathcal{F}_t = \mathcal{F}_t^q \cup \mathcal{F}_t^{\mathrm{a}} \cup \mathcal{F}_t^{\beta} \cup \mathcal{F}_t^u$. Also consider the following assumptions:

Assumption 1 (Martingale difference processes)

(a) $\mathbb{E}\left[\mathbf{q}_t\mathbf{q}_t' - \mathbb{E}(\mathbf{q}_t\mathbf{q}_t')|\mathcal{F}_{t-1}\right] = 0$ *for* $t = 1, 2, \cdots, T$.

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

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

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

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

Assumption 2 (Exponential decaying probability tails)

There exist sufficiently large positive constants C_0 *and* C_1, *and* $s > 0$ *such that*

(a) $\sup_{j,t} \Pr(|q_{jt}| > \alpha) \leq C_0 \exp(-C_1\alpha^s)$, *for all* $\alpha > 0$.

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

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

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

Assumption 3 (Coefficients of signals)

(a) The number of signals, k, *is a finite fixed integer.*

(b) β_{it}, $i = 1, 2, \cdots, k$, *are distributed independently of* $q_{jt'}$, $j = 1, \cdots, m+N$, *and* $u_{t'}$ *for all* t *and* t'.

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

Assumption 4 (Coefficients of conditioning variables)

(a) The number of conditioning variables, $m = \dim(\mathbf{z}_{it})$, *is finite.*

(b) $\mathrm{a}_{\ell t}$, $\ell = 1, 2, \cdots, m$, *are independent of* $q_{jt'}$,$j = 1, \cdots, m + N$, *and* $u_{t'}$ *for all* t *and* t'.

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

Before presenting the theoretical results, we briefly discuss the pros and cons of our assumptions and compare them with the assumptions typically made in the high-dimensional linear regressions and the time-varying parameters 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), some of these assumptions can be relaxed to allow for weak serial correlation in $z_{\ell t}$, $\mathrm{a}_{\ell t}$, x_{it}, β_{it} and u_t. Note that part (e) of Assumption 1 accommodates both randomly and discretely changing parameter models, since $\mathbb{E}(\beta_{it})$ is allowed to be time varying. For examples, see the time-varying parameter models used in the Monte Carlo experiments.

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 parameter instabilities, but it is required to drive upper and lower probability bounds for selection of the approximating model. 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(a) and 4(a) 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(b) and 4(b) 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 time-varying parameters. Assumptions 3(c) ensures the average value of the coefficients of the signal variables does not approach zero too fast. This is an identification assumption that allows us to distinguish signal from noise variables. Finally, Assumption 4(c) constrains the expected values of coefficients of pre-selected covariates to be time-invariant.

4.2 Theoretical results

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 at the variable selection stage in linear high-dimensional settings subject to parameter instability. 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 noise variables; $\{x_{it} : k + k_T^* + 1, k + k_T^* + 2, \cdots, N\}$. The event of choosing the approximating model is defined by

$$\mathcal{A}_0 = \left\{\sum_{i=1}^{k} \hat{\mathcal{J}}_i = k\right\} \cap \left\{\sum_{i=k+k_T^*+1}^{N} \hat{\mathcal{J}}_i = 0\right\}. \tag{4}$$

Note the the approximating model can contain pseudo-signals. 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_T^* pseudo-signals, and $N - k - k_T^*$ noise variables. Consider the OCMT procedure with the critical value function $c_p(N, \delta)$ given by (2), 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 (4), is given by*

$$\Pr(\mathcal{A}_0) = 1 - O(N^{1-2C_0\delta}) - O[\exp(-N^{C_1\kappa_1})]. \tag{5}$$

See Appendix A.1 for a proof.

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 signals are non-zero or tend to zero sufficiently slowly in T, as defined formally by Assumption 3. In the next step, we focus on estimation of the coefficients of the selected model. To simply the exposition we assume that there are no pre-selected covariates, in which case, the DGP (1) simplifies to

$$y_t = \sum_{i=1}^{k} \beta_{it} x_{it} + u_t = \boldsymbol{\beta}_t' \mathbf{x}_{kt} + u_t, \text{ for } t = 1, 2, \cdots, T, \tag{6}$$

where $\mathbf{x}_{kt} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, and $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \cdots, \beta_{kt})'$. For the next set of results the following additional assumption is also needed.

Assumption 5 (Eigenvalues) *Denote the total number of signals (k) and pseudo-signals (k_T^*) by $\tilde{k}_T$ and let $\mathbf{x}_{\tilde{k}_T,t}$ be the $\tilde{k}_T \times 1$ vector of signals and pseudo-signals. then*

$$\lambda_{\min}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{x}_{\tilde{k}_T,t}\mathbf{x}_{\tilde{k}_T,t}')\right] > c > 0.$$

This assumption ensures that the post OCMT selected model can be consistently estimated subject to certain regularity conditions to be discussed below.

The post OCMT selected model can be written as

$$y_t = \sum_{i=1}^{N} \hat{\mathcal{J}}_i x_{it} b_i + \eta_t$$

where $\hat{\mathcal{J}}_i = I\left[|t_{i,T}| > c_p(N,\delta)\right]$, defined by (3). Also $\sum_{i=1}^{N} \hat{\mathcal{J}}_i = \hat{k}_T$, where $\hat{k}_T$ is the number of covariates selected by OCMT. By Theorem 1 the probability that the selected model contains the signals tends to unity as $T \to \infty$. We can further write

$$y_t = \sum_{i=1}^{N} \hat{\mathcal{J}}_i x_{it} b_i + \eta_t = \sum_{\ell=1}^{\hat{k}_T} \gamma_\ell w_{\ell t} + \eta_t, \tag{7}$$

where $\mathbf{w}_t = \left(w_{1t}, w_{2t}, \cdots, w_{\hat{k}_T t}\right)'$. The least squares (LS) estimator of selected coefficients,

$\boldsymbol{\gamma}_T = \left(\gamma_1, \gamma_2, \cdots, \gamma_{\hat{k}_T}\right)'$, is given by

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\sum_{t=1}^{T}\mathbf{w}_t\mathbf{w}_t'\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\mathbf{w}_t y_t\right), \tag{8}$$

In establishing the rate of convergence of $\hat{\gamma}_T$ we distinguish between two cases: when the vector of signals, $\mathbf{x}_{k,t} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$ is included in $\mathbf{w}_t$ as a subset, and when this is not the case. But we know from Theorem 1 the probability of the latter tends to zero at a sufficiently fast rate. The following theorem provides the conditions under which the estimates of the coefficients of the selected pseudo-signals and signals tend to their mean values, defined formally below.

Theorem 2 *Let the DGP for y_t, $t = 1, 2, \cdots, T$ be given by (6) and write down the regression model selected by the OCMT procedure as (7). Suppose that Assumptions 1-5 hold and the number of pseudo-signals, k_T^*, grow with T such that $k_T^* = \ominus(T^d)$ with $0 \leq d < \frac{1}{2}$. Consider the LS estimator of $\boldsymbol{\gamma}_T = \left(\gamma_1, \gamma_2, \cdots, \gamma_{\hat{k}_T}\right)'$, given by (8).*

(i) If $\mathbb{E}(\beta_{it}) = \beta_i$ for all t, then,

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| = O_p\left(T^{\frac{d-1}{2}}\right),$$

where $\boldsymbol{\gamma}_T^* = (\gamma_1^*, \gamma_2^*, \cdots, \gamma_{\hat{k}_T}^*)'$, and

$$\begin{cases} \gamma_\ell^* \in \boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)', & \text{if } w_{\ell t} \in \mathbf{x}_{kt} \\ \gamma_\ell^* = 0, & \text{otherwise.} \end{cases}$$

(ii) If $\mathbb{E}\left(\mathbf{x}_{\tilde{k}_T,t}\mathbf{x}_{\tilde{k}_T,t}'\right)$ is a fixed time-invariant matrix, where $\tilde{k}_T = k + k_T^*$, then,

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^\diamond\| = O_p\left(T^{\frac{d-1}{2}}\right),$$

where $\boldsymbol{\gamma}_T^\diamond = (\gamma_{1T}^\diamond, \gamma_{2T}^\diamond, \cdots, \gamma_{\hat{k}_T,T}^\diamond)'$, and

$$\begin{cases} \gamma_{\ell,T}^\diamond \in \bar{\boldsymbol{\beta}}_T = (\bar{\beta}_{1T}, \bar{\beta}_{2T}, \cdots, \bar{\beta}_{kT})', & \text{if } w_{\ell t} \in \mathbf{x}_{kt} \\ \gamma_{\ell,T}^\diamond = 0, & \text{otherwise,} \end{cases}$$

and $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{it})$, $i = 1, 2, \cdots, k$.

See Appendix A.2 for a proof.

Remark 1 *The above theorem builds on Theorem 1 and establishes that in the post OCMT selected model only signals will end up having non-zero limiting values, as N and $T \to \infty$, so long as $0 \leq d < 1/2$ and δ is sufficiently large. d controls the rate at which number of pseudo-signals is allowed to rise with T. The latter condition rules out the possibility of signals and pseudo-signals sharing the same unobserved common factors. To deal with such a possibility, following Sharifvaghefi (2022), one can first filter out the common factors using principle components (PC) and then apply the OCMT procedure to the least squares residuals of the regressions of the covariates on one or more of their top PCs.*

Remark 2 *The conditions of Theorem 2 are met in the case of random coefficient models where $\beta_{it} = \beta_i + \sigma_{it}\xi_{it}$, and ξ_{it} are distributed independently of the signals (and of the pre-selected covariates, if any), and the LS estimator of $\boldsymbol{\gamma}_T^*$ is consistent, so long as $0 \leq d < 1/2$. Interestingly, if signal and pseudo-signal variables are generated by a stationary process, and hence they satisfy condition (ii) of Theorem 2, then we can extend the random coefficient model to have time-variant means, and still estimate $\boldsymbol{\gamma}_T^*$ consistently by LS.*

Lastly, we consider the residuals of the post OCMT selected model, estimated by LS, that is

$$\hat{\eta}_t = y_t - \sum_{\ell=1}^{\hat{k}_T} \hat{\gamma}_\ell w_{\ell t}, \text{ for } t = 1, 2, ..., T. \tag{9}$$

To obtain the asymptotic properties of the sum of square of residuals (SSR) of the selected model, $\sum_{t=1}^{T} \hat{\eta}_t^2$, we need the following assumption.

Assumption 6 (Weak time dependence) *$h_{ij,t} = x_{it}x_{jt}(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})$ is weakly correlated over time such that*

$$\sum_{t=1}^{T}\sum_{t'=1}^{T} cov(h_{ij,t}, h_{ij,t'}) = O(T), \text{ for } i, j = 1, 2, .., k,$$

where $cov(.,.)$ is the covariance operator.

Remark 3 *Assumption 6 is a high-level assumption. Here is an example of conditions under which this assumption holds. Suppose, Assumptions 1 and 3 hold, and the cross products of coefficients of the signals follow martingale difference processes such that*

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

Then, $\sum_{t=1}^{T}\sum_{t'=1}^{T} cov(h_{ij,t}, h_{ij,t'}) = O(T)$. *To show this, let* $\tilde{h}_{ij,t} = h_{ij,t} - \mathbb{E}\left(h_{ij,t}\right)$. *We have*

$$\begin{aligned}\sum_{t=1}^{T}\sum_{t'=1}^{T} cov(h_{ij,t}, h_{ij,t'}) &= \sum_{t=1}^{T}\mathbb{E}\left(\tilde{h}_{ij,t}^2\right) + 2\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left(\tilde{h}_{ij,t}\tilde{h}_{ij,t'}\right) \\ &= \sum_{t=1}^{T}\mathbb{E}\left(\tilde{h}_{ij,t}^2\right) + 2\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left[\tilde{h}_{ij,t'}\mathbb{E}\left(\tilde{h}_{ij,t}|\mathcal{F}_{t-1}\right)\right].\end{aligned}$$

But, $\mathbb{E}\left(\tilde{h}_{ij,t}|\mathcal{F}_{t-1}\right) = \mathbb{E}\left(h_{ij,t}|\mathcal{F}_{t-1}\right) - \mathbb{E}\left(h_{ij,t}\right)$ *and under the conditions mentioned in this remark,*

$$\begin{aligned}\mathbb{E}\left(h_{ij,t}|\mathcal{F}_{t-1}\right) &= \mathbb{E}\left(x_{it}x_{jt}|\mathcal{F}_{t-1}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})|\mathcal{F}_{t-1}\right] \\ &= \mathbb{E}\left(x_{it}x_{jt}\right)\left\{\mathbb{E}(\beta_{it}\beta_{jt}|\mathcal{F}_{t-1}) - \bar{\beta}_{jT}\mathbb{E}(\beta_{it}|\mathcal{F}_{t-1}) - \bar{\beta}_{iT}\mathbb{E}(\beta_{jt}|\mathcal{F}_{t-1}) + \bar{\beta}_{iT}\bar{\beta}_{jT}\right\} \\ &= \mathbb{E}\left(x_{it}x_{jt}\right)\left\{\mathbb{E}(\beta_{it}\beta_{jt}) - \bar{\beta}_{jT}\mathbb{E}(\beta_{it}) - \bar{\beta}_{iT}\mathbb{E}(\beta_{jt}) + \bar{\beta}_{iT}\bar{\beta}_{jT}\right\} \\ &= \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})\right] = \mathbb{E}\left(h_{ij,t}\right).\end{aligned}$$

Therefore, $\mathbb{E}\left(\tilde{h}_{ij,t}|\mathcal{F}_{t-1}\right) = 0$. *Hence,* $\sum_{t=1}^{T}\sum_{t'=1}^{T} cov(h_{ij,t}, h_{ij,t'}) = \sum_{t=1}^{T}\mathbb{E}\left(\tilde{h}_{ij,t}^2\right) = O\left(T\right)$.

The following theorem establishes the limiting property of SSR of the post OCMT selected model.

Theorem 3 *Let the DGP for* y_t, $t = 1, 2, \cdots, T$ *be given by (6) and write down the regression model selected by the OCMT procedure as (7). The residual of the selected model, estimated by LS, is given by (9). Suppose that Assumptions 1-6 hold and the number of pseudo-signals,* k_T^*, *grow with* T *such that* $k_T^* = \ominus(T^d)$ *with* $0 \leq d < \frac{1}{2}$.

(i) If $\mathbb{E}(\beta_{it}) = \beta_i$ *for all* t, *then*

$$T^{-1}SSR = \bar{\sigma}^2_{u,T} + \bar{\Delta}_{\beta,T} + O_p\left(T^{-\frac{1}{2}}\right) + O_p\left(T^{d-1}\right), \tag{10}$$

where $\bar{\sigma}^2_{u,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(u_t^2\right)$, *and* $\bar{\Delta}_{\beta,T} = T^{-1}\sum_{t=1}^{T} tr(\mathbf{\Sigma}_{\mathbf{x}_k,t}\mathbf{\Omega}_{\beta,t})$ *are non-negative, with* $\mathbf{\Sigma}_{\mathbf{x}_k,t} \equiv (\sigma_{ijt,x})$, $\mathbf{\Omega}_{\beta,t} \equiv (\sigma_{ijt,\beta})$ *for* $i,j = 1,2,\cdots,k$, *and* $\sigma_{ijt,x} = \mathbb{E}\left(x_{it}x_{jt}\right)$, $\sigma_{ijt,\beta} = \mathbb{E}\left[(\beta_{it}-\beta_i)(\beta_{jt}-\beta_j)\right]$.

(ii) If $\mathbb{E}\left(\mathbf{x}_{\tilde{k}_T,t}\mathbf{x}'_{\tilde{k}_T,t}\right)$ *is a fixed time-invariant matrix, where* $\tilde{k}_T = k + k_T^*$, *then,*

$$T^{-1}SSR = \bar{\sigma}^2_{u,T} + \bar{\Delta}^*_{\beta,T} + O_p\left(T^{-\frac{1}{2}}\right) + O_p\left(T^{d-1}\right), \tag{11}$$

where $\bar{\Delta}^*_{\beta,T} = T^{-1}\sum_{t=1}^{T} tr\left(\mathbf{\Sigma}_{\mathbf{x}_k,t}\mathbf{\Omega}^*_{\beta,t}\right)$ *is non-negative, with* $\mathbf{\Omega}^*_{\beta,t} \equiv \left(\sigma^*_{ijt,\beta}\right)$ *for* $i,j = 1,2,\cdots,k$, *and* $\sigma^*_{ijt,\beta} = \mathbb{E}\left[(\beta_{it}-\bar{\beta}_{i,T})(\beta_{jt}-\bar{\beta}_{j,T})\right]$.

See Appendix A.3 for a proof.

Remark 4 *The condition* $d < \frac{1}{2}$ *in Theorem 3 ensures that the number of pseudo-signals grows sufficiently slowly in* T, *which in turn ensures that* $T^{1-d} < T^{-\frac{1}{2}}$ *and hence from equations (10) and (11), we can conclude that the average of square of residuals* ($T^{-1}SSR$) *of the Post OCMT selected model convergences at the same rate of* $T^{-\frac{1}{2}}$ *under both scenarios (i) and (ii).*

Remark 5 *Results (10) and (11) show that the SSR of the selected model depends on (i) the inherent uncertainty due to the unobserved error term,* u_t, *of the DGP, as given by the term* $\bar{\sigma}^2_{u,T}$, *(ii) the cost (in terms of fit) of ignoring the time variation in the coefficients of the signals,* β_{it}, $i = 1,2,\cdots,k$,*as given by the term* $\bar{\Delta}_{\beta,T}$ *and* $\bar{\Delta}^*_{\beta,T}$, *respectively, and (iii) the traditional* $O_p\left(T^{-1/2}\right)$ *sampling uncertainty, which dominates the additional* $O_p\left(T^{d-1}\right)$ *uncertainty due to inclusion of* $k_T^* = \ominus(T^d)$ *pseudo-signals. Clearly, the time variation cost is present even if only the signals are selected in the selection step. But, the cost could be lower if* $\mathbf{\Omega}_{\beta,t}$ *is close to zero in some periods, or if there are cancelling effects from negative* $\sigma_{ijt,x}$ ($\sigma^*_{ijt,x}$) *when* $\sigma_{ijt,\beta}$ *is positive, namely* $\sigma_{ijt,x}\sigma_{ijt,\beta} < 0$ ($\sigma^*_{ijt,x}\sigma_{ijt,\beta} < 0$), *for some* $i \neq j$ *and some* t. *This finding for the in-sample fit is similar to the results for mean square forecast*

errors (MSFE) in the presence of breaks in the literature, such as Proposition 2 of Pesaran and Timmermann (2007) or equation (20) of Pesaran et al. (2013), where the main focus is to minimize the MSFE by mitigating the cost from the time variation in parameters at the expense of increased sampling uncertainty by weighting the observations, such as the use of optimal estimation windows or down-weighting of observations.

Remark 6 *The above three theorems require the exponent δ in the critical value function, (2), to be sufficiently large such that $\delta > \frac{1}{2C_0}$, for some positive constant C_0. The extensive Monte Carlo Studies in Chudik et al. (2018) suggest that setting $\delta = 1$ preforms well in practice.*

5 Monte Carlo evidence

We use Monte Carlo (MC) techniques to compare finite sample performance of OCMT with and without down-weighting at the selection stage, as well as comparing the OCMT results with those of Lasso, A-Lasso, and Boosting. In these comparisons we consider the number of selected covariates ($\hat{k}_T$), the true positive rate (TPR), the false positive rate (FPR), and the one-step-ahead mean square forecast error (MSFE) of the selected models. Sub-section 5.1 outlines the MC designs, sub-section 5.2 provides a summary of how the OCMT, Lasso, A-Lasso, and Boosting procedures are implemented, and finally sub-section 5.3 presents the main MC findings.

5.1 Simulation design

We consider the following data generating process (DGP):

$$y_t = c_t + \rho_{y,t} y_{t-1} + \sum_{j=1}^{k} \beta_{jt} \tilde{x}_{jt} + \tau_u u_t,$$

where the four signals $\tilde{x}_{jt}$, $j = 1, 2, 3, 4$ have non-zero, time-varying means $\mu_{jt} = \mathbb{E}(\tilde{x}_{jt})$. To simplify the exposition of the DGP we consider the demeaned covariates, $x_{jt} = \tilde{x}_{jt} - \mu_{jt}$ (so

that $\mathbb{E}(x_{jt}) = 0$), and write the DGP equivalently as

$$y_t = d_t + \rho_{y,t} y_{t-1} + \sum_{j=1}^{k} \beta_{jt} x_{jt} + \tau_u u_t, \tag{12}$$

where

$$d_t = c_t + \sum_{j=1}^{k} \beta_{jt} \mu_{jt}. \tag{13}$$

Since c_t is a free parameter, without loss of generality we also treat $\{d_t,\ t = 1, 2, ..., T\}$ as free parameters.

For each MC replication, $r = 1, 2, ..., R$, the target variable, y_t, is generated as random draws using (12). The signal variables x_{jt}, $j = 1, 2, 3, 4$, are unknown and belong to a set $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. The vector of covariates $\mathbf{x}_t = (x_{1t}, x_{2t}, \cdots, x_{Nt})'$ is generated as $\mathbf{x}_t = \boldsymbol{R}_t^{1/2} \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{Nt})'$. $\{\varepsilon_{it}\}$ are generated as AR(1) processes with GARCH(1,1) innovations

$$\varepsilon_{it} = \rho_{i\varepsilon} \varepsilon_{i,t-1} + \left(1 - \rho_{i\varepsilon}^2\right)^{1/2} e_{\varepsilon_i t}, \text{ for } t = 1, 2, \cdots, T, \text{and } i = 1, 2, ..., N,$$

using the starting values $\varepsilon_{i,0} \sim IIDN(0, 1)$. The parameters were generated heterogeneously as independent draws, $\rho_{i\varepsilon} \sim IIDU(0, 0.95)$. $e_{\varepsilon_i t} \sim IIDN\left(0, \sigma^2_{\varepsilon_i,t}\right)$, with $\sigma^2_{\varepsilon_i,t}$ given by

$$\sigma^2_{\varepsilon_i,t} = (1 - \alpha_{1\varepsilon_i} - \alpha_{2\varepsilon_i}) + \alpha_{1\varepsilon_i} e^2_{\varepsilon_i t-1} + \alpha_{2\varepsilon_i} \sigma^2_{\varepsilon_i,t-1},$$

where $\alpha_{1\varepsilon_i} \sim IIDU(0, 0.2)$, and $\alpha_{2\varepsilon_i} \sim IIDU(0.6, 0.75)$. The error terms, $\{u_t\}_{t=1}^T$, in (12) are generated as $IIDN(0, \sigma^2_{ut})$ with σ^2_{ut} following the GARCH(1,1) specification

$$\sigma^2_{ut} = (1 - \alpha_{1u} - \alpha_{2u}) + \alpha_{1u} u^2_{t-1} + \alpha_{2u} \sigma^2_{u,t-1},$$

using $u_0 \sim \mathcal{N}(0, 1)$, $\alpha_{1u} = 0.2$ and $\alpha_{2u} = 0.75$.

As our baseline DGP we consider a model with stable parameters, and set $\beta_{jt} = 1$ for $j = 1, 2, 3, 4$. We also set $c_t = 0$ and $\mu_{jt} = 1$ in (13), which yields $d_t = 4$. In addition,

we set $\rho_{y,t} = 0$ when the baseline model is static and $\rho_{y,t} = 0.3$ when the baseline model is dynamic. In the dynamic case we set $y_0 = (1 - \rho_{y,1})^{-1} d_1$. In the case of models with parameter instability we consider a mixed deterministic-stochastic model and generate β_{jt} as

$$\beta_{jt} = b_{jt} + \tau_{\eta_j} \eta_{jt}, \text{ for } j = 1, 2, 3, 4,$$

where b_{jt} are deterministic and η_{jt} are AR(1) processes with GARCH(1,1) innovations,

$$\eta_{jt} = \rho_{\eta j} \eta_{j,t-1} + \left(1 - \rho_{\eta j}^2\right)^{1/2} e_{\eta_j t},$$

using the starting values $\eta_{j,0} \sim IID\mathcal{N}(0, 1)$, and $\rho_{\eta j} = 0.5$, for all j. $\left\{e_{\eta_j t}\right\}$ follows a normal distribution with mean zero, and variance $\sigma^2_{\eta_j t}$ given by

$$\sigma^2_{\eta_j t} = (1 - \alpha_{1\eta_j} - \alpha_{2\eta_j}) + \alpha_{1\eta_j} e^2_{\eta_j, t-1} + \alpha_{2\eta_j} \sigma^2_{\eta_j, t-1}, \text{ for } j = 1, 2, 3, 4,$$

where $\alpha_{1\eta_j} = 0.2$ and $\alpha_{2\eta_j} = 0.75$. We set τ_{η_j} such that deterministic variations in β_{jt} are quite large relative to the stochastic variations. To this end we set τ_{η_j} (using simulations) so that

$$\frac{T^{-1} \sum_{t=1}^{T} b_{jt}^2}{T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[\left(\beta_{jt}^{(r)}\right)^2\right]} = 0.95, \text{ for } j = 1, 2, 3, 4.$$

For the deterministic components of the slope coefficients (b_{jt}, for $j = 1, 2, 3, 4$), we consider the following specifications

$$b_{1t} = b_{2t} = \begin{cases} 2 & \text{if } t \in \{1, 2, \cdots, [T/3]\}, \\ 0 & \text{if } t \in \{[T/3] + 1, [T/3] + 2, \cdots, [2T/3]\}, \\ 1 & \text{if } t \in \{[2T/3] + 1, [2T/3] + 2, \cdots, T\}, \end{cases} \tag{14}$$

and

$$b_{3t} = b_{4t} = \begin{cases} 0.5 & \text{if } t \in \{1, 2, \cdots, [T/2]\}, \\ 1.5 & \text{if } t \in \{[T/2]+1, [T/2]+2, \cdots, T\}, \end{cases} \tag{15}$$

where [.] is the nearest integer function.

We also set $c_t = 0$ in (13) and generate the intercept as $d_t = \sum_{j=1}^{k} \beta_{jt}\mu_{jt}$, where

$$\mu_{1t} = \mu_{2t} = \begin{cases} 0.6 & \text{if } t \in \{1, 2, \cdots, [T/3]\}, \\ 1.5 & \text{if } t \in \{[T/3]+1, [T/3]+2, \cdots, [2T/3]\}, \\ 0.9 & \text{if } t \in \{[2T/3]+1, [2T/3]+2, \cdots, T\}, \end{cases} \tag{16}$$

and

$$\mu_{3t} = \mu_{4t} = \begin{cases} 0.9 & \text{if } t \in \{1, 2, \cdots, [T/2]\}, \\ 1.1 & \text{if } t \in \{[T/2]+1, [T/2]+2, \cdots, T\}, \end{cases} \tag{17}$$

In this design, the jumps in b_{jt} and μ_{jt}, for $j = 1, 2$, have opposite signs and the jumps in b_{jt} and μ_{jt}, for $j = 3, 4$, have the same sign.

The $N \times N$ correlation matrix of the covariates, $\mathbf{R}_t \equiv (r_{ij,t})$, are set as $r_{ij,t} = r_t^{|i-j|}$, for all $i, j = 1, 2, \cdots, N$. We allow for a break in the correlation matrix

$$r_t = \begin{cases} 0.8 & \text{if } t \in \{1, 2, \cdots, [T/2]\}, \text{ and} \\ 0.4 & \text{if } t \in \{[T/2]+1, [T/2]+2, \cdots, T\}. \end{cases}$$

Also, we consider two possibilities for $\rho_{y,t}$. In the static scenario we set $\rho_{y,t} = 0$ for all t. In the dynamic scenario we allow for a switch in $r_{y,t}$ and set it as

$$\rho_{y,t} = \begin{cases} 0.2 & \text{if } t \in \{1, 2, \cdots, [T/2]\}, \\ 0.4 & \text{if } t \in \{[T/2]+1, [T/2]+2, \cdots, T\}. \end{cases} \tag{18}$$

For the static and dynamic models with parameter instabilities, the parameter τ_u is calibrated

by simulations to ensure that the R-squared of the linear regression of y_t on a constant term, the signal variables $\{x_{1t}, x_{2t}, x_{3t}, x_{4t}\}$, and (in experiments with $\rho_{y,t} \neq 0$) the lagged dependent variable is equal to 30% (low fit) and 50% (high fit). The same value of τ_u is used for the corresponding static and dynamic models without parameter instabilities.

We base the MC results on $R = 2,000$ replications, and consider $N \in \{20, 40, 100\}$ and $T \in \{100, 200, 500\}$, combinations. These choices of (N, T) cover our empirical applications. For each pair of (N, T), there are four experiments in case of the models with no parameter instabilities, and four experiments in the case of models with parameter instabilities, corresponding to the two choices of τ_u (low and high fit), ρ_{yt} (static to dynamic). In total, we carry out eight different experiments.

5.2 Selection and estimation methods using weighted and unweighted observations

Let $\mathbf{w}_t = (\mathbf{x}_t', y_t)'$, $t = 1, 2, \cdots, T$ be the (unweighted) set of available observations, and denote the corresponding set of down-weighted observations by $\hat{\mathbf{w}}_t(\lambda) = \lambda^{T-t}\mathbf{w}_t$ where $0 < \lambda \leq 1$ is the down-weighting coefficient.[3] We will consider the following selection/estimation methods: (1) OCMT with down-weighted observations $\{\hat{\mathbf{w}}_t(\lambda)\}_{t=1}^T$ used at both selection and estimation stages; (2) OCMT with the unweighted observations, $\{\mathbf{w}_t\}_{t=1}^T$, used at the selection stage and down-weighted observations, $\{\hat{\mathbf{w}}_t(\lambda)\}_{t=1}^T$, used at the estimation stage; (3) OCMT using unweighted observations, $\{\mathbf{w}_t\}_{t=1}^T$, at both selection and estimation stages; (4,5 & 6) Lasso, A-Lasso, and Boosting also using unweighted observations, $\{\mathbf{w}_t\}_{t=1}^T$; and (7,8 & 9) Lasso, A-Lasso, and Boosting with down-weighted observations, $\{\hat{\mathbf{w}}_t(\lambda)\}_{t=1}^T$ used as inputs.[4]

[3] We are not arguing for the use of exponential down-weighting – but use it as an example. There are also non-exponential type down-weighting schemes that one can use, e.g. Pesaran et al. (2013)

[4] We also consider a two-step procedures based on Lasso, A-Lasso and Boosting. In the first step, we apply Lasso, A-Lasso and Boosting to the original (unweighted) observations and select the variables with non-zero coefficients. In the second step, we estimate the corresponding post-selected model by LS using the weighted observations. Overall, the MSFEs of these procedures were higher than that of direct application

We consider two sets of values for the down-weighting coefficient, λ: (1) Light down-weighting with $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$, and (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 (light/heavy) we focus on simple average forecasts computed over the individual forecasts obtained for each value of λ in the set under consideration.

5.3 Simulation results

A summary of the main results are provided in Tables 1 to 3, with additional summary tables highlighting the effects of down-weighting at the selection stage, and the differences between static versus dynamic models provided in the online MC supplement. Table 1 give the number of selected covariates ($\hat{k}_T$), TPR and FPR of OCMT, Lasso, A-Lasso and Boosting without down-weighting. Panel A of this table reports the results for different N and T combinations, averaged across the four experiments without parameter instabilities, and panel B of the table gives the corresponding results for the four experiments with parameter instabilities. The results show that all the methods under consideration have higher average TPR for models with stable parameters compared to the ones with parameter instabilities. This is to be expected, as the models with parameter instabilities are subject to an additional source of uncertainty.

We further observe that the lower average TPR of OCMT in the models with parameter instabilities is associated with a lower average number of selected covariates, and hence a lower average FPR. On the other hand, the other procedures tend, on average, to select more covariates in the models with parameter instabilities and hence have a higher average FPR relative to the models without parameter instabilities. Lastly, OCMT selects fewer covariates relative to Lasso, A-Lasso, and Boosting, while maintaining the TPR at a similar level. As a result, OCMT has the lowest average FPR among the selection methods under

of Lasso, A-Lasso and Boosting to the weighted observations. The results are available in Section S-2 of the online MC supplement.

consideration across all N and T combinations, with one exception, namley in the case of models with stable parameters for $T = 500$ and $N = 20$, where the FPR for A-Lasso (0.11) is slighly lower than that of OCMT (0.13). Summary Tables S.1 and S.2 in the online MC supplement provide further results on the effects of down-weighting on TPR and FPR. The results consistently show that down-weighting of observations provides no gains for OCMT in terms of average TPR and FPR. This is also true for other methods in majority but not all cases.

Table 2 focusses on the one-step-ahead MSFEs and provides comparative results on the effects of down-weighting across the methods (OCMT, Lasso, A-Lasso and Boosting). As in Table 1, Panel A of Table 2 gives average MSFEs for the four experiments without parameter instabilities, and Panel B gives the corresponding results for the experiments with parameter instabilities. As expected, in the absence of parameter instabilities, using unweighted observations gives the lowest MSFE across all the methods. Moreover, for all N and T combinations and different down-weighting scenarios, the average MSFE of each method is lower in the case of models with stable parameters as compared to those with parameter instabilities. This observation is consistent with our finding in Theorem 3 about the cost of time-variation in the coefficients on the in-sample fit of the estimated model. As can be seen, for models with parameter instabilities, down-weighting does improve the forecasting performance of OCMT (with and without down-weighting in the selection stage), Lasso, A-Lasso, and Boosting. However, by comparing the MSFEs of OCMT with and without down-weighting at the selection stage, we see that the down-weighting at the selection stage always results in deterioration of the forecast accuracy of OCMT, which is in line with our main theoretical result. Last but not least, the results in Table 2 show that OCMT with down-weighting only at the estimation stage has the lowest average MSFE among all the methods for all choices of N, T, and different down-weighting scenarios. In fact, in the case of experiments with parameter instabilities OCMT with down-weighting (light or heavy) at

the estimation stage only, always beats Lasso, A-Lasso and Boosting with light or heavy down-weighting in terms of the one-step-ahead MSFE.

Table 3 compares the performance of OCMT with the down-weighting option at the estimation stage to that of the other procedures, using the same set of down-weighting parameter (λ). Specifically, we report the MSFE of Lasso, A-Lasso, and Boosting relative to that of OCMT. Since the relative MSFE ranking of OCMT, Lasso, A-Lasso, and Boosting does not appear to be affected by no/light/heavy down-weighting options, as a summary measure, we simply average relative MSFE values across individual experiments and the three (no/light/heavy) down-weighting options. However, we provide the relative MSFE results for the models without and with paramter instabilities separately, on left and right panels of Table 3. Two observations stand out from this table. First, the reported average relative MSFEs are greater than one for all the N and T choices, indicating that OCMT outperforms Lasso, A-Lasso, and Boosting in all cases. Second, the degree of the outperformance of OCMT over Lasso and A-Lasso generally increases (particularly for $T > 100$) with paramter instability. This is less so if we compare OCMT with Boosting.

Tables S.4, S.5, and S.6 in the online MC supplement provide further details about the performance of the methods under consideration in static and dynamic experiments. In Table S.4, we compare the number of selected covariates, the TPR, and the FPR of each method without down-weighting across static and dynamic models. For various N and T combinations the reported results are averaged across four experiments (with/without parameter instabilities and with/without high-fit). The results show that all the methods tend to select fewer covariates in the dynamic models relative to the static ones, and hence have a lower TPR and FPR. This is expected, as in the dynamic models, part of the variation in the target variable is explained by its own lag rather than the signal variables. Consequently, in Tables S.5 and S.6, which are about the MSFE in static and dynamic models, respectively, we see that all the methods have a higher MSFE in dynamic models relative to the static

ones. Additionally, the results in Tables S.5 and S.6 show that the MSFE for models with stable parameters is always lower than the ones with parameter instabilities, regardless of whether the model is static or not.

Overall, the results of our MC studies suggest that the OCMT procedure without down-weighting at the selection stage is a useful method to deal with variable selection in linear regression settings with parameter instability.

6 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 A-Lasso. In particular, we consider the following two applications:[5]

- Forecasting monthly rate of price changes for 28 (out of 30) stocks in Dow Jones using a relatively large number of financial, economic, as well as technical indicators.
- 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 A-Lasso, with and without down-weighting. For down-weighting we make use of exponentially down-weighted observations, namely $\hat{x}_{it}(\lambda) = \lambda^{T-t}x_{it}$, and $\hat{y}_t(\lambda) = \lambda^{T-t}y_t$, 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 the same two sets

[5]We also consider forecasting Euro Area quarterly output growth using the European Central Bank (ECB) survey of professional forecasters as our third application. The results of this application can be found in Section S-3 of the online empirical supplement.

of values for the degree of exponential decay, λ, as in the MC section: (1) Light down-weighting with $\lambda = \{0.975, 0.98, 0.985, 0.99, 0.995, 1\}$, and (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 again 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 $q_{lt} \equiv e_{ltA}^2 - e_{ltB}^2$ be the difference in the squared forecasting errors of procedures A and B, for the target variable y_{lt} $(l = 1, 2, ..., L)$ and $t = 1, 2, ..., T_l^f$, where T_l^f 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 $\varepsilon_{lt} \sim \mathcal{N}(0, \sigma_l^2)$. Then under the null hypothesis of $H_0 : \alpha_l = 0$ for all l we have

$$\overline{DM} = \frac{\bar{q}}{\sqrt{V(\bar{q})}} \overset{a}{\sim} \mathcal{N}(0,1), \text{ for } T_{Lf} \to \infty, \text{ where } T_{Lf} = \sum_{l=1}^{L} T_l^f, \bar{q} = T_{Lf}^{-1} \sum_{l=1}^{L} \sum_{t=1}^{T_l^f} q_{lt}, \text{ and}$$

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

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 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 significance 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

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

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

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.

6.1 Forecasting monthly returns 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.[6] 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 ($T = 120$). The active set for predicting $r_{i,t+1}$ consists of 40 financial, economic, and technical variables.[7] The full list and the description of the indicators considered can be found in Section S-1 of online empirical supplement.

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 4 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 5 shows that the OCMT procedure with no down-weighting at the selection stage dominates Lasso and A-Lasso in terms of MSFE and the differences are statistically highly significant.

Further, OCMT outperforms Lasso and A-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 6. As can be seen from this table, OCMT with no down-weighting performs the best; correctly predicting

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

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

the direction of 56.057% of 8,659 forecasts, as compared to 55.33%, which we obtain for Lasso and A-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 A-Lasso. As can be seen from the last column of Table 6, Lasso and A-Lasso on average select many more covariates than OCMT (1-3 variables as compared to 0.072 for OCMT).

So far we have focused on average performance across all the 28 stocks. Table 7 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 A-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 A-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 5 and 6) is not driven by a few stocks and holds more generally.

6.2 Forecasting quarterly output growth rates across 33 countries

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 $2016Q4$ taken from the GVAR dataset.[8] 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. We adopt the following direct forecasting equations:

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

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

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 in Table S.2 in the online empirical supplement. 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 A-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 A-Lasso. Table 12 summarizes country-specific MSFE and DM findings for OCMT relative to Lasso and A-Lasso. The results show OCMT under-performs Lasso in more than half of the countries for one-year ahead horizon, but outperforms Lasso and A-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 statistically significantly better. 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 A-Lasso. Overall, OCMT has a slightly higher MDFA and hence predicts the direction of real output growth better than Lasso and A-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 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, A-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.

7 Conclusion

The penalized regression approach has become the *de facto* benchmark in the literature on variable selection in the context of linear regression models. But, barring a few exceptions (such as Kapetanios and Zikes, 2018), these studies focus on models with stable parameters, and do not consider the implications of parameter instabilities for variable selection.

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, 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 whether to down-weight and how to down-weight the observations at the forecasting stage.

We have provided theoretical arguments for using the unweighted observations at the selection stage of OCMT, and down-weighted observations 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 suggest 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 MC results suggest that overall OCMT without down-weighting at the selection stage outperforms penalized regression methods such as Lasso and A-Lasso, as well as Boosting which tend to be prone to over-fitting.

Table 1: The number of selected variables ($\hat{k}_T$), True Positive Rate (TRP), and False Positive Rate (FPR) averaged across Monte Carlo experiments with and without parameter instabilities.

	$\hat{k}_T$			TPR			FPR		
$N\backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Without parameter instabilities									
	OCMT								
20	3.91	5.15	6.68	0.80	0.95	1.00	0.03	0.07	0.13
40	3.69	5.01	6.52	0.77	0.94	1.00	0.02	0.03	0.06
100	3.47	4.74	6.26	0.73	0.91	1.00	0.01	0.01	0.02
	Lasso								
20	6.93	7.33	7.54	0.86	0.94	0.99	0.18	0.18	0.18
40	8.48	8.89	8.99	0.83	0.93	0.99	0.13	0.13	0.13
100	11.17	10.95	10.98	0.80	0.91	0.98	0.08	0.07	0.07
	A-Lasso								
20	5.39	5.81	6.06	0.77	0.88	0.97	0.12	0.11	0.11
40	6.75	7.28	7.45	0.76	0.89	0.97	0.09	0.09	0.09
100	9.27	9.45	9.70	0.75	0.88	0.97	0.06	0.06	0.06
	Boosting								
20	9.19	9.58	9.74	0.90	0.95	0.99	0.28	0.29	0.29
40	16.04	16.52	16.78	0.89	0.96	0.99	0.31	0.32	0.32
100	35.32	36.64	37.72	0.88	0.95	0.99	0.32	0.33	0.34
B. With parameter instabilities									
	OCMT								
20	3.25	4.55	5.93	0.69	0.90	0.99	0.02	0.05	0.10
40	3.10	4.41	5.85	0.66	0.88	0.99	0.01	0.02	0.05
100	2.96	4.23	5.71	0.61	0.85	0.99	0.01	0.01	0.02
	Lasso								
20	7.60	8.39	9.20	0.78	0.89	0.97	0.22	0.24	0.27
40	10.16	11.71	12.83	0.75	0.88	0.97	0.18	0.20	0.22
100	14.54	16.61	19.82	0.72	0.85	0.96	0.12	0.13	0.16
	A-Lasso								
20	5.84	6.58	7.40	0.68	0.81	0.93	0.16	0.17	0.18
40	7.97	9.40	10.51	0.68	0.82	0.94	0.13	0.15	0.17
100	11.58	13.70	16.73	0.66	0.81	0.94	0.09	0.10	0.13
	Boosting								
20	9.76	10.39	10.95	0.84	0.92	0.98	0.32	0.34	0.35
40	17.50	18.72	19.50	0.84	0.92	0.98	0.35	0.38	0.39
100	37.33	40.37	43.22	0.83	0.91	0.98	0.34	0.37	0.39

Notes: There are $k = 4$ signal variables out of N observed covariates. The reported results for OCMT, Lasso, A-Lasso, and Boosting in the table are based on the original (not down-weighted) observations. Each experiment is based on 2000 Monte Carlo replications. See Section 5 for the detailed description of the Monte Carlo design.

Table 2: The effects of down-weighting on one-step-ahead MSFE of OCMT, Lasso, A-Lasso and Boosting averaged across all MC experiments with and without parameter instabilities.

Down-weighting†:	No	Light	Heavy	No	Light	Heavy	No	Light	Heavy
$N\backslash T$		**100**			**200**			**300**	
A. Without parameter instabilities									
	OCMT(Down-weighling only at the estimation stage)								
20	**29.35**	30.16	31.47	**24.08**	24.76	25.90	**22.97**	24.00	25.53
40	**26.84**	27.22	28.03	**29.45**	31.00	32.77	**22.98**	23.70	25.12
100	**27.37**	27.57	28.29	**26.72**	27.22	28.68	**22.91**	24.45	26.59
	OCMT(Down-weighling at the variable selection and estimation stages)								
20	**29.35**	30.30	32.63	**24.08**	25.31	28.44	**22.97**	25.07	30.62
40	**26.84**	27.30	29.63	**29.45**	31.82	38.12	**22.98**	27.52	41.30
100	**27.37**	28.12	31.56	**26.72**	28.88	36.87	**22.91**	33.50	62.79
	Lasso								
20	**29.54**	31.02	33.03	**24.29**	25.26	27.02	**23.02**	24.31	26.21
40	**27.39**	28.66	31.70	**29.49**	31.38	35.20	**23.10**	24.84	27.83
100	**28.13**	30.70	34.19	**27.09**	29.29	32.73	**23.12**	26.34	30.70
	A-Lasso								
20	**30.85**	32.20	34.40	**24.95**	25.82	28.01	**23.14**	24.72	26.99
40	**29.34**	30.47	33.70	**30.48**	32.62	36.27	**23.58**	25.71	29.36
100	**32.28**	34.76	37.97	**29.51**	31.97	35.50	**23.86**	28.35	33.14
	Boosting								
20	**31.69**	35.26	40.62	**25.07**	28.70	34.11	**23.54**	28.08	32.73
40	**30.11**	34.24	38.90	**31.83**	38.98	43.98	**23.91**	30.32	35.23
100	**35.20**	39.43	42.23	**31.33**	36.17	39.97	**24.78**	33.06	37.88
B. With parameter instabilities									
	OCMT(Down-weighling only at the estimation stage)								
20	33.68	**32.77**	33.24	27.18	**25.75**	26.51	26.26	**24.58**	26.04
40	30.55	**29.54**	29.77	32.65	**31.93**	33.22	25.68	**24.23**	25.47
100	31.60	**30.75**	30.90	30.68	**28.96**	30.00	26.63	**25.00**	26.84
	OCMT(Down-weighling at the variable selection and estimation stages)								
20	33.68	**33.38**	34.79	27.18	**26.48**	29.45	26.26	**26.00**	31.46
40	30.55	**30.08**	32.15	**32.65**	33.29	40.09	**25.68**	28.63	42.35
100	31.60	**31.46**	35.07	**30.68**	31.10	40.90	**26.63**	34.56	64.47
	Lasso								
20	34.50	**34.48**	35.64	27.84	**26.76**	28.28	26.58	**25.25**	27.31
40	31.31	**31.26**	33.65	33.27	**33.01**	36.52	26.30	**26.09**	29.17
100	**32.24**	33.68	36.73	31.76	**31.52**	34.56	**27.18**	27.63	31.98
	A-Lasso								
20	35.79	**35.35**	36.76	28.28	**26.99**	29.06	26.60	**25.49**	28.02
40	33.29	**33.05**	35.76	34.34	**34.04**	37.73	**26.46**	26.69	30.62
100	**36.14**	37.72	40.65	34.54	**34.29**	37.60	**27.95**	29.59	34.51
	Boosting								
20	**35.23**	37.90	42.76	**27.68**	29.95	35.67	**26.42**	29.29	34.28
40	**33.68**	37.73	42.27	**34.26**	40.31	45.49	**26.21**	31.80	37.08
100	**37.49**	42.43	45.35	**34.32**	38.41	42.21	**27.68**	34.52	39.77

Notes: The reported results are averaged across four experiments (with/without dynamics and with/with high-fit) for models with and without parameter instabilities. See Section 5 for the description of the Monte Carlo design. Full set of results is presented in the online Monte Carlo supplement.

†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.

Table 3: One-step-ahead MSFE of Lasso, A-Lasso and Boosting relative to OCMT averaged across MC experiments with and without parameter instabilities.

$N\backslash T$	**100**	**200**	**500**	**100**	**200**	**500**
	A. Without parameter instabilities			B. With parameter instabilities		
	Lasso					
20	1.038	1.024	1.015	1.067	1.047	1.033
40	1.072	1.027	1.051	1.090	1.062	1.081
100	1.128	1.083	1.080	1.119	1.111	1.116
	A-Lasso					
20	1.077	1.052	1.031	1.093	1.061	1.042
40	1.144	1.068	1.089	1.157	1.101	1.107
100	1.268	1.179	1.149	1.244	1.211	1.187
	Boosting					
20	1.191	1.182	1.160	1.168	1.186	1.175
40	1.260	1.215	1.241	1.275	1.226	1.262
100	1.408	1.304	1.279	1.338	1.297	1.302

Notes: This table reports MSFE of Lasso, A-Lasso and Boosting relative to MSFE of OCMT. Relative MSFE values are averaged across experiments and across the three options for down-weighting: no down-weighting (for all methods), light down-weighting of observations prior to Lasso, A-Lasso and Boosting procedures relative to OCMT with light down-weighting only at the estimation stage, and heavy down-weighting of observations prior to Lasso, A-Lasso and Boosting methods compared with OCMT with heavy down-weighting only at the estimation stage. See Section 5 for the description of the Monte Carlo design. Full set of results is presented in the online Monte Carlo supplement.

Table 4: 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 S.3.

Table 5: Mean square forecast error (MSFE) and panel DM test of OCMT versus Lasso and A-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 6: Mean directional forecast accuracy (MDFA) and the average number of selected variables ($\hat{k}$) of OCMT, Lasso and A-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
A-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 7: The number of stocks out of the 28 stocks in Dow Jones index where OCMT outperforms/underperforms Lasso, and A-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 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) 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 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 A-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 A-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 A-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
A-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
A-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, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

Table 13: Mean directional forecast accuracy (MDFA) and PT test of OCMT, Lasso and A-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
A-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 A-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
A-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, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

A Appendix A: Mathematical Derivations

This appendix provides the proofs of Theorems 1 to 3. The proofs are based on lemmas presented in the online theory supplement. Among these, Lemmas S-1.6 and S-1.7 are key. For each covariate $i = 1, 2, \cdots, N$, Lemmas S-1.6 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 $0 \leq \vartheta_i < 1/2$. For DGP given by (6), Lemma S-1.7 provides asymptotic properties of LS estimator of coefficients and SSR of a regression model that includes all the signals and pseudo-signals. This lemma establishes that the coefficients of pseudo-signals estimated by LS converges to zero so long as $k_T^* = \ominus(T^d)$ grows at a slow rate relative to T, i.e. $0 \leq d < 1/2$. This lemma also shows that the SSR of the regression model converges to that of the oracle model, which includes only the signals.

Additional notations and definitions: Throughout this appendix we consider the following events:

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

where $\{\hat{\mathcal{J}}_i \text{ for } i = 1, 2, \cdots, N\}$ are the selection indicators defined by (3). $\mathcal{A}_0$ is the event of selecting the approximating model, defined by (4). $\mathcal{H}$ is the event that all signals are selected, and $\mathcal{G}$ is the event that no 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 (5), 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 $\mathcal{G}^c = \{\sum_{i=k+k_T^*+1}^{N} \hat{\mathcal{J}}_i > 0\}$. Let's consider $\Pr(\mathcal{H}^c)$ and $\Pr(\ \mathcal{G}^c)$ in turn. We have $\Pr(\mathcal{H}^c) \leq \sum_{i=1}^{k} \Pr(\hat{\mathcal{J}}_i = 0)$. 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 S-1.6, 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

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

But for any noise variable $\Pr(\hat{\mathcal{J}}_i = 1) = \Pr\left[|t_{i,T}| > c_p(N,\delta)|\bar{\theta}_{i,T} = \ominus(T^{-\epsilon_i})\right]$, where $\epsilon_i > 1/2$ and hence by Lemma S-1.6, we can conclude that there exist sufficiently large positive constants C_0, C_1 and C_2 such that $\Pr(\hat{\mathcal{J}}_i = 1) \leq \exp\left[-C_0 c_p^2(N,\delta)\right] + \exp(-C_1 T^{C_2})$. 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-2.2 in online theory 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 Proof of Theorem 2

For any $B > 0$,

$$\begin{aligned}\Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B\right) = &\Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B|\mathcal{A}_0\right) \Pr(\mathcal{A}_0) + \\ &\Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B|\mathcal{A}_0^c\right) \Pr(\mathcal{A}_0^c).\end{aligned}$$

Since $\Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B|\mathcal{A}_0^c\right)$ and $\Pr(\mathcal{A}_0)$ are less than or equal to one, we can further write,

$$\Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B\right) \leq \Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B|\mathcal{A}_0\right) + \Pr(\mathcal{A}_0^c).$$

By conditioning on $\mathcal{A}_0$ the dimension of vector $\hat{\boldsymbol{\gamma}}_T$ is at most equal to $k + k_T^*$ and by assumption $k_T^* = \ominus(T^d)$ where $0 \leq d < 1/2$. Therefore, by Lemma S-1.7 in online theory supplement, conditional on $\mathcal{A}_0$, $\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\|$ is $O_p\left(T^{\frac{d-1}{2}}\right)$. By Theorem 1, we also have

$\lim_{T\to\infty} \Pr(\mathcal{A}_0^c) = 0$. Hence, for any $\varepsilon > 0$, there exists $B_\varepsilon > 0$ and $T_\varepsilon > 0$ such that

$$\Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B_\varepsilon | \mathcal{A}_0\right) + \Pr(\mathcal{A}_0^c) < \varepsilon \text{ for all } T > T_\varepsilon.$$

Therefore, $\Pr\left(T^{\frac{1-d}{2}} \|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| > B_\varepsilon\right) < \varepsilon$ for all $T > T_\varepsilon$, and we conclude that

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| = O_P\left(T^{\frac{d-1}{2}}\right),$$

as required. Similar lines of arguments can be used to show that if $\mathbb{E}\left(\mathbf{x}_{\tilde{k}_T,t}\mathbf{x}'_{\tilde{k}_T,t}\right)$ is a fixed time-invariant matrix, then $\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^\diamond\| = O_P\left(T^{\frac{d-1}{2}}\right)$, which completes the proof.

A.3 Proof of Theorem 3

Let $D_T = T^{-1}\sum_{t=1}^T \hat{\eta}_t^2 - \left(\bar{\Delta}_{\beta,T} + \bar{\sigma}_{u,T}^2\right)$. For any $B > 0$,

$$\Pr\left(T^{\frac{1}{2}}|D_T| > B\right) = \Pr\left(T^{\frac{1}{2}}|D_T| > B | \mathcal{A}_0\right)\Pr(\mathcal{A}_0) + \Pr\left(T^{\frac{1}{2}}|D_T| > B | \mathcal{A}_0^c\right)\Pr(\mathcal{A}_0^c).$$

Since $\Pr\left(T^{\frac{1}{2}}|D_T| > B|\mathcal{A}_0^c\right)$ and $\Pr(\mathcal{A}_0)$ are less than or equal to one, we can further write,

$$\Pr\left(T^{\frac{1}{2}}|D_T| > B\right) \leq \Pr\left(T^{\frac{1}{2}}|D_T| > B | \mathcal{A}_0\right) + \Pr(\mathcal{A}_0^c).$$

By conditioning on $\mathcal{A}_0$, the number of selected covariates is at most equal to $k + k_T^*$ and by assumption $k_T^* = \ominus(T^d)$, where $0 \leq d < 1/2$. Therefore, by Lemma S-1.7 in online theory supplement, conditional on $\mathcal{A}_0$, D_T is $O_p\left(T^{-\frac{1}{2}}\right)$. By Theorem 1, we also have $\lim_{T\to\infty}\Pr(\mathcal{A}_0^c) = 0$. Hence, for any $\varepsilon > 0$, there exists $B_\varepsilon > 0$ and $T_\varepsilon > 0$ such that $\Pr\left(T^{\frac{1}{2}}|D_T| > B_\varepsilon|\mathcal{A}_0\right) + \Pr(\mathcal{A}_0^c) < \varepsilon$, for all $T > T_\varepsilon$. Therefore, $\Pr\left(T^{\frac{1}{2}}|D_T| > B_\varepsilon\right) < \varepsilon$ for all $T > T_\varepsilon$, and we conclude that

$$T^{-1}\sum_{t=1}^T \hat{\eta}_t^2 - \left(\bar{\Delta}_{\beta,T} + \bar{\sigma}_{u,T}^2\right) = O_p\left(T^{-\frac{1}{2}}\right),$$

as required. Following similar lines of argument we get that if $\mathbb{E}\left(\mathbf{x}_{\tilde{k}_T,t}\mathbf{x}'_{\tilde{k}_T,t}\right)$ is a fixed time-invariant matrix, then,

$$T^{-1}\sum_{t=1}^T \hat{\eta}_t^2 - \left(\bar{\Delta}_{\beta,T}^* + \bar{\sigma}_{u,T}^2\right) = O_p\left(T^{-\frac{1}{2}}\right),$$

which completes the proof.

References

Alaíz, C. M., Á. Barbero, and J. R. Dorronsoro (2013). Group fused Lasso. In V. Mladenov, P. Koprinkova-Hristova, G. Palm, A. E. P. Villa, B. Appollini, and N. Kasabov (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2013*, Berlin, Heidelberg, pp. 66–73. Springer Berlin Heidelberg.

Buhlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics 34*, 559–583.

Caner, M. and K. Knight (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*, 691–715.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*, 221–241.

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

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

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*, 157–181.

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & economic statistics 20*, 134–144.

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

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 849–911.

Fan, J. and J. Lv (2018). Sure independence screening. *Wiley StatsRef*, 1–8.

Fan, J., Y. Ma, and W. Dai (2014). Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient models. *Journal of the American Statistical Association 109*, 1270–1284.

Fan, Y., J. Lv, M. Sharifvaghefi, and Y. Uematsu (2020). Ipad: stable interpretable forecasting with knockoffs inference. *Journal of the American Statistical Association 115*, 1822–1834.

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.

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

Inoue, A., L. Jin, and B. Rossi (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics 196*, 55–67.

Kapetanios, G. and F. Zikes (2018). Time-varying Lasso. *Economics Letters 169*, 1–6.

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

Koo, B., H. M. Anderson, M. H. Seo, and W. Yao (2020). High-dimensional predictive regression in the presence of cointegration. *Journal of Econometrics, forthcoming*.

Koop, G. and S. Potter (2004). Forecasting in dynamic factor models using Bayesian model averaging. *The Econometrics Journal 7*, 550–565.

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

Lütkepohl, H. (1996). *Handbook of Matrices*. West Sussex, UK: John Wiley & Sons.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies 73*, 1057–1084.

Pesaran, M. H. and A. Pick (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics 29*, 307–318.

Pesaran, M. H., A. Pick, and M. Pranovich (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics 177*, 134–152.

Pesaran, M. H., T. Schuermann, and L. V. Smith (2009). Forecasting economic and financial variables with global VARs. *International journal of forecasting 25*, 642–675.

Pesaran, M. H. and A. Timmermann (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics 10*, 461–465.

Pesaran, M. H. and A. Timmermann (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics 137*, 134–161.

Qian, J. and L. Su (2016). Shrinkage estimation of regression models with multiple structural changes. *Econometric Theory 32*(6), 1376–1433.

Rossi, B. (2013). Advances in forecasting under instability. In *Handbook of Economic Forecasting*, Volume 2B, Chapter 21, pp. 1203–1324. Elsevier.

Sharifvaghefi, M. (2022). Variable selection in linear regressions with many highly correlated covariates. *Available at SSRN 4159979*.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*, 267–288.

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

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

Yousuf, K. and S. Ng (2021). Boosting high dimensional predictive regressions with time varying parameters. *Journal of Econometrics 224*(1), 60–87.

Zhao, P. and B. Yu (2006). On model selection consistency of Lasso. *Journal of Machine learning research 7*, 2541–2563.

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

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association 101*(476), 1418–1429.

Online Theory Supplement to "Variable Selection in High Dimensional Linear Regressions with Parameter Instability"

Alexander Chudik
Federal Reserve Bank of Dallas

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

Mahrad Sharifvaghefi
University of Pittsburgh

January 10, 2023

This online theory supplement has two sections. Section S-1 provides the main lemmas needed for the proof of Theorems 1-3 in Appendix A of the paper. Section S-2 contains the complementary lemmas needed for the proofs of the main lemmas in the previous section.

Notations: Generic finite positive constants are denoted by C_i for $i = 1, 2, \cdots$ and c. They can take different values in different instances. $\|\mathbf{A}\|_2$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_\infty$ and $\|\mathbf{A}\|_1$ denote the spectral, Frobenius, row, and column norms of matrix $\mathbf{A}$, respectively. $\lambda_i(\mathbf{A})$ denotes the i^{th} eigenvalue of a square matrix $\mathbf{A}$. $\text{tr}(\mathbf{A})$ and $\det(\mathbf{A})$ are the trace and determinant of a square matrix $\mathbf{A}$, respectively. $\|\mathbf{x}\|$ denotes the ℓ_2 norm of vector $\mathbf{x}$. 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 and n_0 such that $|f_n|/g_n \leq C_0$ for all $n > n_0$. $f_n = o(g_n)$ 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 $\sup_{n \geq n_0}(f_n/g_n) \leq C_1$. respectively. If $\{f_n\}_{n=1}^\infty$ is a sequence of random variables and $\{g_n\}_{n=1}^\infty$ is a sequence of positive real numbers, then $f_n = O_p(g_n)$, if for any $\varepsilon > 0$, there exists a positive constant B_ε and n_ε such that $\Pr(|f_n| > g_n B_\varepsilon) < \varepsilon$ for all $n > n_\varepsilon$.

S-1 Main lemmas

Lemma S-1.1 *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 $y_t = \mathbf{q}_t'\boldsymbol{\delta}_t + u_t$, where $\mathbf{q}_t = (\mathbf{z}_t', \mathbf{x}_{k,t}')'$ and $\boldsymbol{\delta}_t = (\mathbf{a}_t', \boldsymbol{\beta}_t')'$. We have,

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

■

Lemma S-1.2 *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-2.3 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-2.4, we have

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

and hence

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

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 S-1.3 *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*

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

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 $[T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{z}_t \mathbf{z}_t')^{-1}][T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{z}_t x_{it})]$. 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* $\mathbf{M}_z = \mathbf{I}_T - T^{-1}\mathbf{Z}\hat{\mathbf{\Sigma}}_{zz}^{-1}\mathbf{Z}'$ *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 pre-selected covariates m is finite. Therefore by Lemma S-2.20, 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 S-1.4 *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,* $\{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$*. 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 S-1.1, 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 S-1.2 , 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 pre-selected covariates m is finite. Therefore by Lemma S-2.20, 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 S-1.5 *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 $\mathbf{q}_{it} \equiv (\mathbf{z}_t', x_{it})'$ as*

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

where $\bar{\boldsymbol{\phi}}_{i,T} \equiv \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_{it}\mathbf{q}_{it}')\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_{it}y_t)]$ 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*

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

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

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

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

Proof. Note that $\boldsymbol{\eta}_i'\mathbf{M}_{q_i}\boldsymbol{\eta}_i = \mathbf{y}'\mathbf{M}_{q_i}\mathbf{y}$, where $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$. By Lemma S-1.1 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 S-1.2, 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 pre-selected covariates m is finite. Therefore by Lemma S-2.20, 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

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

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

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

for all $i = 1, 2, \cdots, N$. ■

Lemma S-1.6 *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 $\bar{\boldsymbol{\psi}}_{i,T} = (\psi_{1i,T}, \cdots, \psi_{mi,T})' = [T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')^{-1}][T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t x_{it})]$ *is the* $m \times 1$ *vector of projection coefficients. 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 $\bar{\boldsymbol{\psi}}_{y,T} = (\psi_{1y,T}, \cdots, \psi_{my,T})' = \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t y_t)]$. *Lastly, define the projection regression of* y_t *on* $\mathbf{q}_{it} \equiv (\mathbf{z}_t', x_{it})'$ *as*

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

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

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

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

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

In the first case where $\theta_{i,T} = \ominus(T^{1-\epsilon_i})$ for some $\epsilon_i > 1/2$, by using Lemma S-2.4 we have

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

where $d_T \to 0$ as $T \to \infty$. By using Lemma S-2.6,

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

where $\mathcal{R}_{iT} = |(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_{q_i}\boldsymbol{\eta}_i)/\sigma_{\eta_i}^2 - 1|$ 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-2.3 and S-2.4 , 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 S-1.3, 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

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

Therefore, by Lemma S-1.5, 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 S-1.4, 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-2.8,

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

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-2.5, 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 S-1.4, there exist some finite positive constants C_0 and C_1 such that,

$$\Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > |\theta_{i,T}| - \sigma_{\eta_i}\sigma_{\tilde{x}_i}[(1+d_T)T]^{1/2}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0T^{C_1}),$$

and therefore

$$\Pr\left[\mathcal{A}_{iT} < \sqrt{1+d_T}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0T^{C_1}).$$

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

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

where $\mathcal{R}_{iT} = |(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_{q_i}\boldsymbol{\eta}_i)/\sigma_{\eta_i}^2 - 1|$ 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-2.3 and S-2.4 , 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

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

So, by Lemma S-1.3, we know that there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0T^{C_1}).$$

Similarly,

$$\Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0T^{C_1}).$$

Also note that

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

Therefore, by Lemma S-1.5, 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

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

■

Lemma S-1.7 *Suppose y_t are generated by*

$$y_t = \sum_{i=1}^{k} x_{it}\beta_{it} + u_t \text{ for } t = 1, 2, \cdots, T, \tag{S.1}$$

and consider the LS estimator of the following regression augmented with the additional l_T regressors from the active set:

$$y_t = \mathbf{x}_{kt}'\boldsymbol{\phi} + \mathbf{s}_t'\boldsymbol{\delta}_T + \eta_t,$$

where $\mathbf{x}_{kt} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, is the $k \times 1$ vector of signals, $\mathbf{s}_t$ is the $l_T \times 1$ vector of additional regressors, $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_k)'$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_{l_T})'$ are the associated coefficients. The LS estimator of $\boldsymbol{\gamma}_T = (\boldsymbol{\phi}', \boldsymbol{\delta}_T')'$ is

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{y}\right), \tag{S.2}$$

where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_T)'$, $\mathbf{w}_t = (\mathbf{x}_{kt}', \mathbf{s}_t')'$ and $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$. The model error is

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{W}\hat{\boldsymbol{\gamma}}_T. \tag{S.3}$$

Suppose that $\lambda_{\min}\left[T^{-1}\mathbb{E}(\mathbf{W}'\mathbf{W})\right] > c > 0$, and $l_T = \ominus(T^d)$, where $0 \leq d < \frac{1}{2}$. Moreover suppose that Assumptions 1-4 holds. Now,

(i) If $\mathbb{E}(\beta_{it}) = \beta_i$ *for all* t*, then*

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| = O_p\left(T^{\frac{d-1}{2}}\right),$$

where $\boldsymbol{\gamma}_T^* = (\boldsymbol{\beta}', \ \mathbf{0}'_{l_T})'$ *and* $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)'$*. Under Assumption 6 we also have*

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = \bar{\sigma}^2_{u,T} + \bar{\Delta}_{\beta,T} + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{T}\right),$$

where $\bar{\sigma}^2_{u,T} = T^{-1}\sum_{t=1}^T \mathbb{E}\,(u_t^2)$*, and* $\bar{\Delta}_{\beta,T} = T^{-1}\sum_{t=1}^T tr(\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\boldsymbol{\Omega}_{\beta,t})$ *are non-negative, with* $\boldsymbol{\Sigma}_{\mathbf{x}_k,t} \equiv (\sigma_{ijt,x})$*,* $\boldsymbol{\Omega}_{\beta,t} \equiv (\sigma_{ijt,\beta})$ *for* $i,j = 1,2,\cdots,k$*, and* $\sigma_{ijt,x} = \mathbb{E}\,(x_{it}x_{jt})$*,* $\sigma_{ijt,\beta} = \mathbb{E}\,[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)]$*.*

(ii) If $\mathbb{E}\,(\mathbf{w}_t\mathbf{w}_t')$ *is time-invariant, then*

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^\diamond\| = O_p\left(T^{\frac{d-1}{2}}\right),$$

where $\boldsymbol{\gamma}_T^\diamond = (\bar{\boldsymbol{\beta}}_T', \mathbf{0}'_{l_T})'$*,* $\bar{\boldsymbol{\beta}}_T = (\bar{\beta}_{1T}, \bar{\beta}_{2T}, \cdots, \bar{\beta}_{kT})'$*, and* $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^T \mathbb{E}(\beta_{it})$*. If Assumption 6 also holds, then*

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = \bar{\sigma}^2_{u,T} + \bar{\Delta}^*_{\beta,T} + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{T}\right),$$

where $\bar{\Delta}^*_{\beta,T} = T^{-1}\sum_{t=1}^T tr\left(\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\boldsymbol{\Omega}^*_{\beta,t}\right)$ *is non-negative, with* $\boldsymbol{\Omega}^*_{\beta,t} \equiv \left(\sigma^*_{ijt,\beta}\right)$ *for* $i,j = 1,2,\cdots,k$*, and* $\sigma^*_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \bar{\beta}_{i,T})(\beta_{jt} - \bar{\beta}_{j,T})\right]$*.*

Proof. In the first scenario, where $\mathbb{E}(\beta_{it}) = \beta_i$ for all t , we can write (S.1) as

$$y_t = \sum_{i=1}^k x_{it}\beta_i + \sum_{i=1}^k x_{it}\left(\beta_{it} - \beta_i\right) + u_t = \sum_{i=1}^k x_{it}\beta_i + \sum_{i=1}^k r_{it} + u_t = \mathbf{x}'_{kt}\boldsymbol{\beta} + \mathbf{r}'_t\boldsymbol{\tau} + u_t,$$

where $r_{it} = x_{it}\left(\beta_{it} - \beta_i\right)$, $\mathbf{r}_t = (r_{1t}, r_{2t}, \cdots, r_{kt})'$, and $\boldsymbol{\tau}$ is a $k \times 1$ vector of ones. We can further write the DGP in a following matrix format,

$$\mathbf{y} = \mathbf{X}_k\boldsymbol{\beta} + \mathbf{R}\boldsymbol{\tau} + \mathbf{u}, \tag{S.4}$$

where $\mathbf{X}_k = (\mathbf{x}_{k1}, \mathbf{x}_{k2}, \cdots, \mathbf{x}_{kT})'$, $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \ \mathbf{r}_T)'$ and $\mathbf{u} = (u_1, u_2, \cdots, u_T)'$. By substituting (S.4) into (S.2), we obtain

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{X}_k\boldsymbol{\beta}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right),$$

where $\mathbf{W} = (\mathbf{X}_k, \mathbf{S})$, and $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_T)'$. Since $\boldsymbol{\gamma}_T^* = (\boldsymbol{\beta}', \mathbf{0}'_{l_T})'$, $\mathbf{X}_k\boldsymbol{\beta} = \mathbf{X}_k\boldsymbol{\beta} + \mathbf{S}\mathbf{0}_{l_T} =$

$\mathbf{W}\boldsymbol{\gamma}_T^*$, which in turn allows us to write the above result as:

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\boldsymbol{\gamma}_T^* + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right),$$

and hence

$$\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^* = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right). \tag{S.5}$$

We can further write

$$\begin{aligned}
\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^* = &\left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right]\right\} + \\
&\left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left[T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right] + \\
&\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau}\right)\right]\right\} + \\
&\left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\} + \\
&\left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left[T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right] + \\
&\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\}.
\end{aligned}$$

Hence, by the sub-additive property of norms and Lemma S-2.9, we have

$$\begin{aligned}
\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| \leq &\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\left[\left(\ \mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right]\right\| + \\
&\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right\| + \\
&\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\left[\left(\ \mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right]\right\| + \\
&\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\| + \\
&\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\| + \\
&\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\left[\left(\ \mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\ \mathbf{u}\right)\right]\right\|.
\end{aligned}$$

Since, by Assumption 3, β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of $\mathbf{w}_t$ for $t = 1, 2, \cdots, T$,

$$\begin{aligned}
T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) &= \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t r_{it})\right] = \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t x_{it}\left(\beta_{it} - \beta_i\right))\right] \\
&= \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t x_{it})\mathbb{E}\left(\beta_{it} - \beta_i\right)\right] = \mathbf{0}.
\end{aligned}$$

Also,

$$T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right) = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t u_t) = \mathbf{0}.$$

Hence,

$$\begin{aligned}\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| \leq & \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\| + \\ & \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\| + \\ & \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| + \\ & \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\|.\end{aligned}$$

Since Assumptions 1 and 2 imply that $\mathbf{W}$ and $\mathbf{u}$ satisfy condition (i) and (ii) of Lemma S-2.12, by Lemmas S-2.12 and S-2.13, we have

$$\left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

Similarly,

$$\left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{W}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{W}\right)\right]\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right),$$

and since $l_T = \ominus(T^d)$ with $0 \leq d < 1/2$, by Lemma S-2.14,

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right).$$

Now consider $\|T^{-1}\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau}\|$. Note that the row j and column i of $l_T \times p$ matrix $T^{-1}\mathbf{W}'\mathbf{R}$ is equal to $T^{-1}\sum_{t=1}^{T} w_{jt}r_{it}$. Hence the j^{th} element of $l_T \times 1$ vector $T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}$ is equal $T^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T} w_{jt}r_{it}$. In other words, $T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau} = T^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T}\mathbf{w}_t r_{it}$. Therefore, (re-

calling that $r_{it} = x_{it}(\beta_{it} - \beta_i)$)

$$\begin{aligned}\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\|^2 =& \left\|T^{-1}\sum_{i=1}^{k}\sum_{t=1}^{T}(\mathbf{w}_t r_{it})\right\|^2 \leq \sum_{i=1}^{k}\left\|T^{-1}\sum_{t=1}^{T}\mathbf{w}_t x_{it}(\beta_{it}-\beta_i)\right\|^2 \\ =& T^{-2}\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbf{w}_t'\mathbf{w}_{t'}x_{it}x_{it'}(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i) \\ =& T^{-2}\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{\ell=1}^{k+l_T} w_{\ell t}w_{\ell t'}x_{it}x_{it'}(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i).\end{aligned}$$

Since, by Assumption 1, β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of $\mathbf{w}_t$ for $t = 1, 2, \cdots, T$, we can further write,

$$\begin{aligned}\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\|^2 &\leq T^{-2}\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{\ell=1}^{k+\ell_T}\mathbb{E}(w_{\ell t}w_{\ell t'}x_{it}x_{it'})\,\mathbb{E}[(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i)] \\ &\leq T^{-2}\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{\ell=1}^{k+\ell_T}|\mathbb{E}(w_{\ell t}w_{\ell t'}x_{it}x_{it'})| \times |\mathbb{E}[(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i)]| \\ &\leq T^{-2}(k+\ell_T)\,sup_{i,\ell,t,t'}|\mathbb{E}(w_{\ell t}w_{\ell t'}x_{it}x_{it'})|\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}|\mathbb{E}[(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i)]|\end{aligned}$$

Since $\mathbf{W}$ satisfy condition (i) of Lemma S-2.12, we have $\sup_{i,\ell,t,t'}|\mathbb{E}(w_{\ell t}w_{\ell t'}x_{it}x_{it'})| < C < \infty$. Also, note that for any $t' < t$,

$$\mathbb{E}[(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i)] = \mathbb{E}[(\beta_{it'}-\beta_i)\,\mathbb{E}(\beta_{it}-\beta_i|\mathcal{F}_{t-1})],$$

and by Assumption 1, $\mathbb{E}(\beta_{it}-\beta_i|\mathcal{F}_{t-1}) = 0$. Therefore,

$$\begin{aligned}\sum_{t=1}^{T}\sum_{t'=1}^{T}|\mathbb{E}[(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i)]| &= \sum_{t=1}^{T}\left|\mathbb{E}\left[(\beta_{it}-\beta_i)^2\right]\right| + 2\sum_{t=2}^{T}\sum_{t'=1}^{t}|\mathbb{E}[(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i)]| \\ &= \sum_{t=1}^{T}\left|\mathbb{E}\left[(\beta_{it}-\beta_i)^2\right]\right| = O(T).\end{aligned}$$

Since, by Assumption 3, k is also a finite fixed integer, we conclude that

$$\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\|^2 = O\left(\frac{l_T}{T}\right),$$

and hence, by Lemma S-2.13,

$$\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

So, we can conclude that

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| = O_p\left(\sqrt{\frac{l_T}{T}}\right),$$

as required.

In the next step, consider the mean square error of the model, $T^{-1}\hat{\boldsymbol{\eta}}_T'\hat{\boldsymbol{\eta}}_T$. By substituting y from (S.4) into equation (S.3) for the model error, we have

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{W}\hat{\boldsymbol{\gamma}}_T = \mathbf{X}_k\boldsymbol{\beta} + \mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \mathbf{W}\hat{\boldsymbol{\gamma}}_T.$$

Since $\mathbf{X}_k\boldsymbol{\beta} = \mathbf{W}\boldsymbol{\gamma}_T^*$, where $\boldsymbol{\gamma}_T^* = (\boldsymbol{\beta}', \mathbf{0}_{l_T}')'$, we can further write,

$$\hat{\boldsymbol{\eta}} = \mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \ \mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right).$$

Therefore,

$$\begin{aligned}
T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} &= T^{-1}\left[\mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \mathbf{W}\left(\ \hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right]'\left[\mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right] \\
&= T^{-1}\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right)'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right) + T^{-1}\left[\mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right]'\left[\mathbf{W}\left(\ \hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right] - \\
&\quad 2T^{-1}\left[\mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right]'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right) \\
&= T^{-1}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} + \left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right) - \\
&\quad 2\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)'\left[T^{-1}\left(\mathbf{W}'\ \mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right].
\end{aligned}$$

By substituting for $\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*$ from (S.5), we get

$$\begin{aligned}
T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} &= T^{-1}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\ \boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} + \\
&\qquad \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right] - \\
&\qquad 2\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right] \\
&= T^{-1}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} - \\
&\qquad \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right].
\end{aligned}$$

we can further write

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + T^{-1}\left\{\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\ \boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] + \left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right]\right\} + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} - \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right] - \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right].$$

Therefore,

$$\begin{aligned} T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} - T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) \leq\ & T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\ \ \mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] + T^{-1}\left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right] + \\ & 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} + \left\|T^{-1}\mathbf{W}'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right)\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 + \\ & \left\|T^{-1}\mathbf{W}'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right)\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F. \end{aligned} \tag{S.6}$$

First, consider $T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right]$. Note that

$$\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} = \boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}_t'\right)\boldsymbol{\tau} = \sum_{t=1}^{T}\left(\ \boldsymbol{\tau}'\mathbf{r}_t\right)\left(\mathbf{r}_t'\boldsymbol{\tau}\right) = \sum_{t=1}^{T}\left(\sum_{i=1}^{k} r_{it}\right)\left(\sum_{j=1}^{k} r_{jt}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{t=1}^{T} r_{it}r_{jt}.$$

Recalling that $r_{it} = x_{it}(\beta_{it} - \beta_i)$, and hence,

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\tilde{r}_{ij,t}\right),$$

where

$$\tilde{r}_{ij,t} = r_{it}r_{jt} - \mathbb{E}(r_{it}r_{jt})$$

Now consider $\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}\tilde{r}_{ij,t}\right)^2$ and note that

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}\tilde{r}_{ij,t}\right)^2 = T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(\tilde{r}_{ij,t}\tilde{r}_{ij,t'}\right).$$

By Assumption 6, $T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(\tilde{r}_{ij,t}\tilde{r}_{ij,t'}\right) = O\left(T^{-1}\right)$, and hence, by Lemma S-2.13, it follows that

$$\left|T^{-1}\sum_{t=1}^{T}\tilde{r}_{ij,t}\right| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Since by Assumption 3, k is a finite fixed integer, we can further conclude that

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}-\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right]=\sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\tilde{r}_{ij,t}\right)=O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.7}$$

Now, consider, $T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u}$. Note that

$$T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u}=T^{-1}\boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{r}_t u_t\right)=T^{-1}\sum_{t=1}^{T}\boldsymbol{\tau}'\mathbf{r}_t u_t=T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{k}r_{it}u_t=\sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T}r_{it}u_t\right).$$

We have

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}r_{it}u_t\right)^2=T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(r_{it}^2u_t^2\right)+2T^{-2}\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left(r_{it}r_{it'}u_tu_{t'}\right).$$

Since $r_{it}=x_{it}(\beta_{it}-\beta_i)$, and β_{it} for $i=1,2,\cdots,k$ are distributed independently of x_{js}, $j=1,2,\cdots,N$, and u_s for all t and s, we can further write for any $t'<t$

$$\begin{aligned}\mathbb{E}\left(r_{it}r_{it'}u_tu_{t'}\right)&=\mathbb{E}\left(x_{it}u_tx_{it'}u_{t'}\right)\mathbb{E}\left[(\beta_{it}-\beta_i)(\beta_{it'}-\beta_i)\right]\\&=\mathbb{E}\left(x_{it}u_tx_{it'}u_{t'}\right)\mathbb{E}\left\{(\beta_{it'}-\beta_i)\mathbb{E}\left[(\beta_{it}-\beta_i)|\mathcal{F}_{t-1}\right]\right\}.\end{aligned}$$

But, by Assumption 1, $\mathbb{E}\left[(\beta_{it}-\beta_i)|\mathcal{F}_{t-1}\right]=0$ and thus $\mathbb{E}\left(r_{it}r_{it'}u_tu_{t'}\right)=0$ for any $t'<t$. Therefore,

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}r_{it}u_t\right)^2=T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(r_{it}^2u_t^2\right)=O\left(\frac{1}{T}\right).$$

Hence, by Lemma S-2.13, $\left|T^{-1}\sum_{t=1}^{T}r_{it}u_t\right|=O_p\left(\frac{1}{\sqrt{T}}\right)$. Since, by Assumption 3, k is a finite fixed integer, we conclude that

$$T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u}=\sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T}r_{it}u_t\right)=O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.8}$$

By substituting (S.7) and (S.8) into (S.6), and noting that $\left\|T^{-1}\mathbf{W}'\left(\mathbf{R}\boldsymbol{\tau}+\mathbf{u}\right)\right\|^2=O_p(l_T/T)$, $\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}-\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F=O_p(l_T/\sqrt{T})$, and $T^{-1}\left[\mathbf{u}'\mathbf{u}-\mathbb{E}\left(\ \mathbf{u}'\mathbf{u}\right)\right]=O_p(1/\sqrt{T})$, we conclude that

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}}=\sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma_{ijt,\beta}\right)+\bar{\sigma}_{u,T}^2+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{l_T}{T}\right),$$

where $\sigma_{ijt,x}=\mathbb{E}\left(x_{it}x_{jt}\right)$, $\sigma_{ijt,\beta}=\mathbb{E}\left[(\beta_{it}-\beta_i)(\beta_{jt}-\beta_j)\right]$, and $\bar{\sigma}_{u,T}^2=T^{-1}\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)$. We further

have

$$\bar{\Delta}_{\beta,T} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma_{ijt,\beta}\right) = T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ijt,x}\sigma_{ijt,\beta}\right) = \frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}\left(\mathbf{\Omega}_{\beta,t}\mathbf{\Sigma}_{\mathbf{x}_k,t}\right),$$

where $\mathbf{\Omega}_{\beta,t} \equiv (\sigma_{ijt,\beta})$ and $\mathbf{\Sigma}_{\mathbf{x}_k,t} \equiv (\sigma_{ijt,x})$ for $i,j = 1,2,\cdots,k$. By result 9(b) on page 44 of Lütkepohl (1996), we can further write

$$\mathrm{tr}\left(\mathbf{\Omega}_{\beta,t}\mathbf{\Sigma}_{\mathbf{x}_k,t}\right) \geq k\left[\det\left(\mathbf{\Omega}_{\beta,t}\right)\right]^{1/k}\left[\det\left(\mathbf{\Sigma}_{\mathbf{x}_k,t}\right)\right]^{1/k}.$$

But k is a finite fixed integer. Furthermore, $\det\left(\mathbf{\Omega}_{\beta,t}\right) \geq 0$ and $\det\left(\mathbf{\Sigma}_{\mathbf{x}_k,t}\right) > 0$, since $\mathbf{\Omega}_{\beta,t}$ and $\mathbf{\Sigma}_{\mathbf{x}_k,t}$ are positive semi-definite and positive definite matrices, respectively. So, we can conclude that $\bar{\Delta}_{\beta,T} \geq 0$ as required.

In the second scenario, where $\mathbb{E}\left(\mathbf{w}_t\mathbf{w}_t'\right)$ is time-invariant, we can write (S.1) as

$$y_t = \sum_{i=1}^{k}x_{it}\bar{\beta}_{iT} + \sum_{i=1}^{k}x_{it}\left(\beta_{it} - \bar{\beta}_{iT}\right) + u_t = \sum_{i=1}^{k}x_{it}\bar{\beta}_{iT} + \sum_{i=1}^{k}h_{it} + u_t = \mathbf{x}_{kt}'\bar{\boldsymbol{\beta}} + \mathbf{h}_t'\boldsymbol{\tau} + u_t,$$

where $h_{it} = x_{it}\left(\beta_{it} - \bar{\beta}_{iT}\right)$, and $\mathbf{h}_t = (h_{1t}, h_{2t}, \cdots, h_{kt})'$. We can further write the DGP in (S.1) in a following matrix format,

$$\mathbf{y} = \mathbf{X}_k\bar{\boldsymbol{\beta}} + \mathbf{H}\boldsymbol{\tau} + \mathbf{u},$$

where $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_T)'$. Now, by using the similar lines of arguments as in the first scenario, we obtain

$$\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^{\diamond} = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right).$$

Notice that

$$\begin{aligned}
T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right) &= \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t h_{it})\right] = \sum_{i=1}^{k}\left\{T^{-1}\sum_{t=1}^{T}\mathbb{E}[\mathbf{w}_t x_{it}(\beta_{it} - \bar{\beta}_{iT})]\right\} \\
&= \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t x_{it})\mathbb{E}(\beta_{it} - \bar{\beta}_{iT})\right] \\
&= \sum_{i=1}^{k}\left[\mathbb{E}(\mathbf{w}_t x_{it})T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{it} - \bar{\beta}_{iT})\right] = \mathbf{0}.
\end{aligned}$$

Hence, we can further use the similar lines of arguments as in the first scenario and conclude

that

$$\begin{aligned}\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^{\diamond}\| \leq & \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\| + \\ & \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\| + \\ & \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| + \\ & \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\|.\end{aligned}$$

We know that

$$\left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right),$$

and

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right).$$

Now consider $\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\|$. By using the similar lines of arguments as in the first scenario, we have

$$\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|^2 \leq T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\sum_{t'=1}^{T} w_{\ell t}w_{\ell t'}x_{it}x_{it'}\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right).$$

Since, by Assumption 3, β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of $\mathbf{w}_t$ for $t = 1, 2, \cdots, T$, we can further write,

$$\begin{aligned}\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|^2 &\leq T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right] \\ &= T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\mathbb{E}\left(w_{\ell t}^2x_{it}^2\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)^2\right] + \\ &\quad T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right].\end{aligned}$$

Since, by Assumption 1, $\mathbb{E}\left[w_{\ell t}w_{\ell' t} - \mathbb{E}(w_{\ell t}w_{\ell' t})|\mathcal{F}_{t-1}\right] = 0$ for all ℓ, ℓ' and $t = 1, 2, \cdots, T$, we have for any $t' \neq t$

$$\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right) = \mathbb{E}\left(w_{\ell t}x_{it}\right)\mathbb{E}\left(w_{\ell t'}x_{it'}\right).$$

Therefore,

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$=\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}x_{it}\right)\mathbb{E}\left(w_{\ell t'}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right].$$

Since $\mathbb{E}\left(\mathbf{w}_t\mathbf{w}_t'\right)$ is time-invariant, we can further write

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$=\mathbb{E}\left(w_{\ell t}x_{it}\right)^2\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right].$$

Note that, by Assumption 1, for any $t'\neq t$, $\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]=\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right]$. Therefore

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$=\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\sum_{t=1}^{T}\sum_{t'\neq t}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right].$$

We can further write,

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$=\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\left\{\sum_{t=1}^{T}\sum_{t'=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right]-\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2\right\}$$
$$=\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\left\{\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\right\}\left\{\sum_{t'=1}^{T}\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right]\right\}-$$
$$\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2.$$

But, $\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]=0$, and therefore,

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]=-\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2.$$

So,

$$\begin{aligned}
&\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|^2 \\
&\quad\leq T^{-2}\sum_{i=1}^{p}\sum_{\ell=1}^{p+l_T}\sum_{t=1}^{T}\left\{\mathbb{E}\left(w_{\ell t}^2 x_{it}^2\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)^2\right]-\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2\right\} \\
&\quad= O\left(\frac{l_T}{T}\right),
\end{aligned}$$

and hence, by Lemma S-2.13,

$$\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

So, we conclude that

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^{\diamond}\| = O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

Lastly, consider the model mean square error for the second scenario. Following the same lines of argument as in the first scenario, we can write,

$$\begin{aligned}
&T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} - T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}+\mathbf{u}'\mathbf{u}\right)\leq \\
&\quad T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\ \ \mathbf{H}\boldsymbol{\tau}-\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right]+T^{-1}\left[\mathbf{u}'\mathbf{u}-\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right]+ \\
&\quad 2T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u}+\left\|T^{-1}\mathbf{W}'\left(\mathbf{H}\boldsymbol{\tau}+\mathbf{u}\right)\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2+ \\
&\quad \left\|T^{-1}\mathbf{W}'\left(\mathbf{H}\boldsymbol{\tau}+\mathbf{u}\right)\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}-\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F.
\end{aligned} \tag{S.9}$$

First, consider $T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}-\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right]$. Note that

$$\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} = \boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{h}_t\mathbf{h}_t'\right)\boldsymbol{\tau} = \sum_{t=1}^{T}\left(\boldsymbol{\tau}'\mathbf{r}_t\right)\left(\mathbf{r}_t'\boldsymbol{\tau}\right) = \sum_{t=1}^{T}\left(\sum_{i=1}^{k}h_{it}\right)\left(\sum_{j=1}^{k}h_{jt}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{t=1}^{T}h_{it}h_{jt}.$$

Recalling that $h_{it} = x_{it}(\beta_{it}-\bar{\beta}_{iT})$, and hence,

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}-\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\tilde{h}_{ij,t}\right),$$

where

$$\tilde{h}_{ij,t} = h_{it}h_{jt}-\mathbb{E}(h_{it}h_{jt}).$$

Now consider $\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}\tilde{h}_{ij,t}\right)^2$ and note that

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}\tilde{h}_{ij,t}\right)^2 = T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(\tilde{h}_{ij,t}\tilde{h}_{ij,t'}\right).$$

By Assumption 6, $T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(\tilde{h}_{ij,t}\tilde{h}_{ij,t'}\right) = O\left(T^{-1}\right)$, and hence, by Lemma S-2.13, it follows that

$$\left|T^{-1}\sum_{t=1}^{T}\tilde{h}_{ij,t}\right| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Since by Assumption 3, k is a finite fixed integer, we can further conclude that

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\tilde{h}_{ij,t}\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.10}$$

Now, consider, $T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u}$. Note that

$$T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u} = T^{-1}\boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{h}_t u_t\right) = T^{-1}\sum_{t=1}^{T}\boldsymbol{\tau}'\mathbf{h}_t u_t = T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{k}h_{it}u_t = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T}h_{it}u_t\right).$$

We have

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}h_{it}u_t\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left[\left(h_{it}u_t\right)^2\right] + T^{-2}\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(h_{it}h_{it'}u_t u_{t'}\right).$$

Since $h_{it} = x_{it}(\beta_{it} - \bar{\beta}_{iT})$, and β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of x_{js}, $j = 1, 2, \cdots, N$, and u_s for all t and s, we can further write for any $t' \neq t$

$$\mathbb{E}\left(h_{it}h_{it'}u_t u_{t'}\right) = \mathbb{E}\left(x_{it}u_t x_{it'}u_{t'}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{it'} - \bar{\beta}_{iT})\right].$$

But, by Assumption 1, $\mathbb{E}\left[x_{it}u_t - \mathbb{E}(x_{it}u_t)|\mathcal{F}_{t-1}\right] = 0$ and we also have $\mathbb{E}(x_{it}u_t) = 0$ for $i = 1, 2, \cdots, k$ and thus for any $t' \neq t$ we have

$$\mathbb{E}\left(x_{it}u_t x_{it'}u_{t'}\right) = \mathbb{E}\left(x_{it}u_t\right)\mathbb{E}\left(x_{it'}u_{t'}\right) = 0.$$

Therefore,

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}h_{it}u_t\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left[\left(h_{it}u_t\right)^2\right] = O\left(\frac{1}{T}\right).$$

Hence, by Lemma S-2.13, $\left|T^{-1}\sum_{t=1}^{T} h_{it}u_t\right| = O_p\left(\frac{1}{\sqrt{T}}\right)$. Since, by Assumption 3, k is a finite fixed integer, we conclude that

$$T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u} = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T} h_{it}u_t\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.11}$$

By substituting (S.10) and (S.11) into (S.9), and noting that $\left\|T^{-1}\mathbf{W}'\left(\mathbf{H}\boldsymbol{\tau}+\mathbf{u}\right)\right\|^2 = O_p(l_T/T)$, $\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p(l_T/\sqrt{T})$, and $T^{-1}\left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\ \mathbf{u}'\mathbf{u}\right)\right] = O_p(1/\sqrt{T})$, we conclude that

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma^*_{ijt,\beta}\right) + \bar{\sigma}^2_{u,T} + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{T}\right),$$

where $\sigma^*_{ijt,\beta} = \mathbb{E}\left[(\beta_{it}-\bar{\beta}_{i,T})(\beta_{jt}-\bar{\beta}_{j,T})\right]$, $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{it})$, and $\bar{\sigma}^2_{u,T} = T^{-1}\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)$. We further have

$$\bar{\Delta}^*_{\beta,T} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma^*_{ijt,\beta}\right) = T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ijt,x}\sigma^*_{ijt,\beta}\right) = \frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}\left(\boldsymbol{\Omega}^*_{\beta,t}\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\right),$$

where $\boldsymbol{\Omega}^*_{\beta,t} \equiv \left(\sigma^*_{ijt,\beta}\right)$ and $\boldsymbol{\Sigma}_{\mathbf{x}_k,t} \equiv \left(\sigma_{ijt,x}\right)$ for $i,j = 1,2,\cdots,k$. By result 9(b) on page 44 of Lütkepohl (1996), we can further write

$$\mathrm{tr}\left(\boldsymbol{\Omega}^*_{\beta,t}\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\right) \geq k\left[\det\left(\boldsymbol{\Omega}^*_{\beta,t}\right)\right]^{1/k}\left[\det\left(\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\right)\right]^{1/k}.$$

But k is a finite fixed integer. Furthermore, $\det\left(\boldsymbol{\Omega}^*_{\beta,t}\right) \geq 0$ and $\det\left(\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\right) > 0$, since $\boldsymbol{\Omega}^*_{\beta,t}$ and $\boldsymbol{\Sigma}_{\mathbf{x}_k,t}$ are positive semi-definite and positive definite matrices, respectively. So, we can conclude that $\bar{\Delta}^*_{\beta,T} \geq 0$ as required. ■

S-2 Supplementary lemmas

Lemma S-2.1 *Let z_t be a martingale difference process with respect to $\mathcal{F}^z_{t-1} = \sigma(z_{t-1}, z_{t-2},\cdots)$, and suppose that there exist some finite positive constants C_0 and C_1, and $s > 0$ such that*

$$\sup_t \Pr(|z_t| > \alpha) \leq C_0\exp(-C_1\alpha^s), \quad \textit{for all } \alpha > 0.$$

Let also $\sigma_{zt}^2 = \mathbb{E}(z_t^2|\mathcal{F}_{t-1}^z)$ *and* $\bar{\sigma}_{z,T}^2 = T^{-1}\sum_{t=1}^T \sigma_{zt}^2$. *Suppose that* $\zeta_T = \ominus(T^\lambda)$, *for some* $0 < \lambda \leq (s+1)/(s+2)$. *Then for any* π *in the range* $0 < \pi < 1$, *we have,*

$$\Pr\left(|\textstyle\sum_{t=1}^T z_t| > \zeta_T\right) \leq \exp\left[\frac{-(1-\pi)^2\zeta_T^2}{2T\bar{\sigma}_{z,T}^2}\right].$$

If $\lambda > (s+1)/(s+2)$, *then for some finite positive constant* C_2,

$$\Pr\left(|\textstyle\sum_{t=1}^T z_t| > \zeta_T\right) \leq \exp\left(-C_2\zeta_T^{s/(s+1)}\right).$$

Proof. The results follow from Lemma A3 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S-2.2 *Let*

$$c_p(n,\delta) = \Phi^{-1}\left(1 - \frac{p}{2f(n,\delta)}\right), \tag{S.12}$$

where $\Phi^{-1}(.)$ *is the inverse of standard normal distribution function,* p $(0 < p < 1)$ *is the nominal size of a test, and* $f(n,\delta) = cn^\delta$ *for some positive constants* δ *and* c. *Moreover, let* $a > 0$ *and* $0 < b < 1$. *Then (I)* $c_p(n,\delta) = O\left[\sqrt{\delta\ln(n)}\right]$ *and (II)* $n^a \exp\left[-bc_p^2(n,\delta)\right] = \ominus(n^{a-2b\delta})$.

Proof. The results follow from Lemma 3 of Bailey et al. (2019) Supplementary Appendix A. ■

Lemma S-2.3 *Let* x_i, *for* $i = 1, 2, \cdots, n$, *be random variables. Then for any constants* π_i, *for* $i = 1, 2, \cdots, n$, *satisfying* $0 < \pi_i < 1$ *and* $\sum_{i=1}^n \pi_i = 1$, *we have*

$$\Pr(\textstyle\sum_{i=1}^n |x_i| > C_0) \leq \sum_{i=1}^n \Pr(|x_i| > \pi_i C_0),$$

where C_0 *is a finite positive constant.*

Proof. The result follows from Lemma A11 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S-2.4 *Let* x, y *and* z *be random variables. Then for any finite positive constants* C_0, C_1, *and* C_2, *we have*

$$\Pr(|x| \times |y| > C_0) \leq \Pr(|x| > C_0/C_1) + \Pr(|y| > C_1),$$

and

$$\Pr(|x| \times |y| \times |z| > C_0) \leq \Pr(|x| > C_0/(C_1C_2)) + \Pr(|y| > C_1) + \Pr(|z| > C_2).$$

Proof. The results follow from Lemma A11 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S-2.5 *Let x be a random variable. Then for some finite constants B, and C, with $|B| \geq C > 0$, we have*

$$\Pr(|x + B| \leq C) \leq \Pr(|x| > |B| - C).$$

Proof. The results follow from Lemma A12 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S-2.6 *Let x_T to be a random variable. Then for a deterministic sequence, $\alpha_T > 0$, with $\alpha_T \to 0$ as $T \to \infty$, there exists $T_0 > 0$ such that for all $T > T_0$ we have*

$$\Pr\left(\left|\frac{1}{\sqrt{x_T}} - 1\right| > \alpha_T\right) \leq \Pr(|x_T - 1| < \alpha_T).$$

Proof. The results follow from Lemma A13 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S-2.7 *Consider random variables x_t and z_t with the exponentially bounded probability tail distributions such that*

$$\sup_t \Pr(|x_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^{s_x}), \text{ for all } \alpha > 0,$$
$$\sup_t \Pr(|z_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^{s_z}), \text{ for all } \alpha > 0,$$

where C_0, and C_1 are some finite positive constants, $s_x > 0$, and $s_z > 0$. Then

$$\sup_t \Pr(|x_t z_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^{s/2}), \text{ for all } \alpha > 0,$$

where $s = \min\{s_x, s_z\}$.

Proof. By using Lemma S-2.4, for all $\alpha > 0$,

$$\Pr(|x_t z_t| > \alpha) \leq \Pr(|x_t| > \alpha^{1/2}) + \Pr(|z_t| > \alpha^{1/2})$$

So,

$$\begin{aligned}\sup_t \Pr(|x_t z_t| > \alpha) &\leq \sup_t \Pr(|x_t| > \alpha^{1/2}) + \sup_t \Pr(|z_t| > \alpha^{1/2}) \\ &\leq C_0 \exp(-C_1 \alpha^{s_x/2}) + C_0 \exp(-C_1 \alpha^{s_z/2}) \\ &\leq C_0 \exp(-C_1 \alpha^{s/2})\end{aligned}$$

where $s = \min\{s_x, s_z\}$. ■

Lemma S-2.8 *Let x, y and z be random variables. Then for some finite positive constants C_0, and C_1, we have*

$$\Pr(|x| \times |y| < C_0) \leq \Pr(|x| < C_0/C_1) + \Pr(|y| < C_1),$$

Proof. Define events $\mathfrak{A} = \{|x| \times |y| < C_0\}$, $\mathfrak{B} = \{|x| < C_0/C_1\}$ and $\mathfrak{C} = \{|y| < C_1\}$. Then $\mathfrak{A} \in \mathfrak{B} \cup \mathfrak{C}$. Therefore, $\Pr(\mathfrak{A}) \leq \Pr(\mathfrak{B} \cup \mathfrak{C})$. But $\Pr(\mathfrak{B} \cup \mathfrak{C}) \leq \Pr(\mathfrak{B}) + \Pr(\mathfrak{C})$ and hence $\Pr(\mathfrak{A}) \leq \Pr(\mathfrak{B}) + \Pr(\mathfrak{C})$. ■

Lemma S-2.9 *Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times p$ and $p \times m$ matrices respectively, then*

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_2, \text{ and } \|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_F.$$

Proof. $\|\mathbf{AB}\|_F^2 = \text{tr}(\mathbf{ABB'A'}) = \text{tr}[\mathbf{A}(\mathbf{BB'})\mathbf{A'}]$, and by result (12) of Lütkepohl (1996, p.44),

$$\text{tr}\left[\mathbf{A}(\mathbf{BB'})\mathbf{A'}\right] \leq \lambda_{\max}(\mathbf{BB'})\text{tr}(\mathbf{AA'}) = \|\mathbf{A}\|_F^2 \|\ \mathbf{B}\|_2^2,$$

where $\lambda_{\max}(\mathbf{BB'})$ is the largest eigenvalue of $\mathbf{BB'}$. Therefore, $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_2$, as required. Similarly,

$$\|\mathbf{AB}\|_F^2 = \text{tr}(\mathbf{B'A'AB}) = \text{tr}[\mathbf{B'}(\mathbf{A'A})\mathbf{B}] \leq \lambda_{\max}(\mathbf{A'A})\text{tr}(\mathbf{B'B}) = \|\mathbf{A}\|_2^2 \|\mathbf{B}\|_F^2,$$

and hence

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_F.$$

■

Lemma S-2.10 *Let $\mathbf{A} = (a_{ij})_{n \times m}$ where $\sup_{ij} |a_{ij}| < C < \infty$, then*

$$\|\mathbf{A}\|_2 = O\left(\sqrt{nm}\right).$$

Proof. This result follows, since $\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_\infty \|\mathbf{A}\|_1}$, $\|\mathbf{A}\|_\infty = O(m)$ and $\|\mathbf{A}\|_1 = O(n)$. ■

Lemma S-2.11 *Consider two $N \times N$ nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$ such that*

$$\|\mathbf{B}^{-1}\|_2 \|\mathbf{A} - \mathbf{B}\|_F < 1.$$

Then

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F \leq \frac{\|\mathbf{B}^{-1}\|_2^2 \|\mathbf{A} - \mathbf{B}\|_F}{1 - \|\mathbf{B}^{-1}\|_2 \|\mathbf{A} - \mathbf{B}\|_F}.$$

Proof. By Lemma S-2.9,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F = \|\mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}\|_F \leq \|\mathbf{A}^{-1}\|_2 \|\mathbf{B} - \mathbf{A}\|_F \|\mathbf{B}^{-1}\|_2$$

Note that

$$\begin{aligned}\|\mathbf{A}^{-1}\|_2 &= \|\mathbf{A}^{-1} - \mathbf{B}^{-1} + \mathbf{B}^{-1}\|_2 \leq \|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_2 + \|\mathbf{B}^{-1}\|_2 \\ &\leq \|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F + \|\ \mathbf{B}^{-1}\|_2,\end{aligned}$$

and therefore,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F \leq (\|\mathbf{A}^{-1} - \quad \mathbf{B}^{-1}\|_F + \|\mathbf{B}^{-1}\|_2)\|\mathbf{B} - \mathbf{A}\|_F \|\mathbf{B}^{-1}\|_2.$$

Hence,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F (1 - \|\mathbf{B}^{-1}\|_2 \|\mathbf{B} - \mathbf{A}\|_F) \leq \|\mathbf{B}^{-1}\|_2^2 \|\mathbf{B} - \mathbf{A}\|_F.$$

Since $\|\mathbf{B}^{-1}\|_2 \|\mathbf{B} - \mathbf{A}\|_F < 1$, we can further write,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F \leq \frac{\|\mathbf{B}^{-1}\|_2^2 \|\mathbf{A} - \mathbf{B}\|_F}{1 - \|\mathbf{B}^{-1}\|_2 \|\mathbf{A} - \mathbf{B}\|_F}.$$

■

Lemma S-2.12 *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be* $T \times N_x$ *and* $T \times N_y$ *matrices of observations on random variables* x_{it} *and* y_{jt}*, for* $i = 1, 2, \cdots, N_x$*,* $j = 1, 2, \cdots, N_y$ *and* $t = 1, 2, \cdots, T$*, respectively. Denote*

$$w_{ij,t} = x_{it} y_{jt} - \mathbb{E}(x_{it} y_{jt}), \text{ for all } i, j \text{ and } t.$$

Suppose that

(i) $\sup_{i,t} \mathbb{E}\,|x_{it}|^4 < C$, $\sup_{j,t} \mathbb{E}\,|y_{jt}|^4 < C$, *and*

(ii) $\sup_{i,j} \left[\sum_{t=1}^{T} \sum_{t'=1}^{T} \mathbb{E}(w_{ij,t} w_{ij,t'})\right] = O(T)$.

Then,

$$\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2 = O\left(\frac{N_x N_y}{T}\right).$$

Proof. The results follow from Lemma A18 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S-2.13 *Let* $\mathbf{X} = (x_{ij})_{T\times N_x}$ *and* $\mathbf{Y} = (y_{ij})_{T\times N_y}$ *be matrices of random variables, respectively. Suppose that,*

$$\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2 = O(a_T),$$

where $a_T > 0$*. Then*

$$\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F = O_p(\sqrt{a_T}).$$

Proof. For any $B > 0$, by the Markov's inequality

$$\Pr\left(\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F > B\sqrt{a_T}\right) \leq \frac{\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2}{a_T B^2}$$

Since $\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2 = O(a_T)$, there exist C and T_0 such that for all $T > T_0$

$$\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2 \leq Ca_T.$$

Hence, for any $\varepsilon > 0$, there exist $B_\varepsilon = \sqrt{\frac{C}{\varepsilon}}$ and $T_\varepsilon = T_0$, such that for all $T > T_\varepsilon$

$$\Pr\left(\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F > B_\varepsilon\sqrt{a_T}\right) \leq \varepsilon.$$

Therefore,

$$\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F = O_p\left(\sqrt{a_T}\right).$$

■

Lemma S-2.14 *Let* $\mathbf{\Sigma}_T$ *be a positive definite matrix and* $\hat{\mathbf{\Sigma}}_T$ *be its corresponding estimator. Suppose that* $\lambda_{\min}\left(\mathbf{\Sigma}_T\right) > c > 0$*, and*

$$\mathbb{E}\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F^2 = O(a_T)$$

where $a_T > 0$*, and* $a_T = o(1)$*. Then*

$$\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F = O_p(\sqrt{a_T})$$

Proof. Let $\mathcal{A}_T = \left\{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F < 1\right\}$, $\mathcal{B}_T = \left\{\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F > B\sqrt{a_T}\right\}$ and $\mathcal{D}_T = \left\{\frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F}{\left(1-\left\|\mathbf{\Sigma}_T^{-1}\right\|_2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F\right)} > B\sqrt{a_T}\right\}$ where $B > 0$ is an arbitrary constant. If $\mathcal{A}_T$

holds, by Lemma S-2.11,

$$\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F \leq \frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 \left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F}{1 - \left\|\mathbf{\Sigma}_T^{-1}\right\|_2 \left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F}.$$

Hence $\mathcal{B}_T \cap \mathcal{A}_T \subseteq \mathcal{D}_T$. Therefore

$$\begin{aligned}\Pr(\mathcal{B}_T \cap \mathcal{A}_T) &\leq \Pr\left(\frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 \left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F}{\left(1 - \left\|\ \mathbf{\Sigma}_T^{-1}\right\|_2 \left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F\right)} > B\sqrt{a_T}\right) \\ &= \Pr\left(\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F > \frac{B\sqrt{a_T}}{\left\|\ \mathbf{\Sigma}_T^{-1}\right\|_2 \left(\left\|\mathbf{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)}\right)\end{aligned}$$

By the Markov's inequality, we can further conclude that

$$\Pr(\mathcal{B}_T \cap \mathcal{A}_T) \leq \frac{\mathbb{E}\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F^2}{a_T} \times \frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 \left(\left\|\mathbf{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2}.$$

Since by assumption $\mathbb{E}\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F^2 = O(a_T)$, there exist C and $T_0 > 0$ such that for all $T > T_0$,

$$\mathbb{E}\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F^2 \leq C a_T.$$

Therefore, for all $T > T_0$,

$$\Pr(\mathcal{B}_T \cap \mathcal{A}_T) \leq \frac{C\left\|\ \mathbf{\Sigma}_T^{-1}\right\|_2^2 \left(\left\|\mathbf{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2}.$$

Moreover,

$$\Pr(\mathcal{A}_T^c) = \Pr\left(\left\|\mathbf{\Sigma}_T^{-1}\right\|_2 \left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F \geq 1\right) = \Pr\left(\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F \geq \frac{1}{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2}\right).$$

By the Markov's inequality, we can further write

$$\Pr(\mathcal{A}_T^c) \leq \left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 \times \mathbb{E}\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F^2,$$

and hence, for all $T > T_0$,

$$\Pr(\mathcal{A}_T^c) \leq C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 a_T.$$

Note that

$$\Pr(\mathcal{B}_T) = \Pr\left(\mathcal{B}_T \cap \mathcal{A}_T\right) + \Pr(\mathcal{B}_T|\mathcal{A}_T^c)\Pr(\mathcal{A}_T^c),$$

and since $\Pr(\mathcal{B}_T \cap \mathcal{A}_T) \leq \Pr(\mathcal{D}_T)$ and $\Pr(\mathcal{B}_T|\mathcal{A}_T^c) \leq 1$, we have

$$\Pr(\mathcal{B}_T) \leq \Pr(\mathcal{B}_T \cap \mathcal{A}_T) + \Pr(\mathcal{A}_T^c).$$

Therefore, for all $T > T_0$,

$$\Pr\left(\left\|\hat{\boldsymbol{\Sigma}}_T^{-1} - \boldsymbol{\Sigma}_T^{-1}\right\|_F > B\sqrt{a_T}\right) \leq \frac{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2\left(\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2} + C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 a_T.$$

Now, for a given $\varepsilon > 0$, we are interested to find $B_\varepsilon > 0$ and $T_\varepsilon > 0$ such that for all $T > T_\varepsilon$,

$$\Pr\left(\left\|\hat{\boldsymbol{\Sigma}}_T^{-1} - \boldsymbol{\Sigma}_T^{-1}\right\|_F > B_\varepsilon\sqrt{a_T}\right) \leq \varepsilon.$$

To do so, we first find a value of B such that

$$\frac{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2\left(\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2} + C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 a_T = \varepsilon.$$

By multiplying both sides of the above equality by B^2 and bringing all the equations to the left hand side we have

$$\left(\varepsilon - 2C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 a_T\right)B^2 - 2C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^3\sqrt{a_T}B - C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^4 = 0.$$

By solving the above quadratic equation of B we have

$$\begin{aligned} B^* &= \frac{2C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^3\sqrt{a_T} \pm \sqrt{4C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^4\varepsilon - 4C^2\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^6 a_T}}{2\left(\varepsilon - 2C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 a_T\right)} \\ &= \frac{\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2\left(\sqrt{a_T} \pm \sqrt{\frac{\varepsilon}{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2} - a_T}\right)}{\frac{\varepsilon}{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2} - 2a_T}. \end{aligned}$$

Notice that $a_T \to 0$ as $T \to \infty$, therefore for large enough T^* we have both $\frac{\varepsilon}{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2} - 2a_T$ and $\frac{\varepsilon}{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2} - a_T$ being greater than zero for all $T > T^*$. Now, by setting $T_\varepsilon = \max\{T^*, T_0\}$ and

$$B_\varepsilon = \frac{\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2\left(\sqrt{a_T} + \sqrt{\frac{\varepsilon}{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2} - a_T}\right)}{\frac{\varepsilon}{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2} - 2a_T} > 0,$$

we achieve our goal that for all $T > T_\varepsilon$,

$$\Pr\left(\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F > B_\varepsilon \sqrt{a_T}\right) \leq \varepsilon.$$

■

Remark 7 *By using Lemma S-2.11 we achieve the probability convergence order for* $\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F$ *that is sharper than the one shown in the proof Lemma A21 of Chudik et al. (2018) (see equations (B.103) and (B.105) of Chudik et al. (2018) Online Theory Supplement).*

Lemma S-2.15 *Let* z_{ij} *be a random variable for* $i = 1, 2, \cdots, N$, *and* $j = 1, 2, \cdots, N$. *Then, for any* $d_T > 0$,

$$\Pr(N^{-2} \textstyle\sum_{i=1}^N \sum_{j=1}^N |z_{ij}| > d_T) \leq N^2 \sup_{i,j} \Pr(|z_{ij}| > d_T).$$

Proof. We know that $N^{-2} \sum_{i=1}^N \sum_{j=1}^N |z_{ij}| \leq \sup_{i,j} |z_{ij}|$. Therefore,

$$\begin{aligned}
\Pr(N^{-2} \textstyle\sum_{i=1}^N \sum_{j=1}^N |z_{ij}| > d_T) &\leq \Pr(\sup_{i,j} |z_{ij}| > d_T) \\
&\leq \Pr[\cup_{i=1}^N \cup_{j=1}^N (|z_{ij}| > d_T)] \leq \textstyle\sum_{i=1}^N \sum_{j=1}^N \Pr(|z_{ij}| > d_T) \\
&\leq N^2 \sup_{i,j} \Pr(|z_{ij}| > d_T).
\end{aligned}$$

■

Lemma S-2.16 *Let* $\hat{\mathbf{\Sigma}}$ *be an estimator of a* $N \times N$ *symmetric invertible matrix* $\mathbf{\Sigma}$. *Suppose that there exits a finite positive constant* C_0, *such that*

$$\sup_{i,j} \Pr(|\hat{\sigma}_{ij} - \sigma_{ij}| > d_T) \leq \exp(-C_0 T d_T^2), \text{ for any } d_T > 0,$$

where σ_{ij} *and* $\hat{\sigma}_{ij}$ *are the elements of* $\mathbf{\Sigma}$ *and* $\hat{\mathbf{\Sigma}}$ *respectively. Then, for any* $b_T > 0$,

$$\begin{aligned}
\Pr(\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F > b_T) \leq N^2 \exp\left[-C_0 \frac{T b_T^2}{N^2 \|\mathbf{\Sigma}^{-1}\|_2^2 (\|\mathbf{\Sigma}^{-1}\|_2 + b_T)^2}\right] + \\
N^2 \exp\left(-C_0 \frac{T}{N^2 \|\mathbf{\Sigma}^{-1}\|_2^2}\right).
\end{aligned}$$

Proof. Let $\mathcal{A}_N = \{\|\mathbf{\Sigma}^{-1}\|_2 \| \hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F \leq 1\}$ and $\mathcal{B}_N = \{\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F > b_T\}$, and note that by Lemma S-2.11 if $\mathcal{A}_N$ holds we have

$$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F \leq \frac{\|\mathbf{\Sigma}^{-1}\|_2^2 \|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F}{1 - \|\mathbf{\Sigma}^{-1}\|_2 \|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F}.$$

Hence

$$\begin{aligned}\Pr(\mathcal{B}_N|\mathcal{A}_N) &\leq \Pr\left(\frac{\|\boldsymbol{\Sigma}^{-1}\|_2^2\|\hat{\boldsymbol{\Sigma}}-\ \boldsymbol{\Sigma}\|_F}{1-\|\boldsymbol{\Sigma}^{-1}\|_2\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_F} > b_T\right)\\ &= \Pr\left[\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_F > \frac{b_T}{\|\boldsymbol{\Sigma}^{-1}\|_2(\|\boldsymbol{\Sigma}^{-1}\|_2+b_T)}\right].\end{aligned}$$

Note that $\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_F = \left(\sum_{i=1}^N\sum_{j=1}^N(\hat{\sigma}_{ij}-\sigma_{ij})^2\right)^{1/2}$. Therefore,

$$\begin{aligned}\Pr(\mathcal{B}_N|\mathcal{A}_N) &\leq \Pr\left[\left(\sum_{i=1}^N\sum_{j=1}^N(\hat{\sigma}_{ij}-\sigma_{ij})^2\right)^{1/2} > \frac{b_T}{\|\boldsymbol{\Sigma}^{-1}\|_2(\|\boldsymbol{\Sigma}^{-1}\|_2+b_T)}\right]\\ &= \Pr\left[\sum_{i=1}^N\sum_{j=1}^N(\hat{\sigma}_{ij}-\sigma_{ij})^2 > \frac{b_T^2}{\|\boldsymbol{\Sigma}^{-1}\|_2^2(\|\boldsymbol{\Sigma}^{-1}\|_2+b_T)^2}\right].\end{aligned}$$

By Lemma S-2.15, we can further write,

$$\begin{aligned}\Pr(\mathcal{B}_N|\mathcal{A}_N) &\leq N^2\sup_{i,j}\Pr\left[(\hat{\sigma}_{ij}-\sigma_{ij})^2 > \frac{b_T^2}{N^2\|\boldsymbol{\Sigma}^{-1}\|_2^2(\|\boldsymbol{\Sigma}^{-1}\|_2+b_T)^2}\right]\\ &= N^2\sup_{i,j}\Pr\left[|\hat{\sigma}_{ij}-\sigma_{ij}| > \frac{b_T}{N\|\boldsymbol{\Sigma}^{-1}\|_2(\|\boldsymbol{\Sigma}^{-1}\|_2+b_T)}\right]\\ &\leq N^2\exp\left[-C_0\frac{Tb_T^2}{N^2\|\boldsymbol{\Sigma}^{-1}\|_2^2(\|\boldsymbol{\Sigma}^{-1}\|_2+b_T)^2}\right]\end{aligned}$$

Furthermore,

$$\begin{aligned}\Pr(\mathcal{A}_N^c) &= \Pr(\|\boldsymbol{\Sigma}^{-1}\|_2\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_F > 1)\\ &= \Pr(\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_F > \|\boldsymbol{\Sigma}^{-1}\|_2^{-1})\\ &= \Pr\left[\left(\sum_{i=1}^N\sum_{j=1}^N(\hat{\sigma}_{ij}-\sigma_{ij})^2\right)^{1/2} > \|\boldsymbol{\Sigma}^{-1}\|_2^{-1}\right]\\ &= \Pr\left[\sum_{i=1}^N\sum_{j=1}^N(\hat{\sigma}_{ij}-\sigma_{ij})^2 > \|\boldsymbol{\Sigma}^{-1}\|_2^{-2}\right]\\ &\leq N^2\sup_{i,j}\Pr\left[(\hat{\sigma}_{ij}-\sigma_{ij})^2 > \frac{1}{N^2\|\boldsymbol{\Sigma}^{-1}\|_2^2}\right]\\ &\leq N^2\sup_{i,j}\Pr\left[|\hat{\sigma}_{ij}-\sigma_{ij}| > \frac{1}{N\|\boldsymbol{\Sigma}^{-1}\|_2}\right]\\ &\leq N^2\exp\left[-C_0\frac{T}{N^2\|\ \boldsymbol{\Sigma}^{-1}\|_2^2}\right].\end{aligned}$$

Note that

$$\Pr(\mathcal{B}_N) = \Pr(\mathcal{B}_N|\mathcal{A}_N)\Pr(\mathcal{A}_N) + \Pr(\mathcal{B}_N|\mathcal{A}_N^c)\Pr(\mathcal{A}_N^c),$$

and since $\Pr(\mathcal{A}_N)$ and $\Pr(\mathcal{B}_N|\mathcal{A}_N^c)$ are less than equal to one, we have

$$\Pr(\mathcal{B}_N) \leq \Pr(\mathcal{B}_N|\mathcal{A}_N) + \Pr(\mathcal{A}_N^c).$$

Therefore,

$$\Pr(\mathcal{B}_{NT}) \leq N^2 \exp\left[-C_0 \frac{Tb_T^2}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2(\|\mathbf{\Sigma}^{-1}\|_2 + b_T)^2}\right] + N^2 \exp\left[-C_0 \frac{T}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2}\right].$$

■

Lemma S-2.17 *Let $\hat{\mathbf{\Sigma}}$ be an estimator of a $N \times N$ symmetric invertible matrix $\mathbf{\Sigma}$. Suppose that there exits a finite positive constant C_0, such that*

$$\sup_{i,j} \Pr(|\hat{\sigma}_{ij} - \sigma_{ij}| > d_T) \leq \exp\left[-C_0(Td_T)^{s/s+2}\right], \text{ for any } d_T > 0,$$

where σ_{ij} and $\hat{\sigma}_{ij}$ are the elements of $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ respectively. Then, for any $b_T > 0$,

$$\Pr(\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F > b_T) \leq N^2 \exp\left[-C_0 \frac{(Tb_T)^{s/s+2}}{N^{s/s+2}\|\mathbf{\Sigma}^{-1}\|_2^{s/s+2}(\|\ \mathbf{\Sigma}^{-1}\|_2 + b_T)^{s/s+2}}\right] + N^2 \exp\left(-C_0 \frac{T^{s/s+2}}{N^{s/s+2}\|\mathbf{\Sigma}^{-1}\|_2^{s/s+2}}\right).$$

Proof. The proof is similar to the proof of Lemma S-2.16. ■

Lemma S-2.18 *Let $\{x_{it}\}_{t=1}^T$ for $i = 1, 2, \cdots, N$ and $\{z_{jt}\}_{t=1}^T$ for $j = 1, 2, \cdots, m$ be time-series processes. Also let $\mathcal{F}_{it}^x = \sigma(x_{it}, x_{i,t-1}, \cdots)$ for $i = 1, 2, \cdots, N$, $\mathcal{F}_{jt}^z = \sigma(z_{jt}, z_{j,t-1}, \cdots)$ for $j = 1, 2, \cdots, m$, $\mathcal{F}_t^x = \cup_{i=1}^N \mathcal{F}_{it}^x$, $\mathcal{F}_t^z = \cup_{j=1}^m \mathcal{F}_{jt}^z$, and $\mathcal{F}_t = \mathcal{F}_t^x \cup \mathcal{F}_t^z$. Define the projection regression of x_{it} on $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{m,t})'$ as*

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

where $\bar{\boldsymbol{\psi}}_{i,T} = (\psi_{1i,T}, \psi_{2i,T}, \cdots, \psi_{mi,T})'$ is the $m \times 1$ vector of projection coefficients which is equal to $\left[T^{-1}\sum_{t=1}^T \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right)\right]^{-1}[T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t x_{it})]$. Suppose, $\mathbb{E}[x_{it}x_{i't} - \mathbb{E}(x_{it}x_{i't})|\mathcal{F}_{t-1}] = 0$ for all $i, i' = 1, 2, \cdots, N$, $\mathbb{E}[z_{jt}z_{j't} - \mathbb{E}(z_{jt}z_{j't})|\mathcal{F}_{t-1}] = 0$ for all $j, j' = 1, 2, \cdots, m$, and $\mathbb{E}[z_{jt}x_{it} - \mathbb{E}(z_{jt}x_{it})|\mathcal{F}_{t-1}] = 0$ for all $j = 1, 2, \cdots, m$ and for all $i = 1, 2, \cdots, N$. Then

$$\mathbb{E}\left[\tilde{x}_{it}\tilde{x}_{i't} - \mathbb{E}(\tilde{x}_{it}\tilde{x}_{i't})|\mathcal{F}_{t-1}\right] = 0,$$

for all $j, j' = 1, 2, \cdots, N$,

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

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

$$T^{-1} \sum_{t=1}^{T} \mathbb{E}(\tilde{x}_{it} z_{jt}) = 0,$$

for all $i = 1, 2, \cdots, N$ *and* $j = 1, 2, \cdots, m$.

Proof.

$$\begin{aligned}
\mathbb{E}(\tilde{x}_{it}\tilde{x}_{i't}|\mathcal{F}_{t-1}) &= \mathbb{E}(x_{it}x_{i't}|\mathcal{F}_{t-1}) - \mathbb{E}(x_{it}\mathbf{z}_t'|\mathcal{F}_{t-1})\bar{\boldsymbol{\psi}}_{i',T} - \\
&\quad \mathbb{E}(x_{i't}\mathbf{z}_t'|\mathcal{F}_{t-1})\bar{\boldsymbol{\psi}}_{i,T} + \bar{\boldsymbol{\psi}}_{i,T}' \, \mathbb{E}(\mathbf{z}_t\mathbf{z}_t'|\mathcal{F}_{t-1})\bar{\boldsymbol{\psi}}_{i',T} \\
&= \mathbb{E}(x_{it}x_{i't}) - \mathbb{E}(x_{it}\mathbf{z}_t')\bar{\boldsymbol{\psi}}_{i',T} - \mathbb{E}(x_{i't}\mathbf{z}_t')\bar{\boldsymbol{\psi}}_{i,T} + \\
&\quad \bar{\boldsymbol{\psi}}_{i,T}'\mathbb{E}(\mathbf{z}_t \, \mathbf{z}_t')\bar{\boldsymbol{\psi}}_{i',T} = \mathbb{E}(\tilde{x}_{it}\tilde{x}_{i't}).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\tilde{x}_{it} z_{jt}|\mathcal{F}_{t-1}) &= \mathbb{E}(x_{it} z_{jt}|\mathcal{F}_{t-1}) - \mathbb{E}(\mathbf{z}_t' z_{jt}|\mathcal{F}_{t-1})\bar{\boldsymbol{\psi}}_{i,T} \\
&= \mathbb{E}(x_{it} z_{jt}) - \mathbb{E}(\mathbf{z}_t' z_{jt})\bar{\boldsymbol{\psi}}_{i,T} = \mathbb{E}(\tilde{x}_{it} z_{it}).
\end{aligned}$$

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

■

Lemma S-2.19 *Let* $\{x_{it}\}_{t=1}^{T}$ *for* $i = 1, 2, \cdots, N$ *and* $\{z_{jt}\}_{t=1}^{T}$ *for* $j = 1, 2, \cdots, m$ *be time-series processes. Define the projection regression of* x_{it} *on* $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{m,t})'$ *as*

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

where $\bar{\boldsymbol{\psi}}_{i,T} = (\psi_{1i,T}, \psi_{2i,T}, \cdots, \psi_{mi,T})'$ *is the* $m \times 1$ *vector of projection coefficients which is equal to* $\left[T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right)\right]^{-1} [T^{-1} \sum_{t=1}^{T} \mathbb{E}(\mathbf{z}_t x_{it})]$. *Suppose that only a finite number of elements in* $\bar{\boldsymbol{\psi}}_{i,T}$ *is different from zero for all* $i = 1, 2, \cdots, N$ *and there exist sufficiently large positive constants* C_0 *and* C_1, *and* $s > 0$ *such that*

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

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

Then, there exist sufficiently large positive constants C_0 *and* C_1, *and* $s > 0$ *such that*

$$\sup_{i,t} \Pr(|\tilde{x}_{it}| > \alpha) \leq C_0 \exp(-C_1\alpha^s), \text{ for all } \alpha > 0.$$

Proof. Without loss of generality assume that the first finite ℓ elements of $\psi_{i,T}$ are different from zero and write

$$x_{it} = \sum_{j=1}^{\ell} \psi_{ji,T} z_{jt} + \tilde{x}_{it}.$$

Now, note that

$$\Pr(|\tilde{x}_{it}| > \alpha) \leq \Pr\left(|x_{it}| + \sum_{j=1}^{\ell} |\psi_{ji,T} z_{jt}| > \alpha\right),$$

and hence by Lemma S-2.3, for any $0 < \pi_j < 1$, $j = 1, 2, \cdots, \ell + 1$ we have,

$$\begin{aligned}
\Pr(|\tilde{x}_{it}| > \alpha) &\leq \sum_{j=1}^{\ell} \Pr\left(|\psi_{ji,T} z_{jt}| > \pi_j \alpha\right) + \Pr\left(|x_{it}| > \pi_{\ell+1}\alpha\right) \\
&= \sum_{j=1}^{\ell} \Pr\left(|z_{jt}| > |\psi_{ji,T}|^{-1} \pi_j \alpha\right) + \Pr\left(|x_{it}| > \pi_{\ell+1}\alpha\right) \\
&\leq \ell \sup_{j,t} \Pr\left(|z_{jt}| > |\psi_T^*|^{-1} \pi^* \alpha\right) + \sup_{i,t} \Pr\left(|x_{it}| > \pi^* \alpha\right),
\end{aligned}$$

where $\psi_T^* = sup_{i,j}\{\psi_{ji,T}\}$ and $\pi^* = \inf_{j \in 1,2,\cdots,\ell+1}\{\pi_j\}$. Therefore, by the exponential decaying probability tail assumptions for x_{it} and z_{jt} we have

$$\Pr(|\tilde{x}_{it}| > \alpha) \leq \ell C_0 \exp(-C_1 \alpha^s) + C_0 \exp(-C_1 \alpha^s),$$

and hence there exist sufficiently large positive constants C_0 and C_1, and $s > 0$ such that

$$\sup_{i,t} \Pr(|\tilde{x}_{it}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s), \ \textit{for all } \alpha > 0.$$

■

Lemma S-2.20 *Let $\{x_{it}\}_{t=1}^T$ for $i = 1, 2, \cdots, N$ and $\{z_{\ell t}\}_{t=1}^T$ for $\ell = 1, 2, \cdots, m$ be time-series processes and $m = \ominus(T^d)$. Also let $\mathcal{F}_{it}^x = \sigma(x_{it}, x_{i,t-1}, \cdots)$ for $i = 1, 2, \cdots, N$, $\mathcal{F}_{\ell t}^z = \sigma(z_{\ell t}, z_{\ell,t-1}, \cdots)$ for $\ell = 1, 2, \cdots, m$, $\mathcal{F}_t^x = \cup_{i=1}^N \mathcal{F}_{it}^x$, $\mathcal{F}_t^z = \cup_{\ell=1}^m \mathcal{F}_{\ell t}^z$, and $\mathcal{F}_t = \mathcal{F}_t^x \cup \mathcal{F}_t^z$. Define the projection regression of x_{it} on $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{m,t})'$ as*

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

where $\bar{\boldsymbol{\psi}}_{i,T} = (\psi_{1i,T}, \psi_{2i,T}, \cdots, \psi_{mi,T})'$ is the $m \times 1$ vector of projection coefficients which is equal to $\left[T^{-1}\sum_{t=1}^T \mathbb{E}\left(\mathbf{z}_t \mathbf{z}_t'\right)\right]^{-1} [T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t x_{it})]$. Suppose, $\mathbb{E}[x_{it}x_{jt} - \mathbb{E}(x_{it}x_{jt})|\mathcal{F}_{t-1}] = 0$ for all $i, j = 1, 2, \cdots, N$, $\mathbb{E}[z_{\ell t} z_{\ell' t} - \mathbb{E}(z_{\ell t} z_{\ell t})|\mathcal{F}_{t-1}] = 0$ for all $\ell, \ell' = 1, 2, \cdots, m$, and $\mathbb{E}[z_{\ell t} x_{it} - \mathbb{E}(z_{\ell t} x_{it})|\mathcal{F}_{t-1}] = 0$ for all $\ell = 1, 2, \cdots, m$ and for all $i = 1, 2, \cdots, N$. Additionally, assume that only a finite number of elements in $\bar{\boldsymbol{\psi}}_{i,T}$ is different from zero for all $i = 1, 2, \cdots, N$ and there exist sufficiently large positive constants C_0 and C_1, and $s > 0$ such that

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

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

Then, there exist some finite positive constants C_0, C_1 *and* C_2 *such that if* $d < \lambda \leq (s+2)/(s+4)$,

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

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

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

for all $i, j = 1, 2, \cdots, N$, *where* $\tilde{\boldsymbol{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, *and* $\mathbf{M}_z = \mathbf{I} - T^{-1}\ \mathbf{Z}\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\mathbf{Z}'$ *with* $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ *and* $\hat{\boldsymbol{\Sigma}}_{zz} = T^{-1} \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}_t')$.

Proof.

$$\begin{aligned}
&\Pr\left[|\mathbf{x}_i' \mathbf{M}_z \mathbf{x}_j - \mathbb{E}(\tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_j)| > \zeta_T\right] = \Pr\left[|\tilde{\boldsymbol{x}}_i' \mathbf{M}_z \tilde{\boldsymbol{x}}_j - \mathbb{E}(\tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_j)| > \zeta_T\right] \\
&\quad = \Pr\left[|\tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_j - \mathbb{E}(\tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_j) - T^{-1} \tilde{\boldsymbol{x}}_i' \mathbf{Z} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}' \tilde{\boldsymbol{x}}_j - T^{-1} \tilde{\boldsymbol{x}}_i' \mathbf{Z} (\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \ \boldsymbol{\Sigma}_{zz}^{-1}) \mathbf{Z}' \tilde{\boldsymbol{x}}_j| > \zeta_T\right],
\end{aligned}$$

where $\boldsymbol{\Sigma}_{zz} = \mathbb{E}[T^{-1} \sum_{t=1}^T (\mathbf{z}_t \mathbf{z}_t')]$. By Lemma S-2.3, we can further write

$$\begin{aligned}
&\Pr\left[|\mathbf{x}_i' \mathbf{M}_z \mathbf{x}_j - \mathbb{E}(\tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_j)| > \zeta_T\right] \\
&\quad \leq \Pr\left[|\tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_j - \mathbb{E}(\tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_j)| > \pi_1 \zeta_T\right] + \Pr(|T^{-1} \tilde{\boldsymbol{x}}_i' \mathbf{Z} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}' \tilde{\boldsymbol{x}}_j| > \pi_2 \zeta_T) + \\
&\qquad \Pr\left[|T^{-1} \tilde{\boldsymbol{x}}_i' \mathbf{Z} (\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1})\ \mathbf{Z}' \tilde{\boldsymbol{x}}_j|) > \pi_3 \zeta_T\right],
\end{aligned}$$

where $0 < \pi_i < 1$ and $\sum_{i=1}^3 \pi_i = 1$. By Lemma S-2.9,

$$\Pr(|T^{-1} \tilde{\boldsymbol{x}}_i' \mathbf{Z} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}' \tilde{\boldsymbol{x}}_j| > \pi_2 \zeta_T) \leq \Pr(\|\tilde{\boldsymbol{x}}_i' \mathbf{Z}\|_F \|\boldsymbol{\Sigma}_{zz}^{-1}\|_2 \|\mathbf{Z}' \tilde{\boldsymbol{x}}_j\|_F > \pi_2 \zeta_T T),$$

and again by Lemma S-2.4, we have

$$\begin{aligned}
&\Pr(|T^{-1} \tilde{\boldsymbol{x}}_i' \mathbf{Z} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}' \tilde{\boldsymbol{x}}_j| > \pi_2 \zeta_T) \\
&\quad \leq \Pr(\|\tilde{\boldsymbol{x}}_i' \mathbf{Z}\|_F > \|\boldsymbol{\Sigma}_{zz}^{-1}\|_2^{-1/2} \pi_2^{1/2} \zeta_T^{1/2} T^{1/2}) + \Pr(\|\mathbf{Z}' \tilde{\boldsymbol{x}}_j\|_F > \|\boldsymbol{\Sigma}_{zz}^{-1}\|_2^{-1/2} \pi_2^{1/2} \zeta_T^{1/2} T^{1/2}).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
&\Pr(|T^{-1} \tilde{\boldsymbol{x}}_i' \mathbf{Z} (\ \hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}) \mathbf{Z}' \tilde{\boldsymbol{x}}_j| > \pi_3 \zeta_T) \\
&\quad \leq \Pr(\|\tilde{\boldsymbol{x}}_i' \mathbf{Z}\|_F \|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F \|\mathbf{Z}' \tilde{\boldsymbol{x}}_j\|_F > \pi_3 \zeta_T T) \\
&\quad \leq \Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \ \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1} \zeta_T) + \Pr(\|\tilde{\boldsymbol{x}}_i' \mathbf{Z}\|_F > \pi_3^{1/2} \delta_T^{1/2} T^{1/2}) \\
&\qquad + \Pr(\|\mathbf{Z}' \tilde{\boldsymbol{x}}_j\|_F > \pi_3^{1/2} \delta_T^{1/2} T^{1/2}),
\end{aligned}$$

where $\delta_T = \ominus(T^\alpha)$ with $0 < \alpha < \lambda$.

Note that $\Pr(\|\mathbf{Z}'\tilde{\boldsymbol{x}}_i\|_F > c) = \Pr(\|\mathbf{Z}'\tilde{\boldsymbol{x}}_i\|_F^2 > c^2) = \Pr[\sum_{\ell=1}^m (\sum_{t=1}^T \tilde{x}_{it} z_{\ell t})^2 > c^2]$, where c is a positive constant. So, by Lemma S-2.3, we have

$$\Pr(\|\mathbf{Z}'\tilde{\boldsymbol{x}}_i\|_F > c) \leq \sum_{\ell=1}^m \Pr[(\sum_{t=1}^T \tilde{x}_{it} z_{\ell t})^2 > m^{-1}c^2].$$

Hence, $\Pr(\|\mathbf{Z}'\tilde{\boldsymbol{x}}_i\|_F > c) \leq \sum_{\ell=1}^m \Pr(|\sum_{t=1}^T \tilde{x}_{it} z_{\ell t}| > m^{-1/2}c)$. Also, by Lemma S-2.18 we have $\sum_{t=1}^T \mathbb{E}(\tilde{x}_{it} z_{\ell t}) = 0$ and hence we can further write

$$\Pr(\|\mathbf{Z}'\tilde{\boldsymbol{x}}_i\|_F > c) \leq \sum_{\ell=1}^m \Pr\{|\sum_{t=1}^T [\tilde{x}_{it} z_{\ell t} - \mathbb{E}(\tilde{x}_{it} z_{\ell t})]| > m^{-1/2}c\}.$$

Note that $\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2$ is equal to the largest eigenvalue of $\boldsymbol{\Sigma}_{zz}^{-1}$ and it is a finite positive constant. So, there exists a positive constant $C > 0$ such that,

$$\begin{aligned}
&\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{x}}_j)| > \zeta_T) \\
&\quad \leq \Pr\{|\textstyle\sum_{t=1}^T [\tilde{x}_{it}\tilde{x}_{jt} - \mathbb{E}(\tilde{x}_{it}\tilde{x}_{jt})]| > CT^\lambda\} + \\
&\quad\quad \textstyle\sum_{\ell=1}^m \Pr\{|\sum_{t=1}^T [\tilde{x}_{it} z_{\ell t} - \mathbb{E}(\tilde{x}_{it} z_{\ell t}]| > CT^{1/2+\lambda/2-d/2}\} + \\
&\quad\quad \textstyle\sum_{\ell=1}^m \Pr\{|\sum_{t=1}^T [\tilde{x}_{jt} z_{\ell t} - \mathbb{E}(\tilde{x}_{jt} z_{\ell t}]| > CT^{1/2+\lambda/2-d/2}\} + \\
&\quad\quad \textstyle\sum_{\ell=1}^m \Pr\{|\sum_{t=1}^T [\tilde{x}_{it} z_{\ell t} - \mathbb{E}(\tilde{x}_{it} z_{\ell t}]| > CT^{1/2+\alpha/2-d/2}\} + \\
&\quad\quad \textstyle\sum_{\ell=1}^m \Pr\{|\sum_{t=1}^T [\tilde{x}_{jt} z_{\ell t} - \mathbb{E}(\tilde{x}_{jt} z_{\ell t}]| > CT^{1/2+\alpha/2-d/2}\} + \\
&\quad\quad \Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T).
\end{aligned}$$

Let

$$\kappa_{T,i}(h,d) = \sum_{\ell=1}^m \Pr\{|\sum_{t=1}^T [\tilde{x}_{it} z_{\ell t} - \mathbb{E}(\tilde{x}_{it} z_{\ell t}]| > CT^{1/2+\kappa/2-d/2}\}, \text{ for } h = \lambda, \alpha,$$

and $i = 1, 2, ..., N$. By Lemmas S-2.7, S-2.18, and S-2.19, we have $\tilde{x}_{it}\tilde{x}_{jt} - \mathbb{E}(\tilde{x}_{it}\tilde{x}_{jt})$ and $\tilde{x}_{it} z_{\ell t} - \mathbb{E}(\tilde{x}_{it} z_{\ell t})$ are martingale difference processes with exponentially bounded probability tail, $\frac{s}{2}$. So, depending on the value of exponentially bounded probability tail parameter, from Lemma S-2.1, we know that either

$$\kappa_{T,i}(h,d) \leq m \exp\left[-\ominus\left(T^{h-d}\right)\right],$$

or

$$\kappa_{T,i}(h,d) \leq m \exp\left[-\ominus\left(T^{s(1/2+h/2-d/2)/(s+2)}\right)\right],$$

for $h = \lambda, \alpha$. Also, depending on the value of exponentially bounded probability tail param-

eter, from Lemmas S-2.16 and S-2.17 we have,

$$\Pr(\|\hat{\mathbf{\Sigma}}_{zz}^{-1} - \mathbf{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \leq m^2 \exp\left[-C_0 \frac{T\delta_T^{-2}\zeta_T^2}{m^2\|\mathbf{\Sigma}_{zz}^{-1}\|_2^2(\|\mathbf{\Sigma}_{zz}^{-1}\|_2 + \delta_T^{-1}\zeta_T)^2}\right] + m^2 \exp\left(-C_0 \frac{T}{m^2\|\mathbf{\Sigma}_{zz}^{-1}\|_2^2}\right),$$

or

$$\Pr(\|\hat{\mathbf{\Sigma}}_{zz}^{-1} - \mathbf{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \leq m^2 \exp\left[-C_0 \frac{(T\delta_T^{-1}\zeta_T)^{s/s+2}}{m^{s/s+2}\| \mathbf{\Sigma}_{zz}^{-1}\|_2^{s/s+2}(\|\mathbf{\Sigma}_{zz}^{-1}\|_2 + \delta_T^{-1}\zeta_T)^{s/s+2}}\right] + m^2 \exp\left(-C_0 \frac{T^{s/s+2}}{m^{s/s+2}\|\mathbf{\Sigma}_{zz}^{-1}\|_2^{s/s+2}}\right).$$

Therefore,

$$\begin{aligned}
&\Pr(\|\hat{\mathbf{\Sigma}}_{zz}^{-1} - \mathbf{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \\
&\quad \leq m \exp[- \ominus (T^{\max\{1-2d+2(\lambda-\alpha), 1-2d+\lambda-\alpha, 1-2d\}})] + \\
&\quad\quad m \exp[- \ominus (T^{1-2d})],
\end{aligned}$$

or,

$$\begin{aligned}
&\Pr(\|\hat{\mathbf{\Sigma}}_{zz}^{-1} - \mathbf{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \\
&\quad \leq m \exp[- \ominus (T^{s(\max\{1-d+\lambda-\alpha, 1-d\})/(s+2)})] + \\
&\quad\quad m \exp[- \ominus (T^{s(1-d)/(s+2)})].
\end{aligned}$$

Setting $d < 1/2$, $\alpha = 1/2$, and $\lambda > d$, we have all the terms going to zero as $T \to \infty$ and there exist some finite positive constants C_1 and C_2 such that

$$\kappa_{T,i}(\lambda, d) \leq \exp\left(-C_1 T^{C_2}\right), \kappa_{T,i}(\alpha, d) \leq \exp\left(-C_1 T^{C_2}\right),$$

and

$$\Pr(\|\hat{\mathbf{\Sigma}}_{zz}^{-1} - \mathbf{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \leq \exp(-C_1 T^{C_2}).$$

Hence, if $d < \lambda \leq (s+2)/(s+4)$,

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

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

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

where C_0, C_1 and C_2 are some finite positive constants. ■

References

Bailey, N., Pesaran, M. H., and Smith, L. V. (2019). A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics*, 208(2): 507-534. https://doi.org/10.1016/j.jeconom.2018.10.006

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

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1-22. https://doi.org/10.18637/jss.v033.i01

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons, West Sussex, UK. ISBN-10: 9780471970156

Online Empirical Supplement to "Variable Selection in High Dimensional Linear Regressions with Parameter Instability"

Alexander Chudik
Federal Reserve Bank of Dallas

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

Mahrad Sharifvaghefi
University of Pittsburgh

January 10, 2023

This online empirical supplement has three sections. Section S-1 provides the full list and description of technical indicators considered in the stock market application. Section S-2 provides the list of variables in the conditioning and active sets in the application on forecasting output growth rates across 33 countries. Last section focuses on the third application, forecasting Euro Area quarterly output growth using the European Central Bank (ECB) survey of professional forecasters. The section starts with description of the data and then discusses the results.

S-1 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.[9] The group membership of stock i is denoted by the set

[9]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.

$\mathfrak{g}_i$, 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 $(\bar{r}^g_{it})$: $\bar{r}^g_{it} = |\mathfrak{g}_i|^{-1} \sum_{j \in \mathfrak{g}_i} r_{jt}$.

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

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

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

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

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).

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

5. Price Gap $(gp_{it}(s))$: $gp_{it}(s) = 100\left[p_{it} - \text{MA}(p_{it}, s)\right]$.

6. Realized Volatility (RV_{it}): $RV_{it} = \sqrt{\sum_{\tau=1}^{D^i_t} \left(R_{it}(\tau) - \bar{R}_{it}\right)^2}$, where

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

7. Group Realized Volatility (RV^g_{it}): $RV^g_{it} = \sqrt{|\mathfrak{g}|^{-1} \sum_{i \in \mathfrak{g}} RV^2_{it}}$.

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

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

9. Realized Volatility Gap $(RVG_{it}(s))$: $RVG_{it}(s) = RV_{it} - \text{MA}(RV_{it}, s)$

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

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

11. Relative Strength Indicator (RSI^s_{it}): 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

$$RS^s_{it} = -\frac{\mathrm{MA}(\Delta P^+_{it}, s)}{\mathrm{MA}(\Delta P^-_{it}, s)}, \text{ and } RSI^s_{it} = 100\left(1 - \frac{1}{1 + RS^s_{it}}\right).$$

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 $ID^-_{it}(s)$ respectively, are computed by

$$ID^+_{it}(s) = 100\left(\frac{\mathrm{MA}(DM^+_{it}, s)}{\mathrm{MA}(TR_{it}, s)}\right), \text{ and } ID^-_{it}(s) = 100\left(\frac{\mathrm{MA}(DM^-_{it}, s)}{\mathrm{MA}(TR_{it}, s)}\right),$$

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

$$DX_{it}(s) = 100\left(\frac{|ID^{+}_{it}(s) - ID^{-}_{it}(s)|}{ID^{+}_{it}(s) + ID^{-}_{it}(s)}\right), \text{ and } ADX_{it}(s) = \text{MA}(DX_{it}(s), s).$$

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

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

where

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

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

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

For further details see Kaufman (2020).

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 S.1.

Table S.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 $s = \{3, 6, 12\}$, RSI_{it}^s for $s = \{3, 6, 12\}$, ADX_{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 $(s_1, s_2, m) = (2, 12, 6)$.

S-2 List of variables used when forecasting output growths

Variables in the conditioning and active sets for forecasting output growth across 33 countries are listed in Table S.2 below.

Table S.2: 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)	

S-3 Forecasting euro area output growths using ECB surveys 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.[10] 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 (2022) 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 S.3 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-

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

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 S.4, we compare OCMT (with no down-weighting at the selection stage) with Lasso and A-Lasso. The results indicate that the OCMT procedure outperforms Lasso and A-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 forecaster is selected for at least part of the evaluation period. As to be expected, A-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 A-Lasso, but the panel DM tests are not statistically significant. Moreover, none of the considered big data methods can beat the simple baseline models.

Table S.3: 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 S.4: Mean square forecast error (MSFE) and panel DM test of OCMT versus Lasso and A-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, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

Online Monte Carlo Supplement to "Variable Selection in High Dimensional Linear Regressions with Parameter Instability"

Alexander Chudik
Federal Reserve Bank of Dallas

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

Mahrad Sharifvaghefi
University of Pittsburgh

January 10, 2023

This online Monte Carlo supplement has three sections. Section S-1 explains the algorithms used for implementing Lasso, A-Lasso, Boosting and Cross-validation. We provide additional summary tables of our Monte Carlo simulation findings in Section S-2. The full set of Monte Carlo results for all the baseline experiments are provided in Section S-3.

S-1 Lasso, A-Lasso, Boosting and cross-validation algorithms

This section explains how Lasso, K-fold cross-validation, A-Lasso, and Boosting are implemented in this paper. Let $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$ be a $T \times 1$ vector of target variable, and let $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ be a $T \times m$ matrix of conditioning covariates where $\{\mathbf{z}_t : t = 1, 2, \cdots, T\}$ are $m \times 1$ vectors and let $\mathbf{X} = (\ \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)'$ be a $T \times N$ matrix of covariates in the active set where $\{\mathbf{x}_t : t = 1, 2, \cdots, T\}$ are $N \times 1$ vectors.

Lasso Procedure

1. *Construct the filtered variables* $\tilde{\mathbf{y}} = \mathbf{M}_z\mathbf{y}$ *and* $\tilde{\mathbf{X}} = \mathbf{M}_z\mathbf{X} = (\ \tilde{\mathbf{x}}_{1\circ}, \tilde{\mathbf{x}}_{2\circ}, ..., \tilde{\mathbf{x}}_{N\circ})$*, where* $\mathbf{M}_z = \mathbf{I}_T - \mathbf{Z}(\ \mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$*, and* $\tilde{\mathbf{x}}_{i\circ} = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$.

2. *Normalize each covariate* $\tilde{\mathbf{x}}_{i\circ} = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$ *by its* ℓ_2 *norm, such that*

$$\tilde{\mathbf{x}}^*_{i\circ} = \tilde{\mathbf{x}}_{i\circ}/\|\tilde{\mathbf{x}}_{i\circ}\|_2,$$

where $\|.\|_2$ *denotes the* ℓ_2 *norm of a vector. The corresponding matrix of normalized covariates in the active set is now denoted by* $\tilde{\mathbf{X}}^*$.

3. *For a given value of* $\varphi \geq 0$*, find* $\hat{\boldsymbol{\gamma}}_x^*(\varphi) \equiv [\hat{\gamma}_{1x}^*(\varphi), \hat{\gamma}_{2x}^*(\varphi), \cdots, \hat{\gamma}_{Nx}^*(\varphi)]'$ *such that*

$$\hat{\boldsymbol{\gamma}}_x^*(\varphi) = \arg\min_{\boldsymbol{\gamma}_x^*} \left\{ \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}^* \boldsymbol{\gamma}_x^*\|_2^2 + \varphi \|\boldsymbol{\gamma}_x^*\|_1 \right\},$$

where $\|.\|_1$ *denotes the* ℓ_1 *norm of a vector.*

4. *Divide* $\hat{\gamma}_{ix}^*(\varphi)$ *for* $i = 1, 2, \cdots, N$ *by* ℓ_2 *norm of the* $\tilde{\mathbf{x}}_{i\circ}$ *to match the original scale of* $\tilde{\mathbf{x}}_{i\circ}$*, namely set*

$$\hat{\gamma}_{ix}(\varphi) = \hat{\gamma}_{ix}^*(\varphi) / \|\tilde{\mathbf{x}}_{i\circ}\|_2,$$

where $\hat{\boldsymbol{\gamma}}_x(\varphi) \equiv [\hat{\gamma}_{1x}(\varphi), \hat{\gamma}_{2x}(\varphi), \cdots, \hat{\gamma}_{Nx}(\varphi)]'$ *denotes the vector of scaled coefficients.*

5. *Compute* $\hat{\boldsymbol{\gamma}}_z(\varphi) \equiv [\hat{\gamma}_{1z}(\varphi), \hat{\gamma}_{2z}(\varphi), \cdots, \hat{\gamma}_{mz}(\varphi)]'$ *by* $\hat{\boldsymbol{\gamma}}_z(\varphi) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}(\varphi)$ *where* $\hat{\mathbf{e}}(\varphi) = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}}_x(\varphi)$.

For a given set of values of φ's, say $\{\varphi_j : j = 1, 2, \cdots, h\}$, the optimal value of φ is chosen by K-fold cross-validation as described below.

K-fold Cross-validation

1. *Create a* $T \times 1$ *vector* $\mathbf{w} = (1, 2, \cdots, K, 1, 2, \cdots, K, \cdots)'$ *where* K *is the number of folds.*

2. *Let* $\mathbf{w}^* = (w_1^*, w_2^*, \cdots, w_T^*)'$ *be a* $T \times 1$ *vector generated by randomly permuting the elements of* $\mathbf{w}$.

3. *Group observations into* K *folds such that*

$$g_k = \{t : t \in \{1, 2, \cdots, T\} \text{ and } w_t^* = k\} \text{ for } k = 1, 2, \cdots, K.$$

4. *For a given value of* φ_j *and each fold* $k \in \{1, 2, \cdots, K\}$,

   (a) *Remove the observations related to fold* k *from the set of all observations.*

   (b) *Given the value of* φ_j*, use the remaining observations to estimate the coefficients of the model.*

   (c) *Use the estimated coefficients to compute predicted values of the target variable for the observations in fold* k *and hence compute mean square forecast error of fold* k *denoted by* $MSFE_k(\varphi_j)$.

5. *Compute the average mean square forecast error for a given value of φ_j by*

$$\overline{MSFE}(\varphi_j) = \sum_{k=1}^{K} MSFE_k(\varphi_j)/K.$$

6. *Repeat steps 1 to 5 for all values of $\{\varphi_j : j = 1, 2, \cdots, h\}$.*

7. *Select φ_j with the lowest corresponding average mean square forecast error as the optimal value of φ.*

In this study, following Friedman et al. (2010), we consider a sequence of 100 values of φ's decreasing from $\varphi_{\max}$ to $\varphi_{\min}$ on log scale where $\varphi_{\max} = \max_{i=1,2,\cdots,N}\left\{|\sum_{t=1}^{T} \tilde{x}_{it}^{*}\tilde{y}_t|\right\}$ and $\varphi_{\min} = 0.001\varphi_{\max}$. We use 10-fold cross-validation ($K = 10$) to find the optimal value of φ.

Denote $\hat{\boldsymbol{\gamma}}_x \equiv \hat{\boldsymbol{\gamma}}_x(\varphi_{op})$ where φ_{op} is the optimal value of φ obtained by the K-fold cross-validation. Given $\hat{\boldsymbol{\gamma}}_x$, we implement A-Lasso as described below.

A-Lasso

1. *Let $\mathcal{S} = \{i : i \in \{1, 2, \cdots, N\}$ and $\hat{\gamma}_{ix} \neq 0\}$ and $\mathbf{X}_{\mathcal{S}}$ be the $T \times s$ set of covariates in the active set with $\hat{\gamma}_{ix} \neq 0$ (from the Lasso step) where $s = |\mathcal{S}|$. Additionally, denote the corresponding $s \times 1$ vector of non-zero Lasso coefficients by $\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}} = (\hat{\gamma}_{1x,\mathcal{S}}, \hat{\gamma}_{2x,\mathcal{S}}, \cdots, \hat{\gamma}_{sx,\mathcal{S}})'$.*

2. *For a given value of $\psi \geq 0$, find $\hat{\boldsymbol{\delta}}^{*}_{x,\mathcal{S}}(\psi) \equiv [\hat{\delta}^{*}_{1x,\mathcal{S}}(\psi), \hat{\delta}^{*}_{2x,\mathcal{S}}(\psi), \cdots, \hat{\delta}^{*}_{sx,\mathcal{S}}(\psi)]'$ such that*

$$\hat{\boldsymbol{\delta}}^{*}_{x,\mathcal{S}}(\psi) = \arg\min_{\boldsymbol{\delta}^{*}_{x,\mathcal{S}}} \left\{ \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_{\mathcal{S}} diag(\hat{\boldsymbol{\gamma}}_{x,\ \mathcal{S}})\boldsymbol{\delta}^{*}_{x,\mathcal{S}}\|_2^2 + \psi\|\boldsymbol{\delta}^{*}_{x,\mathcal{S}}\|_1 \right\},$$

   where $diag(\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}})$ is an $s \times s$ diagonal matrix with its diagonal elements given by the corresponding elements of $\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}}$.

3. *Post multiply $\hat{\boldsymbol{\delta}}^{*}_{x,\mathcal{S}}(\psi)$ by $diag(\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}})$ to match the original scale of $\tilde{\mathbf{X}}_{\mathcal{S}}$, such that*

$$\hat{\boldsymbol{\delta}}_{x,\mathcal{S}}(\psi) = diag(\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}})\hat{\boldsymbol{\delta}}^{*}_{x,\mathcal{S}}(\psi).$$

   The coefficients of the covariates in the active set that belong to $\mathcal{S}^c$ are set equal to zero. In other words, $\hat{\boldsymbol{\delta}}_{x,\mathcal{S}^c}(\psi) = 0$ for all $\psi \geq 0$.

4. *Compute $\hat{\boldsymbol{\delta}}_z(\psi) \equiv [\hat{\delta}_{1z}(\psi), \hat{\delta}_{2z}(\psi), \cdots, \hat{\delta}_{mz}(\psi)]'$ by $\hat{\boldsymbol{\delta}}_z(\psi) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}(\psi)$ where $\hat{\mathbf{e}}(\psi) = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}_{\mathcal{S}}\hat{\boldsymbol{\delta}}_{x,\mathcal{S}}(\psi)$.*

As in the Lasso step, the optimal value ψ is set using 10-fold cross-validation as described before.[11]

Boosting

We implement boosting algorithm proposed by Buhlmann (2006).

1. *Consider the matrix of normalized filtered covariates* $\tilde{\mathbf{X}}^* = (\tilde{\mathbf{x}}^*_{1\circ}, \tilde{\mathbf{x}}^*_{2\circ}, ..., \tilde{\mathbf{x}}^*_{n\circ})$*, defined in Step 2 of the Lasso procedure above. Let the row* t *(for* $t = 1, 2, ..., T$*) of* $\tilde{\mathbf{X}}^*$ *be denoted as* $\tilde{\mathbf{x}}^{*\prime}_{\circ t} = (\tilde{x}^*_{1t}, \tilde{x}^*_{2t}, ..., \tilde{x}^*_{nt})$*. Given the normalized covariates matrix* $\tilde{\mathbf{X}}^*$ *and any vector* $\mathbf{e} = (e_1, e_2, .., e_T)'$*, define the least squares base procedure:*

$$\hat{g}_{\tilde{\mathbf{X}}^*,\mathbf{e}}(\tilde{\mathbf{x}}^*_{\circ t}) = \hat{\delta}_{\hat{s}}\tilde{x}^*_{\hat{s}t},\ \hat{s} = \arg\min_{1\le i\le n}\left(\mathbf{e} - \hat{\delta}_i\tilde{\mathbf{x}}^*_i\right)'\left(\mathbf{e} - \hat{\delta}_i\tilde{\mathbf{x}}^*_i\right),\ \ \hat{\delta}_i = \frac{\mathbf{e}'\tilde{\mathbf{x}}^*_i}{\tilde{\mathbf{x}}^{*\prime}_i\tilde{\mathbf{x}}^*_i},$$

2. *Given the normalized filtered covariates data* $\tilde{\mathbf{X}}^*$ *and the filtered target variable* $\tilde{\mathbf{y}} = \mathbf{M}_z\mathbf{y}$*, apply the base procedure to obtain* $\hat{g}^{(1)}_{\tilde{\mathbf{X}}^*,\tilde{\mathbf{y}}}(\tilde{\mathbf{x}}^*_{\circ t})$*. Set* $\hat{F}^{(1)}(\tilde{\mathbf{x}}^*_{\circ t}) = v\hat{g}^{(1)}_{\tilde{\mathbf{X}}^*,\tilde{\mathbf{y}}}(\tilde{\mathbf{x}}^*_{\circ t})$*, for some* $v > 0$*. Set* $\hat{s}^{(1)} = \hat{s}$ *and* $m = 1$*.*

3. *Compute the residual vector* $\mathbf{e}^{(m)} = \tilde{\mathbf{y}} - \hat{F}^{(m)}(\tilde{\mathbf{X}}^*)$*, where* $\hat{F}^{(m)}(\tilde{\mathbf{X}}^*) = [\hat{F}^{(m)}(\tilde{\mathbf{x}}^*_{\circ 1}), \hat{F}^{(m)}(\tilde{\mathbf{x}}^*_{\circ 2}),$ *...,* $\hat{F}^{(m)}(\tilde{\mathbf{x}}^*_{\circ T})]'$*, and fit the base procedure to these residuals to obtain the fit values* $\hat{g}^{(m+1)}_{\tilde{\mathbf{X}}^*,\mathbf{e}^{(m)}}(\tilde{\mathbf{x}}^*_{\circ t})$ *and* $\hat{s}^{(m)}$*. Update*

$$\hat{F}^{(m+1)}(\tilde{\mathbf{x}}^*_{\circ t}) = \hat{F}^{(m)}(\tilde{\mathbf{x}}^*_{\circ t}) + v\hat{g}^{(m+1)}_{\tilde{\mathbf{X}}^*,\mathbf{e}^{(m)}}(\tilde{\mathbf{x}}^*_{\circ t}).$$

4. *Increase the iteration index* m *by one and repeat Step 3 until the stopping iteration* M *is achieved. The stopping iteration is given by*

$$M = \arg\min_{1\le m\le m_{\max}} AIC_C(m),$$

for some predetermined large $m_{\max}$*, where*

$$AIC_C(m) = \log(\hat{\sigma}^2) + \frac{1 + tr\left(\mathcal{B}_m\right)/T}{1 - \left(tr\left(\mathcal{B}_m\right) + 2\right)/T},$$

$$\hat{\sigma}^2 = \frac{1}{T}\left(\tilde{\mathbf{y}} - \mathcal{B}_m\tilde{\mathbf{y}}\right)'\left(\tilde{\mathbf{y}} - \mathcal{B}_m\tilde{\mathbf{y}}\right),$$
$$\mathcal{B}_m = I - \left(I - v\mathcal{H}^{(\hat{s}_m)}\right)\left(I - v\mathcal{H}^{(\hat{s}_{m-1})}\right)...\left(I - v\mathcal{H}^{(\hat{s}_1)}\right),$$
$$\mathcal{H}^{(j)} = \frac{\tilde{\mathbf{x}}^*_{j\circ}\tilde{\mathbf{x}}^{*\prime}_{j\circ}}{\tilde{\mathbf{x}}^{*\prime}_{j\circ}\tilde{\mathbf{x}}^*_{j\circ}}.$$

[11] To implement Lasso, A-Lasso and 10-fold cross-validation we take advantage of glmnet package (Matlab version) available at `http://web.stanford.edu/~hastie/glmnet_matlab/`

We set $m_{\max} = 500$ and consider the same value for the tuning parameter $v = 0.1$ as suggested in Buhlmann (2006).

S-2 Additional Monte Carlo summary tables

Table S.1: Comparison of the effects of down-weighting for TPR performance in MC experiments with and without parameter instability.

	Average TPR								
Down-weighting:	No	Light	Heavy	No	Light	Heavy	No	Light	Heavy
$N\backslash T$	**100**			**200**			**500**		
A. Without parameter instability									
	OCMT (down-weighting at the selection stage)								
20	**0.80**	0.66	0.60	**0.95**	0.77	0.74	**1.00**	0.88	0.88
40	**0.77**	0.63	0.57	**0.94**	0.76	0.74	**1.00**	0.87	0.89
100	**0.73**	0.58	0.53	**0.91**	0.72	0.71	**1.00**	0.87	0.91
	Lasso								
20	**0.86**	0.80	0.75	**0.94**	0.84	0.76	**0.99**	0.86	0.78
40	**0.83**	0.78	0.74	**0.93**	0.82	0.77	**0.99**	0.83	0.77
100	**0.80**	0.75	0.72	**0.91**	0.81	0.77	**0.98**	0.82	0.78
	A-Lasso								
20	**0.77**	0.72	0.66	**0.88**	0.77	0.69	**0.97**	0.81	0.72
40	**0.76**	0.72	0.68	**0.89**	0.78	0.71	**0.97**	0.80	0.72
100	**0.75**	0.70	0.66	**0.88**	0.77	0.72	**0.97**	0.80	0.73
	Boosting								
20	**0.90**	0.89	0.89	**0.95**	0.94	0.93	**0.99**	0.97	0.95
40	**0.89**	0.89	0.87	**0.96**	0.93	0.90	**0.99**	0.95	0.91
100	**0.88**	0.86	0.82	**0.95**	0.91	0.85	**0.99**	0.92	0.86
B. With parameter instability									
	OCMT (down-weighting at the selection stage)								
20	**0.69**	0.58	0.57	**0.90**	0.76	0.75	**0.99**	0.90	0.90
40	**0.66**	0.54	0.54	**0.88**	0.75	0.75	**0.99**	0.90	0.91
100	**0.61**	0.50	0.49	**0.85**	0.70	0.72	**0.99**	0.90	0.93
	Lasso								
20	**0.78**	0.74	0.73	**0.89**	0.83	0.79	**0.97**	0.89	0.82
40	**0.75**	0.72	0.72	**0.88**	0.81	0.79	**0.97**	0.87	0.81
100	**0.72**	0.69	0.68	**0.85**	0.79	0.78	**0.96**	0.86	0.82
	A-Lasso								
20	**0.68**	0.66	0.64	**0.81**	0.76	0.71	**0.93**	0.84	0.76
40	**0.68**	0.65	0.65	**0.82**	0.76	0.73	**0.94**	0.83	0.77
100	**0.66**	0.63	0.61	**0.81**	0.75	0.73	**0.94**	0.83	0.78
	Boosting								
20	0.84	0.84	**0.87**	0.92	0.93	**0.93**	**0.98**	0.97	0.96
40	0.84	0.85	**0.85**	0.92	**0.92**	0.90	**0.98**	0.96	0.92
100	**0.83**	0.81	0.79	**0.91**	0.88	0.85	**0.98**	0.94	0.89

Notes: Down-weighting column label "No" stands for no down-weighting, "Light" stands for light down-weighting given by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$, and "Heavy" stands for heavy down-weighting given by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For each of the two sets of exponential down-weighting (light/heavy) we average TPR across the choices for λ. Best results are highlighted by bold fonts. The reported results are based on 4 experiments for models without parameter instabilities (panel A) and 4 experiments with parameter instabilities (panel B). See Section 5 for the description of the Monte Carlo design.

Table S.2: Comparison of the effects of down-weighting for FPR performance in MC experiments with and without parameter instability.

	Average FPR								
Down-weighting:	No	Light	Heavy	No	Light	Heavy	No	Light	Heavy
$N\backslash T$	**100**			**200**			**500**		
A. Without parameter instability									
	OCMT (down-weighting at the selection stage)								
20	**0.03**	0.04	0.09	**0.07**	0.10	0.21	**0.13**	0.27	0.42
40	**0.02**	0.03	0.08	**0.03**	0.09	0.22	**0.06**	0.27	0.46
100	**0.01**	0.02	0.07	**0.01**	0.07	0.21	**0.02**	0.26	0.51
	Lasso								
20	**0.18**	0.20	0.24	**0.18**	0.21	0.24	**0.18**	0.21	0.24
40	**0.13**	0.17	0.24	**0.13**	0.18	0.25	**0.13**	0.17	0.23
100	**0.08**	0.14	0.20	**0.07**	0.15	0.23	**0.07**	0.14	0.22
	A-Lasso								
20	**0.12**	0.14	0.17	**0.11**	0.14	0.17	**0.11**	0.15	0.17
40	**0.09**	0.13	0.19	**0.09**	0.13	0.19	**0.09**	0.13	0.17
100	**0.06**	0.11	0.15	**0.06**	0.12	0.18	**0.06**	0.12	0.17
	Boosting								
20	**0.28**	0.38	0.47	**0.29**	0.48	0.55	**0.29**	0.57	0.60
40	**0.31**	0.43	0.48	**0.32**	0.51	0.53	**0.32**	0.57	0.56
100	**0.32**	0.36	0.37	**0.33**	0.40	0.39	**0.34**	0.43	0.39
B. With parameter instability									
	OCMT (down-weighting at the selection stage)								
20	**0.02**	0.04	0.10	**0.05**	0.11	0.22	**0.10**	0.27	0.42
40	**0.01**	0.03	0.09	**0.02**	0.09	0.22	**0.05**	0.27	0.46
100	**0.01**	0.02	0.07	**0.01**	0.08	0.21	**0.02**	0.26	0.51
	Lasso								
20	**0.22**	0.24	0.27	**0.24**	0.24	0.26	0.27	**0.24**	0.26
40	**0.18**	0.21	0.27	**0.20**	0.21	0.27	0.22	**0.21**	0.26
100	**0.12**	0.17	0.21	**0.13**	0.18	0.25	**0.16**	0.17	0.25
	A-Lasso								
20	**0.16**	0.17	0.19	**0.17**	0.17	0.19	0.18	**0.17**	0.19
40	**0.13**	0.16	0.21	**0.15**	0.16	0.21	0.17	**0.16**	0.20
100	**0.09**	0.13	0.16	**0.10**	0.14	0.19	**0.13**	0.14	0.19
	Boosting								
20	**0.32**	0.41	0.49	**0.34**	0.49	0.56	**0.35**	0.59	0.61
40	**0.35**	0.45	0.50	**0.38**	0.53	0.54	**0.39**	0.58	0.57
100	**0.34**	0.38	0.38	**0.37**	0.42	0.40	**0.39**	0.44	0.40

Notes: Down-weighting column label "No" stands for no down-weighting, "Light" stands for light down-weighting given by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$, and "Heavy" stands for heavy down-weighting given by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For each of the two sets of exponential down-weighting (light/heavy) we average FPR across the choices for λ. Best results are highlighted by bold fonts. The reported results are based on 4 experiments for models without parameter instabilities (panel A) and 4 experiments with parameter instabilities (panel B). See Section 5 for the description of the Monte Carlo design.

Table S.3: Comparison of the effects of down-weighting for the number of selected variables $\hat{k}$ in MC experiments with and without parameter instability.

	Average $\hat{k}$								
Down-weighting:	No	Light	Heavy	No	Light	Heavy	No	Light	Heavy
$N\backslash T$		**100**			**200**			**500**	
A. Without parameter instability									
	OCMT (down-weighting at the selection stage)								
20	3.91	3.43	4.22	5.15	5.19	7.23	6.68	8.89	11.87
40	3.69	3.58	5.54	5.01	6.60	11.66	6.52	14.12	21.89
100	3.47	4.11	8.90	4.74	9.91	23.84	6.26	29.45	54.25
	Lasso								
20	6.93	7.24	7.75	7.33	7.45	7.76	7.54	7.70	7.83
40	8.48	9.95	12.71	8.89	10.30	12.87	8.99	10.29	12.22
100	11.17	16.86	22.60	10.95	18.07	26.51	10.98	17.63	25.57
	A-Lasso								
20	5.39	5.71	6.09	5.81	5.94	6.11	6.06	6.15	6.18
40	6.75	8.04	10.20	7.28	8.43	10.34	7.45	8.48	9.85
100	9.27	13.69	17.78	9.45	14.97	20.85	9.70	14.79	20.16
	Boosting								
20	9.19	11.19	13.00	9.58	13.27	14.70	9.74	15.34	15.80
40	16.04	20.55	22.83	16.52	24.12	24.88	16.78	26.64	25.92
100	35.32	39.91	40.07	36.64	44.02	42.22	37.72	46.28	42.89
B. With parameter instability									
	OCMT (down-weighting at the selection stage)								
20	3.25	3.15	4.18	4.55	5.18	7.34	5.93	9.00	11.94
40	3.10	3.44	5.64	4.41	6.72	11.88	5.85	14.32	21.99
100	2.96	4.28	9.29	4.23	10.32	24.20	5.71	29.78	54.33
	Lasso								
20	7.60	7.75	8.22	8.39	8.16	8.44	9.20	8.41	8.59
40	10.16	11.27	13.64	11.71	11.80	14.13	12.83	11.73	13.66
100	14.54	19.31	23.86	16.61	20.93	28.46	19.82	20.45	28.10
	A-Lasso								
20	5.84	6.06	6.42	6.58	6.46	6.63	7.40	6.72	6.81
40	7.97	8.97	10.86	9.40	9.55	11.30	10.51	9.63	11.01
100	11.58	15.40	18.64	13.70	17.14	22.32	16.73	17.04	22.14
	Boosting								
20	9.76	11.51	13.18	10.39	13.49	14.87	10.95	15.60	16.06
40	17.50	21.21	23.25	18.72	24.80	25.35	19.50	27.13	26.41
100	37.33	40.90	40.78	40.37	45.17	43.03	43.22	47.27	43.88

Notes: Down-weighting column label "No" stands for no down-weighting, "Light" stands for light down-weighting given by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$, and "Heavy" stands for heavy down-weighting given by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For each of the two sets of exponential down-weighting (light/heavy) we average $\hat{k}$ across the choices for λ. The reported results are based on 4 experiments for models without parameter instabilities (panel A) and 4 experiments with parameter instabilities (panel B). See Section 5 for the description of the Monte Carlo design.

Table S.4: The number of selected variables ($\hat{k}$), True Positive Rate (TRP), and False Positive Rate (FPR) averaged across MC experiments with and without dynamics.

	$\hat{k}$			TPR			FPR		
$N\backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Static									
	OCMT								
20	4.51	5.64	7.07	0.90	0.99	1.00	0.05	0.08	0.15
40	4.37	5.55	6.97	0.88	0.98	1.00	0.02	0.04	0.07
100	4.31	5.42	6.81	0.85	0.98	1.00	0.01	0.02	0.03
	Lasso								
20	7.83	8.32	8.75	0.89	0.96	1.00	0.21	0.22	0.24
40	9.88	11.15	11.83	0.88	0.96	1.00	0.16	0.18	0.20
100	13.25	14.96	17.44	0.85	0.95	0.99	0.10	0.11	0.13
	A-Lasso								
20	6.12	6.63	7.02	0.81	0.92	0.99	0.14	0.15	0.15
40	7.83	9.05	9.64	0.81	0.93	0.99	0.11	0.13	0.14
100	10.77	12.47	14.71	0.80	0.92	0.99	0.08	0.09	0.11
	Boosting								
20	9.74	10.27	10.61	0.92	0.97	1.00	0.30	0.32	0.33
40	16.76	17.90	18.55	0.92	0.97	1.00	0.33	0.35	0.36
100	35.57	38.09	40.72	0.91	0.97	1.00	0.32	0.34	0.37
B. Dynamic									
	OCMT								
20	2.65	4.06	5.54	0.60	0.86	0.99	0.01	0.03	0.08
40	2.41	3.87	5.40	0.55	0.84	0.99	0.01	0.01	0.04
100	2.13	3.55	5.16	0.49	0.79	0.98	0.00	0.00	0.01
	Lasso								
20	6.71	7.40	8.00	0.74	0.87	0.96	0.19	0.20	0.21
40	8.76	9.45	9.99	0.71	0.86	0.96	0.15	0.15	0.15
100	12.46	12.60	13.36	0.66	0.82	0.95	0.10	0.09	0.10
	A-Lasso								
20	5.12	5.76	6.44	0.63	0.78	0.92	0.13	0.13	0.14
40	6.89	7.64	8.31	0.62	0.79	0.92	0.11	0.11	0.12
100	10.08	10.67	11.72	0.60	0.78	0.93	0.08	0.08	0.08
	Boosting								
20	9.21	9.70	10.08	0.81	0.90	0.97	0.30	0.30	0.31
40	16.78	17.35	17.73	0.81	0.90	0.97	0.34	0.34	0.35
100	37.09	38.92	40.22	0.80	0.89	0.97	0.34	0.35	0.36

Notes: There are $k = 4$ signal variables out of N observed covariates. The top panel reports results averaged across 4 static experiments, which do not feature lagged dependent variable. The bottom panel reports results averaged across 4 dynamic experiments featuring lagged dependent variable. Each experiment is based on 2000 Monte Carlo simulations. OCMT, Lasso and A-Lasso methods in this table are based on original (not down-weighted) observations. See Section 5 of the paper for the detailed description of the Monte Carlo design.

Table S.5: Comparison of the effects of down-weighting on one-step-ahead MSFE of OCMT, Lasso, A-Lasso and Boosting averaged across all the static MC experiments without and with parameter instabilities.

Down-weighting:	No	Light	Heavy	No	Light	Heavy	No	Light	Heavy
$N\backslash T$		**100**			**200**			**300**	
A. Without parameter instabilities									
	OCMT(Down-weighling only at the estimation stage)								
20	**15.01**	15.38	16.11	**12.60**	12.98	13.69	**11.98**	12.42	13.21
40	**13.92**	14.32	14.97	**14.76**	15.99	17.17	**12.10**	12.61	13.44
100	**14.18**	14.33	14.81	**13.91**	14.29	15.22	**12.04**	12.81	14.10
	OCMT(Down-weighling only at the variable selection and estimation stages)								
20	**15.01**	15.50	16.95	**12.60**	13.30	15.25	**11.98**	13.13	16.14
40	**13.92**	14.54	16.22	**14.76**	16.70	20.71	**12.10**	15.03	23.48
100	**14.18**	14.72	17.70	**13.91**	15.57	22.84	**12.04**	19.17	38.08
	Lasso								
20	**15.35**	15.88	16.92	**12.79**	13.17	14.05	**12.01**	12.63	13.52
40	**14.33**	15.05	16.30	**14.74**	16.10	17.87	**12.09**	13.05	14.28
100	**14.57**	15.91	18.26	**14.22**	15.46	17.02	**12.11**	13.65	15.82
	A-Lasso								
20	**15.92**	16.44	17.58	**13.13**	13.45	14.60	**12.05**	12.82	13.94
40	**15.31**	16.06	17.48	**15.24**	17.06	18.94	**12.30**	13.49	15.03
100	**16.38**	17.90	20.29	**15.29**	16.89	18.51	**12.45**	14.75	17.19
	Boosting								
20	**16.12**	17.42	19.99	**13.06**	14.68	17.39	**12.15**	14.09	16.14
40	**15.28**	17.01	19.25	**15.64**	18.87	21.44	**12.40**	15.47	17.97
100	**17.41**	19.61	21.16	**15.88**	18.41	20.52	**12.60**	16.38	18.65
B. With parameter instabilities									
	OCMT(Down-weighling only at the estimation stage)								
20	18.43	**17.56**	17.59	15.21	**13.90**	14.32	14.67	**13.08**	13.85
40	16.88	**16.22**	16.35	17.37	**16.80**	17.59	14.37	**13.21**	13.95
100	17.41	**16.78**	16.79	16.92	**15.49**	16.05	15.05	**13.28**	14.36
	OCMT(Down-weighling only at the variable selection and estimation stages)								
20	18.43	18.11	18.93	15.21	**14.27**	16.21	14.67	**13.97**	16.93
40	16.88	16.93	18.49	**17.37**	18.23	22.38	**14.37**	16.02	24.45
100	17.41	17.39	20.94	**16.92**	17.71	25.90	**15.05**	20.19	39.64
	Lasso								
20	19.23	**18.65**	18.98	15.65	**14.43**	15.04	14.97	**13.51**	14.37
40	17.73	**17.47**	18.39	17.89	**17.74**	19.29	14.70	**14.08**	15.50
100	**18.31**	18.68	20.45	17.86	**17.35**	18.61	15.38	**14.85**	17.10
	A-Lasso								
20	19.78	**18.91**	19.42	15.83	**14.47**	15.46	14.91	**13.59**	14.73
40	18.77	**18.38**	19.70	18.63	**18.52**	20.33	14.80	**14.37**	16.13
100	**20.21**	20.60	22.46	19.44	**18.96**	20.27	**15.92**	15.98	18.59
	Boosting								
20	**19.40**	19.53	21.42	**15.73**	15.74	18.48	14.99	**14.94**	17.12
40	**18.51**	19.58	21.63	**18.39**	20.15	22.69	**14.76**	16.45	19.15
100	**19.95**	21.75	23.22	**18.95**	20.05	22.05	**15.60**	17.41	20.00

Notes: The reported results are averaged over two experiments (low fit and high fit) for models without and with parameter instabilities. See Section 5 for the description of the Monte Carlo design. Full set of results is presented in the online Monte Carlo supplement.

†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.

Table S.6: Comparison of the effects of down-weighting on one-step-ahead MSFE of OCMT, Lasso, A-Lasso and Boosting averaged across all the dynamic MC experiments without and with parameter instabilities.

Down-weighting:	No	Light	Heavy	No	Light	Heavy	No	Light	Heavy
$N\backslash T$		**100**			**200**			**300**	
A. Without parameter instabilities									
	OCMT(Down-weighling only at the estimation stage)								
20	**43.69**	44.93	46.82	**35.56**	36.54	38.11	**33.96**	35.59	37.85
40	**39.75**	40.11	41.09	**44.15**	46.01	48.36	**33.86**	34.79	36.81
100	**40.55**	40.81	41.77	**39.54**	40.15	42.13	**33.79**	36.09	39.09
	OCMT(Down-weighling only at the variable selection and estimation stages)								
20	**43.69**	45.10	48.31	**35.56**	37.31	41.64	**33.96**	37.00	45.10
40	**39.75**	40.06	43.03	**44.15**	46.93	55.54	**33.86**	40.01	59.11
100	**40.55**	41.52	45.42	**39.54**	42.18	50.90	**33.79**	47.82	87.50
	Lasso								
20	**43.73**	46.17	49.15	**35.78**	37.35	39.99	**34.03**	35.99	38.90
40	**40.45**	42.27	47.10	**44.24**	46.66	52.52	**34.11**	36.63	41.38
100	**41.70**	45.49	50.11	**39.96**	43.11	48.45	**34.12**	39.03	45.57
	A-Lasso								
20	**45.78**	47.97	51.21	**36.76**	38.20	41.43	**34.24**	36.62	40.03
40	**43.37**	44.87	49.92	**45.72**	48.17	53.60	**34.87**	37.93	43.69
100	**48.19**	51.62	55.64	**43.73**	47.05	52.49	**35.28**	41.96	49.09
	Boosting								
20	**47.27**	53.10	61.26	**37.08**	42.71	50.83	**34.93**	42.07	49.33
40	**44.94**	51.47	58.55	**48.02**	59.09	66.52	**35.41**	45.17	52.48
100	**52.98**	59.26	63.31	**46.79**	53.93	59.42	**36.96**	49.74	57.10
B. With parameter instabilities									
	OCMT(Down-weighling only at the estimation stage)								
20	48.94	**47.98**	48.88	39.15	**37.60**	38.71	37.84	**36.08**	38.23
40	44.22	**42.86**	43.19	47.93	**47.06**	48.85	36.99	**35.25**	36.99
100	45.79	**44.71**	45.01	44.45	**42.42**	43.94	38.21	**36.72**	39.32
	OCMT(Down-weighling only at the variable selection and estimation stages)								
20	48.94	**48.64**	50.65	39.15	**38.69**	42.68	**37.84**	38.02	46.00
40	44.22	**43.22**	45.82	**47.93**	48.35	57.80	**36.99**	41.24	60.26
100	45.79	**45.53**	49.20	**44.45**	44.49	55.91	**38.21**	48.93	89.29
	Lasso								
20	**49.78**	50.31	52.30	40.03	**39.08**	41.51	38.18	**37.00**	40.26
40	**44.88**	45.05	48.91	48.65	**48.27**	53.76	**37.90**	38.09	42.85
100	**46.16**	48.68	53.01	**45.66**	45.68	50.52	**38.98**	40.40	46.86
	A-Lasso								
20	51.79	**51.79**	54.11	40.73	**39.51**	42.67	38.30	**37.39**	41.32
40	47.81	**47.72**	51.83	50.05	**49.57**	55.13	**38.13**	39.02	45.11
100	**52.08**	54.84	58.84	49.64	**49.62**	54.93	**39.98**	43.20	50.43
	Boosting								
20	**51.06**	56.26	64.10	**39.63**	44.16	52.87	**37.85**	43.64	51.43
40	**48.85**	55.87	62.90	**50.13**	60.47	68.29	**37.66**	47.15	55.00
100	**55.03**	63.10	67.48	**49.69**	56.77	62.37	**39.75**	51.63	59.53

Notes: The reported results are averaged across two experiments (low fit and high fit) for models without and with parameter instabilities. See Section 5 for the description of the Monte Carlo design. Full set of results is presented in the online Monte Carlo supplement.

†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.

S-3 Monte Carlo results for all the experiments

S-3.1 MC findings for baseline experiments without parameter instabilities

Table S.7: MC results for methods using no down-weighting in the baseline experiment with no dynamics ($\rho_y = 0$) and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	22.85	19.62	18.67	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	21.29	22.69	18.67	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	21.66	21.57	18.73	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	23.59	19.75	18.77	4.15	5.33	6.78	0.88	0.99	1.00	0.03	0.07	0.14
40	21.82	23.08	18.96	3.95	5.20	6.62	0.85	0.98	1.00	0.01	0.03	0.07
100	22.28	21.84	18.85	3.71	4.95	6.33	0.80	0.98	1.00	0.01	0.01	0.02
	LASSO											
20	24.00	20.02	18.83	6.91	7.34	7.52	0.87	0.96	1.00	0.17	0.17	0.18
40	22.35	23.04	18.96	8.16	8.89	9.04	0.85	0.96	1.00	0.12	0.13	0.13
100	22.76	22.30	18.99	10.22	10.64	10.83	0.81	0.94	1.00	0.07	0.07	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	25.74	20.68	19.09	6.91	7.34	7.52	0.87	0.96	1.00	0.17	0.17	0.18
40	24.40	24.18	19.59	8.16	8.89	9.04	0.85	0.96	1.00	0.12	0.13	0.13
100	26.74	24.44	19.63	10.22	10.64	10.83	0.81	0.94	1.00	0.07	0.07	0.07
	A-LASSO											
20	24.94	20.62	18.89	5.30	5.84	6.18	0.76	0.90	0.99	0.11	0.11	0.11
40	23.90	23.81	19.30	6.40	7.28	7.65	0.75	0.91	0.99	0.08	0.09	0.09
100	25.74	24.02	19.53	8.49	9.19	9.71	0.75	0.90	0.99	0.06	0.06	0.06
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	25.49	20.72	18.97	5.30	5.84	6.18	0.76	0.90	0.99	0.11	0.11	0.11
40	24.45	24.08	19.46	6.40	7.28	7.65	0.75	0.91	0.99	0.08	0.09	0.09
100	26.54	24.40	19.67	8.49	9.19	9.71	0.75	0.90	0.99	0.06	0.06	0.06
	Boosting											
20	24.95	20.33	18.99	9.05	9.56	9.73	0.91	0.97	1.00	0.27	0.28	0.29
40	23.58	24.34	19.38	15.52	16.31	16.65	0.90	0.97	1.00	0.30	0.31	0.32
100	26.97	24.78	19.71	33.89	35.47	37.07	0.89	0.97	1.00	0.30	0.32	0.33
	Boosting for variable selection only. LS for estimation/forecasting.											
20	26.80	20.96	19.14	9.05	9.56	9.73	0.91	0.97	1.00	0.27	0.28	0.29
40	27.73	27.28	20.32	15.52	16.31	16.65	0.90	0.97	1.00	0.30	0.31	0.32
100	40.82	32.19	22.38	33.89	35.47	37.07	0.89	0.97	1.00	0.30	0.32	0.33

Notes: This table reports one-step-ahead Mean Square Forecast Error (MSFE, $\times 100$) , average number of selected variables ($\hat{k}$), True Positive Rate (TPR), and False Positive Rate (FPR). The baseline model features no parameter instabilities in slopes and intercepts. There are $k = 4$ signals variables out of N observed variables. The DGP is given by $y_t = d + \rho_y y_{t-1} + \sum_{j=1}^{4} \beta_j x_{jt} + \tau_u u_t$. Oracle model assumes the identity of signal variables is known. The reported results are based on 2000 Monte Carlo replications. See Section 5 of the paper for the detailed description of the Monte Carlo design.

Table S.8: MC results for methods using light down-weighting in the baseline experiment with no dynamics ($\rho_y = 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	23.41	20.05	19.10	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	22.08	24.16	19.02	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	22.05	21.94	19.56	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	24.12	20.29	19.38	4.15	5.33	6.78	0.88	0.99	1.00	0.03	0.07	0.14
40	22.40	24.92	19.74	3.95	5.20	6.62	0.85	0.98	1.00	0.01	0.03	0.07
100	22.40	22.41	19.98	3.71	4.95	6.33	0.80	0.98	1.00	0.01	0.01	0.02
	LASSO											
20	26.46	21.71	20.06	6.91	7.34	7.52	0.87	0.96	1.00	0.17	0.17	0.18
40	24.89	26.71	21.31	8.16	8.89	9.04	0.85	0.96	1.00	0.12	0.13	0.13
100	27.75	25.52	21.01	10.22	10.64	10.83	0.81	0.94	1.00	0.07	0.07	0.07
	A-LASSO											
20	25.85	21.50	19.67	5.30	5.84	6.18	0.76	0.90	0.99	0.11	0.11	0.11
40	24.95	26.28	20.85	6.40	7.28	7.65	0.75	0.91	0.99	0.08	0.09	0.09
100	27.36	25.21	20.98	8.49	9.19	9.71	0.75	0.90	0.99	0.06	0.06	0.06
	Boosting											
20	27.67	22.18	20.44	9.05	9.56	9.73	0.91	0.97	1.00	0.27	0.28	0.29
40	28.91	30.74	23.34	15.52	16.31	16.65	0.90	0.97	1.00	0.30	0.31	0.32
100	43.40	36.87	29.47	33.89	35.47	37.07	0.89	0.97	1.00	0.30	0.32	0.33
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	24.19	20.80	20.54	3.67	5.69	9.58	0.71	0.82	0.90	0.04	0.12	0.30
40	22.66	26.01	23.46	3.89	7.44	15.52	0.67	0.80	0.90	0.03	0.11	0.30
100	22.90	24.00	29.85	4.57	11.61	32.76	0.62	0.77	0.89	0.02	0.09	0.29
	LASSO											
20	24.73	20.57	19.76	6.99	7.20	7.51	0.80	0.84	0.86	0.19	0.19	0.20
40	23.39	25.08	20.37	9.10	9.57	9.60	0.77	0.82	0.83	0.15	0.16	0.16
100	24.74	24.09	21.25	14.32	14.88	14.77	0.73	0.79	0.81	0.11	0.12	0.12
	A-LASSO											
20	25.65	21.05	20.07	5.46	5.73	6.01	0.70	0.77	0.81	0.13	0.13	0.14
40	24.95	26.65	21.05	7.30	7.83	7.96	0.70	0.77	0.80	0.11	0.12	0.12
100	27.88	26.31	22.97	11.62	12.38	12.48	0.68	0.75	0.78	0.09	0.09	0.09
	Boosting											
20	26.98	22.72	21.91	11.03	13.18	15.30	0.89	0.95	0.97	0.37	0.47	0.57
40	26.27	29.37	24.06	20.12	23.96	26.63	0.89	0.94	0.95	0.41	0.50	0.57
100	30.40	28.62	25.50	39.22	43.74	46.27	0.86	0.91	0.93	0.36	0.40	0.43

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.7.

Table S.9: MC results for methods using heavy down-weighting in the baseline experiment with no dynamics ($\rho_y = 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	24.49	20.88	19.90	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	23.26	25.26	19.64	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	22.95	23.00	20.83	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	25.20	21.35	20.56	4.15	5.33	6.78	0.88	0.99	1.00	0.03	0.07	0.14
40	23.33	26.64	20.96	3.95	5.20	6.62	0.85	0.98	1.00	0.01	0.03	0.07
100	22.99	23.86	21.87	3.71	4.95	6.33	0.80	0.98	1.00	0.01	0.01	0.02
	LASSO											
20	27.80	23.30	21.34	6.91	7.34	7.52	0.87	0.96	1.00	0.17	0.17	0.18
40	26.12	28.89	23.27	8.16	8.89	9.04	0.85	0.96	1.00	0.12	0.13	0.13
100	29.45	27.80	23.68	10.22	10.64	10.83	0.81	0.94	1.00	0.07	0.07	0.07
	A-LASSO											
20	26.77	22.73	20.63	5.30	5.84	6.18	0.76	0.90	0.99	0.11	0.11	0.11
40	26.09	27.95	22.74	6.40	7.28	7.65	0.75	0.91	0.99	0.08	0.09	0.09
100	28.69	27.13	23.30	8.49	9.19	9.71	0.75	0.90	0.99	0.06	0.06	0.06
	Boosting											
20	29.09	24.37	22.35	9.05	9.56	9.73	0.91	0.97	1.00	0.27	0.28	0.29
40	31.81	33.68	27.32	15.52	16.31	16.65	0.90	0.97	1.00	0.30	0.31	0.32
100	49.34	46.27	41.50	33.89	35.47	37.07	0.89	0.97	1.00	0.30	0.32	0.33
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	26.41	23.78	25.23	4.63	7.89	12.46	0.63	0.77	0.89	0.10	0.24	0.44
40	25.02	31.99	36.73	6.24	12.87	22.94	0.60	0.77	0.90	0.10	0.25	0.48
100	27.02	34.90	58.90	10.41	26.66	56.42	0.55	0.74	0.92	0.08	0.24	0.53
	LASSO											
20	26.26	21.87	21.05	7.20	7.11	7.27	0.72	0.74	0.75	0.22	0.21	0.21
40	25.24	27.66	22.08	11.24	11.20	10.54	0.71	0.73	0.73	0.21	0.21	0.19
100	28.33	26.33	24.46	20.32	21.38	20.20	0.68	0.72	0.73	0.18	0.18	0.17
	A-LASSO											
20	27.35	22.77	21.80	5.62	5.60	5.74	0.63	0.66	0.69	0.15	0.15	0.15
40	27.02	29.35	23.28	8.97	8.97	8.52	0.64	0.67	0.68	0.16	0.16	0.14
100	31.56	28.65	26.63	15.87	16.92	16.06	0.62	0.67	0.68	0.13	0.14	0.13
	Boosting											
20	30.90	26.82	25.02	12.85	14.63	15.77	0.88	0.93	0.95	0.47	0.54	0.60
40	29.66	33.27	27.84	22.55	24.77	25.90	0.87	0.90	0.91	0.48	0.53	0.56
100	32.68	31.73	28.89	39.58	41.97	42.74	0.81	0.84	0.86	0.36	0.39	0.39

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.7.

Table S.10: MC results for methods using no down-weighting in the baseline experiment with no dynamics ($\rho_y = 0$) and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$n\backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	6.30	5.41	5.14	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	5.87	6.25	5.14	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	5.97	5.94	5.16	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	6.44	5.45	5.20	5.43	6.48	8.12	0.99	1.00	1.00	0.07	0.12	0.21
40	6.02	6.43	5.24	5.29	6.40	7.98	0.99	1.00	1.00	0.03	0.06	0.10
100	6.09	5.98	5.22	5.27	6.23	7.77	0.98	1.00	1.00	0.01	0.02	0.04
	LASSO											
20	6.70	5.56	5.19	7.61	7.59	7.54	0.98	1.00	1.00	0.18	0.18	0.18
40	6.31	6.44	5.22	9.11	9.27	9.06	0.98	1.00	1.00	0.13	0.13	0.13
100	6.37	6.15	5.23	11.48	11.13	10.82	0.97	1.00	1.00	0.08	0.07	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	7.12	5.69	5.25	7.61	7.59	7.54	0.98	1.00	1.00	0.18	0.18	0.18
40	6.90	6.82	5.39	9.11	9.27	9.06	0.98	1.00	1.00	0.13	0.13	0.13
100	7.43	6.74	5.44	11.48	11.13	10.82	0.97	1.00	1.00	0.08	0.07	0.07
	A-LASSO											
20	6.90	5.64	5.20	6.17	6.18	5.86	0.96	0.99	1.00	0.12	0.11	0.09
40	6.72	6.67	5.31	7.47	7.63	7.16	0.95	0.99	1.00	0.09	0.09	0.08
100	7.02	6.57	5.37	9.73	9.64	9.18	0.95	0.99	1.00	0.06	0.06	0.05
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	7.03	5.67	5.22	6.17	6.18	5.86	0.96	0.99	1.00	0.12	0.11	0.09
40	6.90	6.77	5.35	7.47	7.63	7.16	0.95	0.99	1.00	0.09	0.09	0.08
100	7.21	6.69	5.39	9.73	9.64	9.18	0.95	0.99	1.00	0.06	0.06	0.05
	Boosting											
20	7.29	5.79	5.32	9.54	9.78	9.84	0.99	1.00	1.00	0.28	0.29	0.29
40	6.97	6.95	5.43	16.03	16.51	16.76	0.99	1.00	1.00	0.30	0.31	0.32
100	7.85	6.97	5.50	34.42	35.69	37.14	0.98	1.00	1.00	0.30	0.32	0.33
	Boosting for variable selection only. LS for estimation/forecasting.											
20	7.50	5.79	5.28	9.54	9.78	9.84	0.99	1.00	1.00	0.28	0.29	0.29
40	7.67	7.56	5.60	16.03	16.51	16.76	0.99	1.00	1.00	0.30	0.31	0.32
100	11.53	8.82	6.16	34.42	35.69	37.14	0.98	1.00	1.00	0.30	0.32	0.33

Notes: See notes to Table S.7.

Table S.11: MC results for methods using light down-weighting in the baseline experiment with no dynamics ($\rho_y = 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
	Oracle											
20	6.45	5.52	5.26	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	6.08	6.66	5.24	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	6.08	6.05	5.39	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	6.65	5.66	5.46	5.43	6.48	8.12	0.99	1.00	1.00	0.07	0.12	0.21
40	6.25	7.05	5.47	5.29	6.40	7.98	0.99	1.00	1.00	0.03	0.06	0.10
100	6.26	6.18	5.64	5.27	6.23	7.77	0.98	1.00	1.00	0.01	0.02	0.04
	LASSO											
20	7.33	5.98	5.53	7.61	7.59	7.54	0.98	1.00	1.00	0.18	0.18	0.18
40	7.14	7.63	5.88	9.11	9.27	9.06	0.98	1.00	1.00	0.13	0.13	0.13
100	7.72	7.12	5.79	11.48	11.13	10.82	0.97	1.00	1.00	0.08	0.07	0.07
	A-LASSO											
20	7.17	5.93	5.42	6.17	6.18	5.86	0.96	0.99	1.00	0.12	0.11	0.09
40	7.10	7.46	5.71	7.47	7.63	7.16	0.95	0.99	1.00	0.09	0.09	0.08
100	7.52	7.02	5.74	9.73	9.64	9.18	0.95	0.99	1.00	0.06	0.06	0.05
	Boosting											
20	7.78	6.17	5.61	9.54	9.78	9.84	0.99	1.00	1.00	0.28	0.29	0.29
40	8.04	8.55	6.46	16.03	16.51	16.76	0.99	1.00	1.00	0.30	0.31	0.32
100	12.16	10.14	7.99	34.42	35.69	37.14	0.98	1.00	1.00	0.30	0.32	0.33
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	6.82	5.80	5.72	5.13	7.13	11.02	0.90	0.92	0.95	0.08	0.17	0.36
40	6.42	7.39	6.59	5.70	9.60	17.97	0.89	0.92	0.95	0.05	0.15	0.35
100	6.55	7.13	8.49	7.20	15.43	37.53	0.87	0.90	0.96	0.04	0.12	0.34
	LASSO											
20	7.03	5.77	5.50	8.16	8.27	8.46	0.96	0.98	0.98	0.22	0.22	0.23
40	6.70	7.11	5.74	10.69	11.07	11.06	0.96	0.97	0.97	0.17	0.18	0.18
100	7.08	6.83	6.06	16.61	17.34	17.10	0.94	0.97	0.97	0.13	0.13	0.13
	A-LASSO											
20	7.23	5.85	5.58	6.60	6.70	6.73	0.93	0.95	0.96	0.14	0.14	0.14
40	7.17	7.48	5.94	8.79	9.12	9.07	0.92	0.95	0.96	0.13	0.13	0.13
100	7.91	7.47	6.52	13.64	14.51	14.36	0.92	0.95	0.96	0.10	0.11	0.11
	Boosting											
20	7.87	6.64	6.27	11.57	13.46	15.42	0.98	0.99	1.00	0.38	0.47	0.57
40	7.75	8.37	6.87	20.63	24.15	26.66	0.98	0.99	0.99	0.42	0.50	0.57
100	8.82	8.20	7.27	39.52	43.69	46.02	0.97	0.99	0.99	0.36	0.40	0.42

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.7.

Table S.12: MC results for methods using heavy down-weighting in the baseline experiment with no dynamics ($\rho_y = 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	6.75	5.75	5.48	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	6.41	6.96	5.41	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	6.32	6.34	5.74	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	7.02	6.03	5.86	5.43	6.48	8.12	0.99	1.00	1.00	0.07	0.12	0.21
40	6.61	7.71	5.92	5.29	6.40	7.98	0.99	1.00	1.00	0.03	0.06	0.10
100	6.63	6.59	6.34	5.27	6.23	7.77	0.98	1.00	1.00	0.01	0.02	0.04
	LASSO											
20	7.70	6.45	5.87	7.61	7.59	7.54	0.98	1.00	1.00	0.18	0.18	0.18
40	7.57	8.38	6.44	9.11	9.27	9.06	0.98	1.00	1.00	0.13	0.13	0.13
100	8.23	7.79	6.51	11.48	11.13	10.82	0.97	1.00	1.00	0.08	0.07	0.07
	A-LASSO											
20	7.47	6.30	5.68	6.17	6.18	5.86	0.96	0.99	1.00	0.12	0.11	0.09
40	7.44	8.05	6.19	7.47	7.63	7.16	0.95	0.99	1.00	0.09	0.09	0.08
100	7.96	7.61	6.37	9.73	9.64	9.18	0.95	0.99	1.00	0.06	0.06	0.05
	Boosting											
20	8.22	6.77	6.14	9.54	9.78	9.84	0.99	1.00	1.00	0.28	0.29	0.29
40	8.85	9.47	7.58	16.03	16.51	16.76	0.99	1.00	1.00	0.30	0.31	0.32
100	13.67	12.67	11.09	34.42	35.69	37.14	0.98	1.00	1.00	0.30	0.32	0.33
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	7.48	6.73	7.04	6.18	9.32	13.59	0.80	0.87	0.94	0.15	0.29	0.49
40	7.42	9.43	10.23	8.46	15.23	24.69	0.79	0.88	0.95	0.13	0.29	0.52
100	8.39	10.78	17.26	14.22	31.12	59.27	0.77	0.87	0.96	0.11	0.28	0.55
	LASSO											
20	7.58	6.24	5.98	8.75	8.70	8.77	0.92	0.93	0.93	0.25	0.25	0.25
40	7.36	8.08	6.49	13.34	13.45	12.89	0.92	0.92	0.92	0.24	0.24	0.23
100	8.20	7.70	7.18	22.70	24.34	23.42	0.91	0.93	0.92	0.19	0.21	0.20
	A-LASSO											
20	7.80	6.43	6.09	6.99	6.92	6.92	0.87	0.89	0.90	0.18	0.17	0.17
40	7.94	8.52	6.78	10.81	10.83	10.36	0.88	0.89	0.89	0.18	0.18	0.17
100	9.02	8.38	7.76	17.73	19.20	18.49	0.88	0.90	0.90	0.14	0.16	0.15
	Boosting											
20	9.07	7.96	7.26	13.29	14.85	15.89	0.97	0.98	0.99	0.47	0.55	0.60
40	8.84	9.61	8.09	22.93	24.95	26.00	0.96	0.97	0.97	0.48	0.53	0.55
100	9.64	9.31	8.42	39.81	41.99	42.68	0.95	0.96	0.96	0.36	0.38	0.39

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.7.

Table S.13: MC results for methods using no down-weighting in the baseline experiment with dynamics ($\rho_y \neq 0$) and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$n \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	65.06	54.52	52.06	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	59.58	67.25	51.88	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	61.06	60.22	51.99	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	67.41	54.94	52.45	1.97	3.62	5.22	0.46	0.81	0.99	0.01	0.02	0.06
40	61.62	68.21	52.23	1.65	3.41	5.09	0.38	0.78	0.99	0.00	0.01	0.03
100	62.82	61.15	52.17	1.36	3.02	4.83	0.32	0.70	0.98	0.00	0.00	0.01
	LASSO											
20	67.13	55.13	52.54	5.89	6.73	7.41	0.67	0.83	0.96	0.16	0.17	0.18
40	62.31	68.14	52.68	7.44	8.15	8.75	0.62	0.81	0.95	0.12	0.12	0.12
100	63.95	61.52	52.65	10.54	10.35	10.93	0.57	0.76	0.94	0.08	0.07	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	71.16	57.19	53.33	5.89	6.73	7.41	0.67	0.83	0.96	0.16	0.17	0.18
40	67.75	71.58	54.27	7.44	8.15	8.75	0.62	0.81	0.95	0.12	0.12	0.12
100	77.15	68.38	55.00	10.54	10.35	10.93	0.57	0.76	0.94	0.08	0.07	0.07
	A-LASSO											
20	70.45	56.61	52.91	4.43	5.09	5.90	0.55	0.71	0.89	0.11	0.11	0.12
40	66.78	70.52	53.83	5.81	6.51	7.26	0.53	0.71	0.90	0.09	0.09	0.09
100	73.92	67.11	54.50	8.56	8.82	9.74	0.50	0.70	0.91	0.07	0.06	0.06
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	71.87	57.19	53.15	4.43	5.09	5.90	0.55	0.71	0.89	0.11	0.11	0.12
40	68.30	71.26	54.27	5.81	6.51	7.26	0.53	0.71	0.90	0.09	0.09	0.09
100	76.07	68.21	54.76	8.56	8.82	9.74	0.50	0.70	0.91	0.07	0.06	0.06
	Boosting											
20	71.62	56.38	53.22	8.72	9.22	9.59	0.76	0.87	0.97	0.28	0.29	0.29
40	67.92	73.13	54.06	16.07	16.41	16.75	0.76	0.87	0.97	0.33	0.32	0.32
100	80.36	71.25	56.06	36.51	37.74	38.42	0.74	0.86	0.97	0.34	0.34	0.35
	Boosting for variable selection only. LS for estimation/forecasting.											
20	76.86	57.22	53.33	8.72	9.22	9.59	0.76	0.87	0.97	0.28	0.29	0.29
40	80.99	77.62	56.84	16.07	16.41	16.75	0.76	0.87	0.97	0.33	0.32	0.32
100	114.48	93.27	62.23	36.51	37.74	38.42	0.74	0.86	0.97	0.34	0.34	0.35

Notes: See notes to Table S.7.

Table S.14: MC results for methods using light down-weighting in the baseline experiment with dynamics ($\rho_y \neq 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	67.66	56.22	54.33	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	61.57	69.79	52.83	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	63.05	61.72	55.30	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	69.20	56.30	54.92	1.97	3.62	5.22	0.46	0.81	0.99	0.01	0.02	0.06
40	61.85	70.78	53.55	1.65	3.41	5.09	0.38	0.78	0.99	0.00	0.01	0.03
100	63.20	61.91	55.68	1.36	3.02	4.83	0.32	0.70	0.98	0.00	0.00	0.01
	LASSO											
20	74.12	59.84	56.58	5.89	6.73	7.41	0.67	0.83	0.96	0.16	0.17	0.18
40	68.98	76.56	58.74	7.44	8.15	8.75	0.62	0.81	0.95	0.12	0.12	0.12
100	79.16	70.68	59.90	10.54	10.35	10.93	0.57	0.76	0.94	0.08	0.07	0.07
	A-LASSO											
20	74.62	59.65	55.89	4.43	5.09	5.90	0.55	0.71	0.89	0.11	0.11	0.12
40	69.47	75.51	57.88	5.81	6.51	7.26	0.53	0.71	0.90	0.09	0.09	0.09
100	77.14	70.04	59.19	8.56	8.82	9.74	0.50	0.70	0.91	0.07	0.06	0.06
	Boosting											
20	80.02	60.00	57.00	8.72	9.22	9.59	0.76	0.87	0.97	0.28	0.29	0.29
40	83.18	82.28	65.19	16.07	16.41	16.75	0.76	0.87	0.97	0.33	0.32	0.32
100	119.04	101.18	79.09	36.51	37.74	38.42	0.74	0.86	0.97	0.34	0.34	0.35
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	69.34	57.56	57.13	1.65	3.30	6.98	0.34	0.56	0.77	0.01	0.05	0.20
40	61.40	72.29	61.59	1.54	3.97	10.96	0.29	0.52	0.75	0.01	0.05	0.20
100	63.83	65.01	73.43	1.53	5.56	23.20	0.23	0.46	0.74	0.01	0.04	0.20
	LASSO											
20	70.69	57.46	55.40	6.04	6.28	6.56	0.59	0.64	0.68	0.18	0.19	0.19
40	64.66	71.49	56.25	8.91	9.20	9.12	0.56	0.62	0.65	0.17	0.17	0.16
100	69.73	65.98	59.74	17.11	18.49	17.73	0.52	0.60	0.64	0.15	0.16	0.15
	A-LASSO											
20	73.47	58.83	56.33	4.69	4.90	5.21	0.50	0.55	0.61	0.14	0.13	0.14
40	68.65	73.64	58.24	7.10	7.44	7.49	0.48	0.56	0.60	0.13	0.13	0.13
100	79.18	71.84	64.09	13.74	15.17	14.78	0.46	0.55	0.60	0.12	0.13	0.12
	Boosting											
20	80.36	64.67	63.84	10.79	13.00	15.21	0.77	0.86	0.93	0.39	0.48	0.57
40	77.73	89.87	68.46	20.52	24.03	26.57	0.77	0.85	0.89	0.44	0.52	0.58
100	89.89	81.83	75.23	40.39	44.35	46.51	0.72	0.79	0.83	0.38	0.41	0.43

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.7.

Table S.15: MC results for methods using heavy down-weighting in the baseline experiment with dynamics ($\rho_y \neq 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	71.35	58.74	57.44	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	64.90	72.62	55.19	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	66.62	65.52	59.08	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	71.83	58.47	58.24	1.97	3.62	5.22	0.46	0.81	0.99	0.01	0.02	0.06
40	62.95	73.94	56.43	1.65	3.41	5.09	0.38	0.78	0.99	0.00	0.01	0.03
100	64.46	64.55	60.12	1.36	3.02	4.83	0.32	0.70	0.98	0.00	0.00	0.01
	LASSO											
20	78.37	63.15	61.50	5.89	6.73	7.41	0.67	0.83	0.96	0.16	0.17	0.18
40	71.96	81.94	63.55	7.44	8.15	8.75	0.62	0.81	0.95	0.12	0.12	0.12
100	83.51	77.14	65.96	10.54	10.35	10.93	0.57	0.76	0.94	0.08	0.07	0.07
	A-LASSO											
20	78.24	62.58	59.71	4.43	5.09	5.90	0.55	0.71	0.89	0.11	0.11	0.12
40	71.83	79.75	62.49	5.81	6.51	7.26	0.53	0.71	0.90	0.09	0.09	0.09
100	80.42	75.58	64.78	8.56	8.82	9.74	0.50	0.70	0.91	0.07	0.06	0.06
	Boosting											
20	83.95	65.15	62.34	8.72	9.22	9.59	0.76	0.87	0.97	0.28	0.29	0.29
40	89.75	88.32	74.92	16.07	16.41	16.75	0.76	0.87	0.97	0.33	0.32	0.32
100	131.63	122.52	106.79	36.51	37.74	38.42	0.74	0.86	0.97	0.34	0.34	0.35
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	74.05	64.20	69.45	2.40	5.32	10.33	0.34	0.56	0.80	0.05	0.15	0.36
40	65.61	85.19	90.87	3.06	8.68	19.55	0.29	0.54	0.80	0.05	0.16	0.41
100	69.13	77.90	133.63	4.87	18.16	50.30	0.25	0.50	0.83	0.04	0.16	0.47
	LASSO											
20	75.06	61.29	59.50	6.67	6.73	6.78	0.55	0.58	0.60	0.22	0.22	0.22
40	72.01	80.21	63.12	12.08	12.41	11.60	0.55	0.60	0.60	0.25	0.25	0.23
100	76.58	74.23	69.51	22.80	28.89	27.97	0.51	0.61	0.63	0.21	0.26	0.25
	A-LASSO											
20	78.34	63.57	61.19	5.21	5.24	5.33	0.46	0.49	0.53	0.17	0.16	0.16
40	76.23	81.69	66.63	9.64	9.91	9.34	0.47	0.52	0.53	0.19	0.20	0.18
100	85.00	80.34	74.77	18.02	22.61	21.95	0.45	0.54	0.56	0.16	0.20	0.20
	Boosting											
20	92.65	76.90	74.73	12.71	14.50	15.69	0.80	0.87	0.92	0.48	0.55	0.60
40	88.46	101.02	79.38	22.75	24.76	25.81	0.77	0.81	0.83	0.49	0.54	0.56
100	95.90	89.98	86.33	40.39	42.46	43.07	0.67	0.72	0.74	0.38	0.40	0.40

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.7.

Table S.16: MC results for methods using no down-weighting in the baseline experiment with dynamics ($\rho_y \neq 0$) and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	19.13	16.07	15.36	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	17.53	19.75	15.32	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	17.93	17.75	15.34	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	19.98	16.19	15.48	4.08	5.16	6.59	0.89	0.99	1.00	0.03	0.06	0.13
40	17.89	20.09	15.49	3.85	5.02	6.40	0.86	0.99	1.00	0.01	0.03	0.06
100	18.27	17.92	15.40	3.55	4.77	6.11	0.81	0.98	1.00	0.00	0.01	0.02
	LASSO											
20	20.33	16.43	15.51	7.32	7.65	7.70	0.90	0.97	1.00	0.19	0.19	0.19
40	18.59	20.33	15.54	9.19	9.26	9.12	0.88	0.97	1.00	0.14	0.13	0.13
100	19.45	18.40	15.60	12.44	11.68	11.34	0.84	0.96	1.00	0.09	0.08	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	21.68	17.03	15.72	7.32	7.65	7.70	0.90	0.97	1.00	0.19	0.19	0.19
40	20.54	21.34	16.10	9.19	9.26	9.12	0.88	0.97	1.00	0.14	0.13	0.13
100	23.24	20.73	16.22	12.44	11.68	11.34	0.84	0.96	1.00	0.09	0.08	0.07
	A-LASSO											
20	21.10	16.91	15.57	5.66	6.11	6.31	0.80	0.93	0.99	0.12	0.12	0.12
40	19.96	20.92	15.90	7.31	7.69	7.72	0.80	0.94	1.00	0.10	0.10	0.09
100	22.45	20.36	16.05	10.29	10.15	10.19	0.79	0.93	1.00	0.07	0.06	0.06
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	21.52	17.05	15.65	5.66	6.11	6.31	0.80	0.93	0.99	0.12	0.12	0.12
40	20.51	21.20	16.05	7.31	7.69	7.72	0.80	0.94	1.00	0.10	0.10	0.09
100	23.17	20.72	16.18	10.29	10.15	10.19	0.79	0.93	1.00	0.07	0.06	0.06
	Boosting											
20	22.92	17.78	16.64	9.45	9.76	9.79	0.93	0.98	1.00	0.29	0.29	0.29
40	21.96	22.92	16.77	16.53	16.85	16.94	0.92	0.98	1.00	0.32	0.32	0.32
100	25.61	22.33	17.86	36.46	37.66	38.24	0.91	0.98	1.00	0.33	0.34	0.34
	Boosting for variable selection only. LS for estimation/forecasting.											
20	22.96	17.08	15.75	9.45	9.76	9.79	0.93	0.98	1.00	0.29	0.29	0.29
40	23.66	22.88	16.72	16.53	16.85	16.94	0.92	0.98	1.00	0.32	0.32	0.32
100	34.98	27.57	18.32	36.46	37.66	38.24	0.91	0.98	1.00	0.33	0.34	0.34

Notes: See notes to Table S.7.

Table S.17: MC results for methods using light down-weighting in the baseline experiment with dynamics ($\rho_y \neq 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	19.90	16.56	16.01	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	18.13	20.57	15.56	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	18.50	18.15	16.27	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	20.67	16.78	16.25	4.08	5.16	6.59	0.89	0.99	1.00	0.03	0.06	0.13
40	18.36	21.24	16.03	3.85	5.02	6.40	0.86	0.99	1.00	0.01	0.03	0.06
100	18.43	18.39	16.50	3.55	4.77	6.11	0.81	0.98	1.00	0.00	0.01	0.02
	LASSO											
20	22.71	17.99	16.79	7.32	7.65	7.70	0.90	0.97	1.00	0.19	0.19	0.19
40	21.07	22.97	17.35	9.19	9.26	9.12	0.88	0.97	1.00	0.14	0.13	0.13
100	24.18	21.74	17.62	12.44	11.68	11.34	0.84	0.96	1.00	0.09	0.08	0.07
	A-LASSO											
20	22.12	17.91	16.39	5.66	6.11	6.31	0.80	0.93	0.99	0.12	0.12	0.12
40	20.81	22.51	17.03	7.31	7.69	7.72	0.80	0.94	1.00	0.10	0.10	0.09
100	23.95	21.50	17.43	10.29	10.15	10.19	0.79	0.93	1.00	0.07	0.06	0.06
	Boosting											
20	24.12	18.17	16.88	9.45	9.76	9.79	0.93	0.98	1.00	0.29	0.29	0.29
40	24.61	24.80	19.11	16.53	16.85	16.94	0.92	0.98	1.00	0.32	0.32	0.32
100	36.33	29.92	23.55	36.46	37.66	38.24	0.91	0.98	1.00	0.33	0.34	0.34
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	20.85	17.07	16.88	3.27	4.64	7.99	0.70	0.81	0.90	0.02	0.07	0.22
40	18.72	21.57	18.44	3.19	5.39	12.03	0.66	0.80	0.90	0.01	0.06	0.21
100	19.22	19.36	22.22	3.13	7.04	24.30	0.61	0.76	0.90	0.01	0.04	0.21
	LASSO											
20	21.64	17.24	16.58	7.76	8.06	8.28	0.84	0.88	0.90	0.22	0.23	0.23
40	19.88	21.84	17.02	11.12	11.37	11.38	0.83	0.87	0.88	0.20	0.20	0.20
100	21.24	20.24	18.32	19.40	21.57	20.92	0.80	0.87	0.88	0.16	0.18	0.17
	A-LASSO											
20	22.47	17.57	16.92	6.11	6.43	6.66	0.75	0.82	0.85	0.15	0.16	0.16
40	21.10	22.71	17.63	8.96	9.32	9.41	0.76	0.83	0.85	0.15	0.15	0.15
100	24.07	22.26	19.83	15.75	17.83	17.52	0.75	0.83	0.85	0.13	0.15	0.14
	Boosting											
20	25.84	20.75	20.30	11.38	13.44	15.42	0.91	0.96	0.98	0.39	0.48	0.57
40	25.22	28.30	21.89	20.94	24.36	26.73	0.91	0.95	0.96	0.43	0.51	0.57
100	28.63	26.03	24.25	40.51	44.31	46.32	0.89	0.93	0.95	0.37	0.41	0.43

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.7.

Table S.18: MC results for methods using heavy down-weighting in the baseline experiment with dynamics ($\rho_y \neq 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	21.02	17.31	16.91	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	19.11	21.44	16.22	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	19.55	19.24	17.34	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	21.81	17.75	17.46	4.08	5.16	6.59	0.89	0.99	1.00	0.03	0.06	0.13
40	19.23	22.78	17.18	3.85	5.02	6.40	0.86	0.99	1.00	0.01	0.03	0.06
100	19.08	19.72	18.05	3.55	4.77	6.11	0.81	0.98	1.00	0.00	0.01	0.02
	LASSO											
20	24.13	19.34	18.24	7.32	7.65	7.70	0.90	0.97	1.00	0.19	0.19	0.19
40	22.32	24.87	18.92	9.19	9.26	9.12	0.88	0.97	1.00	0.14	0.13	0.13
100	25.95	23.98	19.47	12.44	11.68	11.34	0.84	0.96	1.00	0.09	0.08	0.07
	A-LASSO											
20	23.15	18.99	17.52	5.66	6.11	6.31	0.80	0.93	0.99	0.12	0.12	0.12
40	21.81	23.88	18.47	7.31	7.69	7.72	0.80	0.94	1.00	0.10	0.10	0.09
100	25.24	23.33	19.18	10.29	10.15	10.19	0.79	0.93	1.00	0.07	0.06	0.06
	Boosting											
20	25.57	19.88	18.55	9.45	9.76	9.79	0.93	0.98	1.00	0.29	0.29	0.29
40	27.10	27.60	22.04	16.53	16.85	16.94	0.92	0.98	1.00	0.32	0.32	0.32
100	39.99	36.20	32.23	36.46	37.66	38.24	0.91	0.98	1.00	0.33	0.34	0.34
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	22.57	19.07	20.76	3.69	6.40	11.11	0.62	0.76	0.89	0.06	0.17	0.38
40	20.45	25.88	27.35	4.38	9.87	20.38	0.58	0.75	0.90	0.05	0.17	0.42
100	21.70	23.90	41.37	6.09	19.41	51.02	0.53	0.73	0.92	0.04	0.16	0.47
	LASSO											
20	23.24	18.68	18.30	8.39	8.49	8.51	0.79	0.81	0.82	0.26	0.26	0.26
40	22.18	24.84	19.63	14.19	14.43	13.83	0.79	0.82	0.81	0.28	0.28	0.26
100	23.65	22.67	21.63	24.56	31.42	30.67	0.76	0.83	0.84	0.22	0.28	0.27
	A-LASSO											
20	24.08	19.29	18.87	6.57	6.69	6.74	0.70	0.73	0.75	0.19	0.19	0.19
40	23.61	25.51	20.74	11.39	11.64	11.18	0.71	0.75	0.76	0.21	0.22	0.20
100	26.29	24.63	23.40	19.50	24.67	24.15	0.70	0.78	0.79	0.17	0.22	0.21
	Boosting											
20	29.88	24.77	23.92	13.13	14.82	15.86	0.91	0.95	0.96	0.48	0.55	0.60
40	28.64	32.01	25.59	23.09	25.05	25.99	0.89	0.92	0.92	0.49	0.53	0.56
100	30.72	28.87	27.87	40.49	42.48	43.06	0.85	0.87	0.88	0.37	0.39	0.40

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.7.

S-3.2 MC Findings for experiments with parameter instabilities

Table S.19: MC results for methods using no down-weighting in the experiment with parameter instabilities, no dynamics ($\rho_y = 0$) and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$n \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	26.22	22.07	21.40	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	23.89	25.39	20.79	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	24.59	24.31	21.71	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	27.27	22.27	21.51	3.60	4.89	6.16	0.77	0.96	1.00	0.03	0.05	0.11
40	24.95	25.57	21.10	3.44	4.78	6.08	0.74	0.96	1.00	0.01	0.02	0.05
100	25.67	24.79	21.92	3.31	4.62	5.91	0.69	0.94	1.00	0.01	0.01	0.02
	LASSO											
20	28.02	22.91	21.91	7.41	8.24	8.97	0.80	0.91	0.99	0.21	0.23	0.25
40	25.76	26.30	21.61	9.47	11.21	12.29	0.78	0.91	0.99	0.16	0.19	0.21
100	26.81	26.23	22.40	12.81	15.10	18.15	0.73	0.88	0.98	0.10	0.12	0.14
	LASSO for variable selection only. LS for estimation/forecasting.											
20	29.22	23.42	22.02	7.41	8.24	8.97	0.80	0.91	0.99	0.21	0.23	0.25
40	27.84	27.74	22.14	9.47	11.21	12.29	0.78	0.91	0.99	0.16	0.19	0.21
100	30.99	29.03	23.51	12.81	15.10	18.15	0.73	0.88	0.98	0.10	0.12	0.14
	A-LASSO											
20	28.89	23.26	21.82	5.67	6.46	7.27	0.69	0.83	0.96	0.15	0.16	0.17
40	27.20	27.32	21.77	7.41	9.04	10.18	0.69	0.85	0.97	0.12	0.14	0.16
100	29.56	28.30	23.07	10.24	12.47	15.45	0.67	0.84	0.97	0.08	0.09	0.12
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	29.22	23.47	21.89	5.67	6.46	7.27	0.69	0.83	0.96	0.15	0.16	0.17
40	27.94	27.82	22.06	7.41	9.04	10.18	0.69	0.85	0.97	0.12	0.14	0.16
100	30.55	28.83	23.35	10.24	12.47	15.45	0.67	0.84	0.97	0.08	0.09	0.12
	Boosting											
20	28.34	23.04	21.91	9.56	10.27	10.87	0.85	0.94	0.99	0.31	0.33	0.35
40	26.89	27.07	21.67	16.71	18.27	19.17	0.85	0.94	0.99	0.33	0.36	0.38
100	29.37	27.98	22.78	35.75	38.90	42.31	0.84	0.93	0.99	0.32	0.35	0.38
	Boosting for variable selection only. LS for estimation/forecasting.											
20	30.42	23.84	22.17	9.56	10.27	10.87	0.85	0.94	0.99	0.31	0.33	0.35
40	31.76	30.33	22.82	16.71	18.27	19.17	0.85	0.94	0.99	0.33	0.36	0.38
100	45.11	37.24	26.01	35.75	38.90	42.31	0.84	0.93	0.99	0.32	0.35	0.38

Notes: This table reports one-step-ahead Mean Square Forecast Error (MSFE, $\times 100$), average number of selected variables ($\hat{k}$), True Positive Rate (TPR), and False Positive Rate (FPR). There are $k = 4$ signals variables out of N observed variables. The DGP is given by $y_t = d_t + \rho_{y,t} y_{t-1} + \sum_{j=1}^{4} \beta_{jt} x_{jt} + \tau_u u_t$, where slopes $\beta_{jt} = b_{jt} + \tau_{\eta_j} \eta_{jt}$ feature stochastic AR(1) component η_{jt} and parameter instabilities in mean b_{jt} given by (14)-(15), intercepts are given by $d_t = \sum_{j=1}^{k} \beta_{jt} \mu_{jt}$ where parameter instabilities in μ_{jt} is given by (16)-(17), and $\rho_{y,t}$ is zero in experiments without dynamics, and given by (18) in experiments with dynamics. u_t is given by a GARCH(1,1). See Section 5 of the paper for the detailed description of the Monte Carlo design. The reported results are based on 2000 simulations. Oracle model assumes the identity of signal variables is known.

Table S.20: MC results for methods using light down-weighting in the experiment with parameter instabilities, no dynamics ($\rho_y = 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	25.40	20.77	19.74	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	23.48	25.04	19.67	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	24.02	22.96	20.16	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	26.57	21.21	20.10	3.60	4.89	6.16	0.77	0.96	1.00	0.03	0.05	0.11
40	24.52	25.64	20.21	3.44	4.78	6.08	0.74	0.96	1.00	0.01	0.02	0.05
100	25.14	23.51	20.37	3.31	4.62	5.91	0.69	0.94	1.00	0.01	0.01	0.02
	LASSO											
20	28.85	22.94	20.77	7.41	8.24	8.97	0.80	0.91	0.99	0.21	0.23	0.25
40	27.70	29.01	22.96	9.47	11.21	12.29	0.78	0.91	0.99	0.16	0.19	0.21
100	31.65	29.71	23.70	12.81	15.10	18.15	0.73	0.88	0.98	0.10	0.12	0.14
	A-LASSO											
20	28.69	22.83	20.43	5.67	6.46	7.27	0.69	0.83	0.96	0.15	0.16	0.17
40	27.76	28.40	22.33	7.41	9.04	10.18	0.69	0.85	0.97	0.12	0.14	0.16
100	30.86	28.80	22.96	10.24	12.47	15.45	0.67	0.84	0.97	0.08	0.09	0.12
	Boosting											
20	30.20	23.25	21.11	9.56	10.27	10.87	0.85	0.94	0.99	0.31	0.33	0.35
40	32.30	32.03	24.82	16.71	18.27	19.17	0.85	0.94	0.99	0.33	0.36	0.38
100	47.27	41.29	31.17	35.75	38.90	42.31	0.84	0.93	0.99	0.32	0.35	0.38
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	27.14	21.69	21.50	3.55	5.87	9.84	0.64	0.81	0.92	0.05	0.13	0.31
40	25.19	27.59	24.60	3.96	7.89	16.03	0.60	0.80	0.92	0.04	0.12	0.31
100	25.75	26.45	30.79	5.16	12.74	33.88	0.55	0.76	0.92	0.03	0.10	0.30
	LASSO											
20	27.73	21.95	20.73	7.44	7.89	8.20	0.75	0.84	0.90	0.22	0.23	0.23
40	25.85	26.87	21.56	10.09	10.90	10.97	0.72	0.82	0.88	0.18	0.19	0.19
100	27.71	26.14	22.64	16.16	17.35	17.30	0.68	0.79	0.86	0.13	0.14	0.14
	A-LASSO											
20	28.20	22.11	20.89	5.78	6.23	6.56	0.66	0.77	0.84	0.16	0.16	0.16
40	27.20	28.14	22.05	8.03	8.84	9.03	0.65	0.77	0.84	0.14	0.14	0.14
100	30.58	28.59	24.38	12.95	14.29	14.53	0.62	0.74	0.83	0.10	0.11	0.11
	Boosting											
20	29.15	23.77	22.77	11.33	13.39	15.56	0.85	0.93	0.98	0.40	0.48	0.58
40	28.97	30.66	25.06	20.66	24.60	27.10	0.85	0.92	0.96	0.43	0.52	0.58
100	32.51	30.31	26.57	40.13	44.82	47.19	0.82	0.89	0.94	0.37	0.41	0.43

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.19.

Table S.21: MC results for methods using heavy down-weighting in the experiment with parameter instabilities, no dynamics ($\rho_y = 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	25.73	21.33	20.54	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	24.13	25.88	20.30	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	24.32	23.78	21.39	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	26.96	21.99	21.27	3.60	4.89	6.16	0.77	0.96	1.00	0.03	0.05	0.11
40	24.97	26.94	21.24	3.44	4.78	6.08	0.74	0.96	1.00	0.01	0.02	0.05
100	25.37	24.53	21.92	3.31	4.62	5.91	0.69	0.94	1.00	0.01	0.01	0.02
	LASSO											
20	29.36	24.50	22.54	7.41	8.24	8.97	0.80	0.91	0.99	0.21	0.23	0.25
40	28.65	31.33	26.35	9.47	11.21	12.29	0.78	0.91	0.99	0.16	0.19	0.21
100	33.43	33.27	28.56	12.81	15.10	18.15	0.73	0.88	0.98	0.10	0.12	0.14
	A-LASSO											
20	29.03	24.10	21.71	5.67	6.46	7.27	0.69	0.83	0.96	0.15	0.16	0.17
40	28.41	29.95	25.16	7.41	9.04	10.18	0.69	0.85	0.97	0.12	0.14	0.16
100	32.16	31.44	26.90	10.24	12.47	15.45	0.67	0.84	0.97	0.08	0.09	0.12
	Boosting											
20	30.87	25.11	23.05	9.56	10.27	10.87	0.85	0.94	0.99	0.31	0.33	0.35
40	34.93	35.80	30.47	16.71	18.27	19.17	0.85	0.94	0.99	0.33	0.36	0.38
100	53.08	50.82	45.70	35.75	38.90	42.31	0.84	0.93	0.99	0.32	0.35	0.38
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	28.42	24.75	26.11	4.76	8.14	12.61	0.61	0.79	0.91	0.12	0.25	0.45
40	27.52	34.03	37.65	6.62	13.39	23.24	0.58	0.79	0.92	0.11	0.26	0.49
100	30.86	38.49	60.77	11.43	27.74	56.86	0.53	0.76	0.94	0.09	0.25	0.53
	LASSO											
20	28.54	22.94	21.99	7.70	7.87	8.09	0.72	0.78	0.81	0.24	0.24	0.24
40	27.45	29.40	23.55	12.08	12.51	12.06	0.70	0.77	0.79	0.23	0.24	0.22
100	30.65	28.09	25.98	21.27	23.44	22.90	0.66	0.75	0.79	0.19	0.20	0.20
	A-LASSO											
20	29.33	23.67	22.61	5.98	6.15	6.41	0.63	0.70	0.75	0.17	0.17	0.17
40	29.45	31.07	24.55	9.60	10.00	9.73	0.62	0.71	0.74	0.18	0.18	0.17
100	33.67	30.61	28.26	16.54	18.45	18.17	0.59	0.70	0.75	0.14	0.16	0.15
	Boosting											
20	32.30	27.95	26.01	13.04	14.80	16.03	0.87	0.93	0.96	0.48	0.55	0.61
40	32.20	34.57	29.07	22.92	25.20	26.34	0.85	0.90	0.93	0.49	0.54	0.57
100	34.71	33.34	30.36	40.27	42.73	43.68	0.78	0.85	0.89	0.37	0.39	0.40

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.19.

Table S.22: MC results for methods using no down-weighting in the experiment with parameter instabilities, no dynamics ($\rho_y = 0$) and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	9.45	7.97	7.78	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	8.64	9.06	7.48	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	8.84	8.71	8.03	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	9.58	8.15	7.83	4.87	5.87	7.21	0.95	1.00	1.00	0.05	0.09	0.16
40	8.81	9.18	7.64	4.81	5.84	7.18	0.94	1.00	1.00	0.03	0.05	0.08
100	9.15	9.04	8.17	4.94	5.89	7.23	0.92	0.99	1.00	0.01	0.02	0.03
	LASSO											
20	10.43	8.39	8.04	9.38	10.10	10.97	0.92	0.97	1.00	0.29	0.31	0.35
40	9.70	9.48	7.79	12.75	15.23	16.92	0.91	0.97	1.00	0.23	0.28	0.32
100	9.81	9.49	8.36	18.47	22.95	29.96	0.89	0.97	1.00	0.15	0.19	0.26
	LASSO for variable selection only. LS for estimation/forecasting.											
20	10.97	8.58	8.08	9.38	10.10	10.97	0.92	0.97	1.00	0.29	0.31	0.35
40	10.77	10.18	8.10	12.75	15.23	16.92	0.91	0.97	1.00	0.23	0.28	0.32
100	11.50	11.15	9.07	18.47	22.95	29.96	0.89	0.97	1.00	0.15	0.19	0.26
	A-LASSO											
20	10.67	8.40	8.00	7.33	8.04	8.77	0.84	0.94	0.99	0.20	0.21	0.24
40	10.34	9.94	7.82	10.04	12.23	13.57	0.85	0.95	0.99	0.17	0.21	0.24
100	10.85	10.58	8.77	14.60	18.58	24.51	0.84	0.95	0.99	0.11	0.15	0.21
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	10.92	8.51	8.06	7.33	8.04	8.77	0.84	0.94	0.99	0.20	0.21	0.24
40	10.84	10.16	8.02	10.04	12.23	13.57	0.85	0.95	0.99	0.17	0.21	0.24
100	11.39	11.00	9.05	14.60	18.58	24.51	0.84	0.95	0.99	0.11	0.15	0.21
	Boosting											
20	10.46	8.41	8.08	10.80	11.48	12.00	0.94	0.98	1.00	0.35	0.38	0.40
40	10.14	9.70	7.86	18.77	20.49	21.64	0.94	0.98	1.00	0.38	0.41	0.44
100	10.52	9.93	8.43	38.21	42.29	46.34	0.93	0.98	1.00	0.34	0.38	0.42
	Boosting for variable selection only. LS for estimation/forecasting.											
20	11.13	8.78	8.11	10.80	11.48	12.00	0.94	0.98	1.00	0.35	0.38	0.40
40	11.72	10.80	8.33	18.77	20.49	21.64	0.94	0.98	1.00	0.38	0.41	0.44
100	16.59	13.42	9.82	38.21	42.29	46.34	0.93	0.98	1.00	0.34	0.38	0.42

Notes: See notes to Table S.19.

Table S.23: MC results for methods using light down-weighting in the experiment with parameter instabilities, no dynamics ($\rho_y = 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	8.32	6.36	5.89	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	7.66	7.63	5.87	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	7.93	7.06	6.03	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	8.55	6.59	6.05	4.87	5.87	7.21	0.95	1.00	1.00	0.05	0.09	0.16
40	7.92	7.96	6.21	4.81	5.84	7.18	0.94	1.00	1.00	0.03	0.05	0.08
100	8.42	7.48	6.20	4.94	5.89	7.23	0.92	0.99	1.00	0.01	0.02	0.03
	LASSO											
20	10.15	7.25	6.27	9.38	10.10	10.97	0.92	0.97	1.00	0.29	0.31	0.35
40	10.21	9.44	7.24	12.75	15.23	16.92	0.91	0.97	1.00	0.23	0.28	0.32
100	11.52	10.48	8.44	18.47	22.95	29.96	0.89	0.97	1.00	0.15	0.19	0.26
	A-LASSO											
20	10.14	7.20	6.19	7.33	8.04	8.77	0.84	0.94	0.99	0.20	0.21	0.24
40	10.20	9.20	6.83	10.04	12.23	13.57	0.85	0.95	0.99	0.17	0.21	0.24
100	11.19	10.25	8.00	14.60	18.58	24.51	0.84	0.95	0.99	0.11	0.15	0.21
	Boosting											
20	10.37	7.37	6.33	10.80	11.48	12.00	0.94	0.98	1.00	0.35	0.38	0.40
40	11.32	10.23	7.66	18.77	20.49	21.64	0.94	0.98	1.00	0.38	0.41	0.44
100	17.03	13.79	9.79	38.21	42.29	46.34	0.93	0.98	1.00	0.34	0.38	0.42
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	9.08	6.86	6.45	4.90	7.16	11.03	0.82	0.90	0.96	0.08	0.18	0.36
40	8.67	8.88	7.45	5.73	9.85	18.14	0.80	0.90	0.96	0.06	0.16	0.36
100	9.03	8.97	9.58	7.86	16.24	37.95	0.77	0.88	0.96	0.05	0.13	0.34
	LASSO											
20	9.58	6.91	6.30	9.39	9.64	9.50	0.90	0.96	0.98	0.29	0.29	0.28
40	9.09	8.62	6.59	13.42	13.95	13.51	0.89	0.95	0.98	0.25	0.25	0.24
100	9.64	8.57	7.07	21.04	22.73	22.37	0.86	0.94	0.97	0.18	0.19	0.18
	A-LASSO											
20	9.63	6.83	6.29	7.39	7.73	7.57	0.83	0.92	0.96	0.20	0.20	0.19
40	9.56	8.89	6.68	10.71	11.28	11.04	0.83	0.92	0.97	0.18	0.19	0.18
100	10.62	9.33	7.58	16.71	18.52	18.55	0.82	0.91	0.97	0.13	0.15	0.15
	Boosting											
20	9.92	7.70	7.11	12.18	13.93	15.85	0.93	0.98	1.00	0.42	0.50	0.59
40	10.20	9.64	7.84	21.78	25.40	27.53	0.94	0.98	0.99	0.45	0.54	0.59
100	10.99	9.80	8.25	41.38	45.75	47.72	0.92	0.97	0.99	0.38	0.42	0.44

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.19.

Table S.24: MC results for methods using heavy down-weighting in the experiment with parameter instabilities, no dynamics ($\rho_y = 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	7.91	6.32	6.10	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	7.40	7.65	6.03	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	7.57	7.09	6.34	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	8.22	6.64	6.43	4.87	5.87	7.21	0.95	1.00	1.00	0.05	0.09	0.16
40	7.72	8.23	6.65	4.81	5.84	7.18	0.94	1.00	1.00	0.03	0.05	0.08
100	8.21	7.58	6.80	4.94	5.89	7.23	0.92	0.99	1.00	0.01	0.02	0.03
	LASSO											
20	9.89	7.60	6.91	9.38	10.10	10.97	0.92	0.97	1.00	0.29	0.31	0.35
40	10.24	9.92	9.00	12.75	15.23	16.92	0.91	0.97	1.00	0.23	0.28	0.32
100	12.37	11.55	11.10	18.47	22.95	29.96	0.89	0.97	1.00	0.15	0.19	0.26
	A-LASSO											
20	9.90	7.39	6.61	7.33	8.04	8.77	0.84	0.94	0.99	0.20	0.21	0.24
40	10.04	9.46	7.93	10.04	12.23	13.57	0.85	0.95	0.99	0.17	0.21	0.24
100	11.57	10.91	10.03	14.60	18.58	24.51	0.84	0.95	0.99	0.11	0.15	0.21
	Boosting											
20	10.20	7.76	6.96	10.80	11.48	12.00	0.94	0.98	1.00	0.35	0.38	0.40
40	11.65	11.19	9.66	18.77	20.49	21.64	0.94	0.98	1.00	0.38	0.41	0.44
100	18.88	16.42	14.32	38.21	42.29	46.34	0.93	0.98	1.00	0.34	0.38	0.42
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	9.45	7.68	7.75	6.13	9.41	13.54	0.77	0.88	0.95	0.15	0.30	0.49
40	9.45	10.73	11.24	8.65	15.43	24.74	0.75	0.88	0.96	0.14	0.30	0.52
100	11.01	13.30	18.51	14.90	31.53	59.32	0.72	0.87	0.97	0.12	0.28	0.55
	LASSO											
20	9.43	7.14	6.75	9.60	9.72	9.68	0.88	0.93	0.95	0.30	0.30	0.29
40	9.34	9.18	7.44	14.94	15.48	14.97	0.88	0.93	0.94	0.29	0.29	0.28
100	10.26	9.13	8.22	24.43	27.51	27.76	0.84	0.92	0.94	0.21	0.24	0.24
	A-LASSO											
20	9.52	7.24	6.84	7.52	7.72	7.68	0.81	0.89	0.91	0.21	0.21	0.20
40	9.95	9.59	7.70	11.90	12.41	12.03	0.82	0.89	0.92	0.22	0.22	0.21
100	11.25	9.93	8.93	18.88	21.51	21.81	0.79	0.89	0.93	0.16	0.18	0.18
	Boosting											
20	10.54	9.01	8.24	13.62	15.18	16.29	0.94	0.98	0.99	0.49	0.56	0.62
40	11.06	10.80	9.23	23.63	25.76	26.79	0.94	0.97	0.98	0.50	0.55	0.57
100	11.74	10.77	9.64	41.11	43.39	44.34	0.90	0.95	0.97	0.37	0.40	0.40

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.19.

Table S.25: MC results for methods using no down-weighting in the experiment with parameter instabilities, dynamics ($\rho_y \neq 0$) and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	71.12	58.20	56.40	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	63.61	70.96	55.02	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	66.03	64.88	57.06	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	73.07	58.80	56.68	1.42	2.98	4.70	0.33	0.69	0.97	0.00	0.01	0.04
40	66.35	72.43	55.54	1.21	2.74	4.57	0.28	0.64	0.97	0.00	0.00	0.02
100	68.44	67.14	56.98	0.95	2.35	4.34	0.22	0.56	0.95	0.00	0.00	0.01
	LASSO											
20	73.84	59.94	57.08	5.93	6.89	7.78	0.60	0.76	0.92	0.18	0.19	0.21
40	66.91	72.86	56.78	7.97	8.83	9.74	0.55	0.74	0.91	0.14	0.15	0.15
100	68.42	67.93	58.04	11.53	11.92	13.20	0.51	0.68	0.88	0.10	0.09	0.10
	LASSO for variable selection only. LS for estimation/forecasting.											
20	78.11	61.46	57.65	5.93	6.89	7.78	0.60	0.76	0.92	0.18	0.19	0.21
40	73.10	76.36	58.19	7.97	8.83	9.74	0.55	0.74	0.91	0.14	0.15	0.15
100	81.43	75.12	59.74	11.53	11.92	13.20	0.51	0.68	0.88	0.10	0.09	0.10
	A-LASSO											
20	76.99	61.24	57.30	4.44	5.26	6.17	0.49	0.65	0.84	0.12	0.13	0.14
40	71.31	74.88	57.13	6.27	7.06	8.01	0.47	0.65	0.84	0.11	0.11	0.12
100	76.78	73.62	59.39	9.23	10.01	11.54	0.45	0.63	0.85	0.07	0.08	0.08
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	78.37	61.85	57.46	4.44	5.26	6.17	0.49	0.65	0.84	0.12	0.13	0.14
40	72.62	76.09	57.54	6.27	7.06	8.01	0.47	0.65	0.84	0.11	0.11	0.12
100	79.39	75.11	59.88	9.23	10.01	11.54	0.45	0.63	0.85	0.07	0.08	0.08
	Boosting											
20	75.75	59.22	56.60	8.88	9.45	10.07	0.72	0.83	0.94	0.30	0.31	0.32
40	72.17	75.07	56.44	16.72	17.37	17.96	0.72	0.83	0.94	0.35	0.35	0.36
100	81.53	74.23	59.21	37.31	39.47	41.08	0.70	0.81	0.94	0.34	0.36	0.37
	Boosting for variable selection only. LS for estimation/forecasting.											
20	82.82	62.55	58.10	8.88	9.45	10.07	0.72	0.83	0.94	0.30	0.31	0.32
40	89.34	82.31	60.27	16.72	17.37	17.96	0.72	0.83	0.94	0.35	0.35	0.36
100	121.55	99.11	67.65	37.31	39.47	41.08	0.70	0.81	0.94	0.34	0.36	0.37

Notes: See notes to Table S.19.

Table S.26: MC results for methods using light down-weighting in the experiment with parameter instabilities, dynamics ($\rho_y \neq 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	70.52	56.93	54.94	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	63.14	70.04	53.63	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	65.81	62.90	55.96	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	72.41	57.50	55.25	1.42	2.98	4.70	0.33	0.69	0.97	0.00	0.01	0.04
40	64.73	72.12	53.91	1.21	2.74	4.57	0.28	0.64	0.97	0.00	0.00	0.02
100	67.42	65.12	56.39	0.95	2.35	4.34	0.22	0.56	0.95	0.00	0.00	0.01
	LASSO											
20	78.05	62.09	57.07	5.93	6.89	7.78	0.60	0.76	0.92	0.18	0.19	0.21
40	72.33	78.90	61.20	7.97	8.83	9.74	0.55	0.74	0.91	0.14	0.15	0.15
100	83.09	76.07	60.36	11.53	11.92	13.20	0.51	0.68	0.88	0.10	0.09	0.10
	A-LASSO											
20	77.67	62.15	56.48	4.44	5.26	6.17	0.49	0.65	0.84	0.12	0.13	0.14
40	71.81	77.83	59.15	6.27	7.06	8.01	0.47	0.65	0.84	0.11	0.11	0.12
100	79.93	74.82	59.36	9.23	10.01	11.54	0.45	0.63	0.85	0.07	0.08	0.08
	Boosting											
20	83.76	62.29	58.20	8.88	9.45	10.07	0.72	0.83	0.94	0.30	0.31	0.32
40	89.89	84.96	66.42	16.72	17.37	17.96	0.72	0.83	0.94	0.35	0.35	0.36
100	125.04	107.40	80.23	37.31	39.47	41.08	0.70	0.81	0.94	0.34	0.36	0.37
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	73.26	59.08	58.15	1.44	3.24	7.10	0.29	0.54	0.81	0.01	0.05	0.19
40	64.80	73.72	62.87	1.39	3.91	11.00	0.24	0.51	0.80	0.01	0.05	0.20
100	68.23	67.63	74.07	1.46	5.49	23.08	0.20	0.45	0.79	0.01	0.04	0.20
	LASSO											
20	75.31	59.36	56.52	6.25	6.67	7.18	0.57	0.66	0.74	0.20	0.20	0.21
40	67.42	73.14	57.89	9.60	9.98	10.05	0.53	0.63	0.71	0.19	0.19	0.18
100	72.69	68.91	61.31	18.55	20.22	19.46	0.50	0.61	0.70	0.17	0.18	0.17
	A-LASSO											
20	77.82	60.15	57.26	4.86	5.23	5.71	0.47	0.57	0.67	0.15	0.15	0.15
40	71.49	75.18	59.38	7.62	8.07	8.26	0.45	0.57	0.65	0.15	0.14	0.14
100	81.80	74.79	65.52	14.83	16.53	16.18	0.44	0.55	0.66	0.13	0.14	0.14
	Boosting											
20	83.69	66.02	65.35	10.96	13.10	15.35	0.74	0.86	0.94	0.40	0.48	0.58
40	82.56	90.80	70.49	20.91	24.35	26.80	0.74	0.84	0.90	0.45	0.52	0.58
100	93.38	84.61	77.12	40.85	44.87	47.02	0.69	0.78	0.85	0.38	0.42	0.44

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.19.

Table S.27: MC results for methods using heavy down-weighting in the experiment with parameter instabilities, dynamics ($\rho_y \neq 0$), and low fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	72.79	59.16	58.12	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	65.52	72.42	55.93	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	68.21	66.30	59.58	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	73.97	59.17	58.39	1.42	2.98	4.70	0.33	0.69	0.97	0.00	0.01	0.04
40	65.17	74.75	56.44	1.21	2.74	4.57	0.28	0.64	0.97	0.00	0.00	0.02
100	67.99	67.30	60.42	0.95	2.35	4.34	0.22	0.56	0.95	0.00	0.00	0.01
	LASSO											
20	80.87	65.88	61.66	5.93	6.89	7.78	0.60	0.76	0.92	0.18	0.19	0.21
40	74.94	84.03	67.45	7.97	8.83	9.74	0.55	0.74	0.91	0.14	0.15	0.15
100	87.60	82.88	67.29	11.53	11.92	13.20	0.51	0.68	0.88	0.10	0.09	0.10
	A-LASSO											
20	79.28	65.12	60.40	4.44	5.26	6.17	0.49	0.65	0.84	0.12	0.13	0.14
40	73.62	82.19	63.90	6.27	7.06	8.01	0.47	0.65	0.84	0.11	0.11	0.12
100	82.74	80.16	65.32	9.23	10.01	11.54	0.45	0.63	0.85	0.07	0.08	0.08
	Boosting											
20	86.81	66.63	64.18	8.88	9.45	10.07	0.72	0.83	0.94	0.30	0.31	0.32
40	95.81	93.48	79.66	16.72	17.37	17.96	0.72	0.83	0.94	0.35	0.35	0.36
100	137.08	132.36	110.70	37.31	39.47	41.08	0.70	0.81	0.94	0.34	0.36	0.37
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	76.61	65.05	70.36	2.36	5.34	10.42	0.33	0.57	0.83	0.05	0.15	0.36
40	68.62	89.48	91.60	3.06	8.73	19.57	0.28	0.56	0.84	0.05	0.16	0.41
100	73.35	84.91	135.48	4.89	18.08	50.16	0.24	0.52	0.86	0.04	0.16	0.47
	LASSO											
20	78.58	63.08	61.22	6.96	7.24	7.48	0.56	0.62	0.67	0.24	0.24	0.24
40	73.58	81.56	64.87	12.75	13.18	12.59	0.55	0.63	0.66	0.26	0.27	0.25
100	79.59	76.53	71.00	23.96	30.35	29.40	0.51	0.63	0.68	0.22	0.28	0.27
	A-LASSO											
20	81.67	64.87	62.93	5.43	5.63	5.92	0.47	0.52	0.59	0.18	0.18	0.18
40	77.83	83.55	68.45	10.15	10.48	10.14	0.47	0.55	0.59	0.21	0.21	0.19
100	88.18	83.33	76.36	18.88	23.75	23.15	0.44	0.56	0.62	0.17	0.22	0.21
	Boosting											
20	95.53	78.91	76.84	12.82	14.59	15.85	0.79	0.87	0.92	0.48	0.56	0.61
40	93.10	102.34	81.95	23.02	25.03	26.08	0.75	0.82	0.86	0.50	0.54	0.57
100	99.90	92.87	88.76	40.74	42.87	43.61	0.66	0.73	0.78	0.38	0.40	0.40

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.19.

Table S.28: MC results for methods using no down-weighting in the experiment with parameter instabilities, dynamics ($\rho_y \neq 0$) and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
	Oracle											
20	24.00	19.41	18.95	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	21.15	23.30	18.31	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	21.99	21.40	19.41	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	24.81	19.49	19.01	3.12	4.47	5.67	0.72	0.95	1.00	0.01	0.03	0.08
40	22.09	23.43	18.44	2.95	4.30	5.56	0.68	0.94	1.00	0.01	0.01	0.04
100	23.15	21.76	19.43	2.65	4.08	5.35	0.62	0.92	1.00	0.00	0.00	0.01
	LASSO											
20	25.72	20.13	19.29	7.69	8.33	9.09	0.80	0.91	0.98	0.22	0.23	0.26
40	22.85	24.45	19.02	10.43	11.58	12.36	0.77	0.90	0.98	0.18	0.20	0.21
100	23.89	23.40	19.93	15.34	16.47	17.96	0.74	0.88	0.97	0.12	0.13	0.14
	LASSO for variable selection only. LS for estimation/forecasting.											
20	27.02	20.43	19.45	7.69	8.33	9.09	0.80	0.91	0.98	0.22	0.23	0.26
40	25.01	25.78	19.50	10.43	11.58	12.36	0.77	0.90	0.98	0.18	0.20	0.21
100	28.70	26.47	20.79	15.34	16.47	17.96	0.74	0.88	0.97	0.12	0.13	0.14
	A-LASSO											
20	26.60	20.23	19.30	5.93	6.57	7.38	0.69	0.83	0.95	0.16	0.16	0.18
40	24.32	25.23	19.13	8.15	9.28	10.25	0.69	0.84	0.96	0.14	0.15	0.16
100	27.38	25.67	20.57	12.24	13.71	15.43	0.67	0.84	0.96	0.10	0.10	0.12
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	27.12	20.54	19.39	5.93	6.57	7.38	0.69	0.83	0.95	0.16	0.16	0.18
40	25.11	25.85	19.33	8.15	9.28	10.25	0.69	0.84	0.96	0.14	0.15	0.16
100	28.52	26.45	20.78	12.24	13.71	15.43	0.67	0.84	0.96	0.10	0.10	0.12
	Boosting											
20	26.38	20.05	19.09	9.80	10.35	10.87	0.85	0.93	0.98	0.32	0.33	0.35
40	25.53	25.18	18.88	17.81	18.76	19.25	0.86	0.93	0.99	0.36	0.38	0.38
100	28.53	25.15	20.30	38.06	40.82	43.14	0.84	0.93	0.98	0.35	0.37	0.39
	Boosting for variable selection only. LS for estimation/forecasting.											
20	28.08	21.12	19.65	9.80	10.35	10.87	0.85	0.93	0.98	0.32	0.33	0.35
40	29.29	26.99	20.08	17.81	18.76	19.25	0.86	0.93	0.99	0.36	0.38	0.38
100	41.52	33.76	23.16	38.06	40.82	43.14	0.84	0.93	0.98	0.35	0.37	0.39

Notes: See notes to Table S.19.

Table S.29: MC results for methods using light down-weighting in the experiment with parameter instabilities, dynamics ($\rho_y \neq 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Light down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	22.56	17.49	16.67	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	19.91	21.49	16.31	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	20.96	19.27	17.06	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	23.56	17.70	16.91	3.12	4.47	5.67	0.72	0.95	1.00	0.01	0.03	0.08
40	20.99	22.01	16.59	2.95	4.30	5.56	0.68	0.94	1.00	0.01	0.01	0.04
100	22.01	19.73	17.06	2.65	4.08	5.35	0.62	0.92	1.00	0.00	0.00	0.01
	LASSO											
20	26.19	19.18	17.56	7.69	8.33	9.09	0.80	0.91	0.98	0.22	0.23	0.26
40	24.13	25.30	19.13	10.43	11.58	12.36	0.77	0.90	0.98	0.18	0.20	0.21
100	28.73	25.71	19.47	15.34	16.47	17.96	0.74	0.88	0.97	0.12	0.13	0.14
	A-LASSO											
20	26.17	19.32	17.23	5.93	6.57	7.38	0.69	0.83	0.95	0.16	0.16	0.18
40	24.11	25.29	18.36	8.15	9.28	10.25	0.69	0.84	0.96	0.14	0.15	0.16
100	28.13	25.12	19.09	12.24	13.71	15.43	0.67	0.84	0.96	0.10	0.10	0.12
	Boosting											
20	27.07	19.48	17.95	9.80	10.35	10.87	0.85	0.93	0.98	0.32	0.33	0.35
40	28.63	26.09	20.36	17.81	18.76	19.25	0.86	0.93	0.99	0.36	0.38	0.38
100	42.00	35.22	25.62	38.06	40.82	43.14	0.84	0.93	0.98	0.35	0.37	0.39
B. Light down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	24.02	18.30	17.89	2.70	4.46	8.04	0.57	0.79	0.92	0.02	0.06	0.22
40	21.65	22.98	19.61	2.66	5.22	12.12	0.53	0.77	0.92	0.01	0.05	0.21
100	22.84	21.35	23.79	2.63	6.81	24.20	0.47	0.73	0.92	0.01	0.04	0.21
	LASSO											
20	25.30	18.80	17.47	7.92	8.43	8.74	0.76	0.87	0.92	0.24	0.25	0.25
40	22.67	23.39	18.29	11.97	12.37	12.39	0.75	0.85	0.91	0.22	0.22	0.22
100	24.68	22.46	19.48	21.49	23.42	22.67	0.71	0.82	0.90	0.19	0.20	0.19
	A-LASSO											
20	25.77	18.87	17.52	6.22	6.68	7.03	0.67	0.79	0.88	0.18	0.18	0.18
40	23.96	23.95	18.66	9.51	10.03	10.18	0.67	0.79	0.87	0.17	0.17	0.17
100	27.88	24.45	20.88	17.13	19.23	18.89	0.65	0.78	0.88	0.15	0.16	0.15
	Boosting											
20	28.83	22.29	21.92	11.58	13.52	15.62	0.85	0.94	0.98	0.41	0.49	0.58
40	29.18	30.14	23.81	21.48	24.84	27.09	0.86	0.93	0.97	0.45	0.53	0.58
100	32.82	28.94	26.13	41.24	45.22	47.15	0.82	0.90	0.96	0.38	0.42	0.43

Notes: Light down-weighting is defined by by values $\lambda = 0.975, 0.98, 0.985, 0.99, 0.995, 1$. For this set 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. See notes to Table S.19.

Table S.30: MC results for methods using heavy down-weighting in the experiment with parameter instabilities, dynamics ($\rho_y \neq 0$), and high fit.

	MSFE ($\times 100$)			$\hat{k}$			**TPR**			**FPR**		
$N \backslash T$	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**	**100**	**200**	**500**
A. Heavy down-weighting in the estimation/forecasting stage only.												
Variable selection is based on original (not down-weighted) data.												
Forecasting stage is Least Squares on selected down-weighted covariates for all methods												
	Oracle											
20	22.63	17.95	17.60	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	20.11	22.00	16.96	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	21.16	20.03	18.07	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	23.80	18.25	18.07	3.12	4.47	5.67	0.72	0.95	1.00	0.01	0.03	0.08
40	21.21	22.96	17.54	2.95	4.30	5.56	0.68	0.94	1.00	0.01	0.01	0.04
100	22.03	20.58	18.22	2.65	4.08	5.35	0.62	0.92	1.00	0.00	0.00	0.01
	LASSO											
20	26.58	20.54	19.16	7.69	8.33	9.09	0.80	0.91	0.98	0.22	0.23	0.26
40	24.86	27.18	21.73	10.43	11.58	12.36	0.77	0.90	0.98	0.18	0.20	0.21
100	29.96	28.74	22.70	15.34	16.47	17.96	0.74	0.88	0.97	0.12	0.13	0.14
	A-LASSO											
20	26.36	20.34	18.52	5.93	6.57	7.38	0.69	0.83	0.95	0.16	0.16	0.18
40	24.35	26.76	20.27	8.15	9.28	10.25	0.69	0.84	0.96	0.14	0.15	0.16
100	28.82	26.94	21.71	12.24	13.71	15.43	0.67	0.84	0.96	0.10	0.10	0.12
	Boosting											
20	27.41	20.74	19.96	9.80	10.35	10.87	0.85	0.93	0.98	0.32	0.33	0.35
40	30.45	28.48	24.59	17.81	18.76	19.25	0.86	0.93	0.99	0.36	0.38	0.38
100	45.93	41.56	35.80	38.06	40.82	43.14	0.84	0.93	0.98	0.35	0.37	0.39
B. Heavy down-weighting in both the variable selection and estimation/forecasting stages.												
OCMT uses down-weighted variables for selection as well as for forecasting using Least Squares.												
Remaining forecasts are based on Lasso, A-Lasso and Boosting regressions applied to down-weighted data.												
	OCMT											
20	24.68	20.32	21.63	3.48	6.45	11.17	0.57	0.78	0.92	0.06	0.17	0.37
40	23.01	26.11	28.93	4.22	9.97	20.43	0.53	0.77	0.93	0.05	0.17	0.42
100	25.04	26.90	43.11	5.93	19.43	50.99	0.48	0.75	0.94	0.04	0.16	0.47
	LASSO											
20	26.02	19.95	19.30	8.61	8.94	9.12	0.75	0.82	0.86	0.28	0.28	0.28
40	24.25	25.95	20.82	14.79	15.35	15.03	0.75	0.82	0.86	0.30	0.30	0.29
100	26.43	24.51	22.72	25.76	32.55	32.34	0.70	0.82	0.87	0.23	0.29	0.29
	A-LASSO											
20	26.54	20.46	19.71	6.75	7.01	7.26	0.66	0.74	0.80	0.20	0.20	0.20
40	25.83	26.71	21.76	11.79	12.30	12.13	0.67	0.76	0.81	0.23	0.23	0.22
100	29.49	26.52	24.51	20.25	25.56	25.44	0.64	0.77	0.83	0.18	0.22	0.22
	Boosting											
20	32.67	26.84	26.03	13.25	14.91	16.09	0.87	0.94	0.97	0.49	0.56	0.61
40	32.71	34.24	28.06	23.43	25.41	26.41	0.86	0.91	0.94	0.50	0.54	0.57
100	35.07	31.87	30.31	41.00	43.13	43.90	0.80	0.87	0.91	0.38	0.40	0.40

Notes: Heavy down-weighting is defined by by values $\lambda = 0.95, 0.96, 0.97, 0.98, 0.99, 1$. For this set 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. See notes to Table S.19.