Federal Reserve Bank of Dallas

Variable Selection in High Dimensional Linear Regressions with Parameter Instability

Alexander Chudik, M. Hashem Pesaran and Mahrad Sharifvaghefi

Globalization Institute Working Paper 394 **August 2020**

(Revised August 2024)

Research Department

https://doi.org/10.24149/gwp394r3

Variable Selection in High Dimensional Linear Regressions with Parameter Instability*

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

July 23, 2020
Revised: July 15, 2024

Abstract

This paper considers the problem of variable selection allowing for parameter instability. It distinguishes between signal and pseudo-signal variables that are correlated with the target variable, and noise variables that are not, and investigates the asymptotic properties of the One Covariate at a Time Multiple Testing (OCMT) method proposed by Chudik et al. (2018) under parameter insatiability. It is established that OCMT continues to asymptotically select an approximating model that includes all the signals and none of the noise variables. Properties of post selection regressions are also investigated, and in-sample fit of the selected regression is shown to have the oracle property. The theoretical results support the use of unweighted observations at the selection stage of OCMT, whilst applying down-weighting of observations only at the forecasting stage. Monte Carlo and empirical applications show that OCMT without down-weighting at the selection stage yields smaller mean squared forecast errors compared to Lasso, Adaptive Lasso, and boosting.

Keywords: Lasso, One Covariate at a time Multiple Testing (OCMT), Parameter instability, Variable selection, Forecasting

JEL Classifications: C22, C52, C53, C55

*We are grateful to Elie Tamer (Editor), two anonymous reviewers and an associate editor, for their constructive comments and helpful suggestions. We have also benefited from discussions and comments by George Kapetanios, Oliver Linton, Ron Smith, and seminar participants at Cambridge University. 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.

†Alexander Chudik, Federal Reserve Bank of Dallas.

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

§Corresponding author: Mahrad Sharifvaghefi, University of Pittsburgh, 230 S. Bouquet St., Pittsburgh, PA, USA, 15260. sharifvaghefi@pitt.edu.

1 Introduction

Models fitted to statistical relationships could be subject to parameter instabilities. In an extensive early study, Stock and Watson (1996) find that a large number 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. 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, such as 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 and regime switching models that are used for forecasting. See, for example, Hamilton (1988) and Dangl and Halling (2012). This literature does not address the problem of variable selection and takes the model specification as given.

The theory of variable selection in the presence of parameter instability is still largely underdeveloped. The application of penalized regression methods to variable selection is often theoretically justified under two key parameter stability assumptions: the stability of the coefficients in the data generating process and the stability of the correlation matrix of the covariates in the active set. Under these assumptions, the penalized regression methods can proceed using the full sample without down-weighting or separating the variable selection from the estimation stage. However, in the presence of parameter instability penalized regression methods must be adapted to simultaneously deal with selection and parameter change. 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, 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 by 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. These studies focus on specific forms of discrete or continuous time varying parameter models, and often carry out variable selection and estimation simultaneously using the penalized regression or boosting procedures.

This paper proposes the use of One Covariate at a Time Multiple Testing (OCMT) procedure proposed by Chudik et al. (2018) which is readily adapted to the task of variable selection under parameter instability. The key insight comes from the fact that coefficients of the noise variables that do not enter the data generating process are zero at all times. Consequently, using unweighted observations at the variable selection stage will be most effective in removing noise variables, while using weighted observations at the estimation stage can provide gains in terms of mean squared forecast errors. In this study, we allow the marginal effects of signals on the target variable, as well as the correlation of the covariates under consideration, to vary over time, assuming time variations in the marginal effects are not correlated with the signals. We establish the conditions required for OCMT with unweighted observations to select a model that contains all the signal variables and none of the noise variables with probability approaching one as the sample size, T, and the number of covariates under consideration, N, tend to infinity.

Clearly, it is also possible to use penalized regression methods with unweighted observations for the purpose of variable selection, and then estimate the selected model by the least squares method using weighted observations. However, as far as we know, there are no studies that consider the choice of the penalty term to achieve variable selection consistency under parameter instability. It is hoped that the present paper provides an impetus for further theoretical analysis of penalized regression techniques under parameter instability. Although at this stage a comparison of the assumptions required for variable selection consistency of OCMT and Lasso under parameter instability is not possible, in Section 4 we provide a discussion of the assumptions required for the variable selection consistency of Lasso under parameter stability that are comparable with the those required for the validity of the OCMT procedure.

The OCMT procedure selects variables based on the statistical significance of the net effect of the covariates in the active set on the target variable, one-at-a-time subject to 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, Bühlmann (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 an approximating model that includes all the signals and none of the noise variables. This model can contain covariates that do not enter the data generating process for the target variable but exhibit non-zero correlation with at least one signal, known as pseudo-signals.

Lasso and OCMT exploit different aspects of the low-dimensional structure assumed for the underlying data generating process. Lasso restricts the magnitude of the correlations within signals as well as the correlations between signals and the remaining covariates in the active set. OCMT limits the rate at which the number of pseudo-signals, k_T^*, rises with the sample size, T. Under parameter stability, the variable selection consistency of Lasso has been investigated by Zhao and Yu (2006), Meinshausen and Bühlmann (2006) and more recently by Lahiri (2021). These conditions, and how they compare with the conditions that underlie OCMT, are discussed in Section 4 of the paper. Although Lasso does not directly impose any restrictions on k_T^*, its Irrepresentable Condition (IRC), by restricting the magnitude of correlations within and between the signals and pseudo-signals, does have implications for the number of pseudo-signals that Lasso selects. OCMT requires k_T^* not to rise faster than $\sqrt{T}$. When this condition is violated, then the true signals must end up as common factors for the pseudo-signals, and what matters is the number of residuals (from the regressions of pseudo-signals on the common factors) that are correlated with the residuals of the true signals from the same set of common factors. Sharifvaghefi (2023) shows that such common factors can be estimated from the principal components of the covariates in the active set and the OCMT condition on the number pseudo-signals, now defined in terms of the correlation of the residuals, is no longer restrictive.[1] Once the model is selected, Theorem 2 establishes how the convergence rate of estimated coefficients of the selected variables depends on k_T^*. The regular convergence rate of $\sqrt{T}$ is achieved only if k_T^* is fixed in T. A similar issue also arises for Lasso, as shown by Lahiri (2021) who establishes that the Lasso procedure cannot achieve both variable selection consistency and $\sqrt{T}$-consistency in coefficient estimation. As noted above, the focus of the present paper is on the application of OCMT to variable selection in the presence of parameter instability, broadly defined. To the best of our knowledge, there are no studies that investigate the variable selection properties of Lasso under parameter instability.

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 go to a non-zero value, semi-strong signals whose average marginal effects converge to zero, but sufficiently slow, and weak signals whose average marginal effects approach 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-

[1] Another extension of OCMT is provided by Su et al. (2023) who allow for unknown potentially non-linear relationship between the signals and the target variable.

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 approximating model that contains all the signals (strong and semi-strong) and none of the noise variables approaches to one as T goes to infinity. Our results apply both when N is fixed as well as when N goes to infinity jointly with T, covering the case where $N \gg T$. We also establish conditions under which (a) least squares estimates of the coefficients of selected covariates converge to zero unless they are signals, and (b) the average squared 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 show that the OCMT procedure with weighted observations only at the estimation stage outperforms, in terms of mean squared forecast errors, Lasso and Adaptive Lasso (A-Lasso by Zou (2006)), as well as boosting by Bühlmann (2006), under many different settings.

Finally, we provide three empirical applications, forecasting monthly rates of price changes of 28 stocks in Dow Jones using large number of financial, economic and technical indicators, forecasting output growths across 33 countries using a large number of macroeconomic indicators, and forecasting euro area output growth using ECB surveys of 25 professional forecasters. To save space the third application is included in the online supplement. We generate a large number of forecasts using OCMT with and without down-weighting of the observations at the selection stage and compare the results with the forecasts obtained using Lasso, A-Lasso and boosting. 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 performance in terms of mean squared forecast errors and mean directional forecast accuracy. The empirical results also provide that OCMT with no down-weighting at the selection stage outperforms, in terms of mean squared forecast errors, boosting, Lasso and 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 the OCMT procedure for variable selection without down-weighting in the presence of parameter instability. Section 4 provides a discussion of key assumptions of Lasso and OCMT under parameter stability. Section 5 discusses the technical

assumptions and the asymptotic properties of the OCMT procedure under parameter instability. Section 6 provides the details of the Monte Carlo experiments and a summary of the main results. Section 7 presents the empirical applications, and Section 8 concludes. The paper is also accompanied with three online supplements. A theory supplement contains the mathematical proofs of the theorems and related lemmas. A Monte Carlo supplement 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. Further details of the empirical applications are given in an empirical supplement.

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 we say $f_n = \ominus(g_n)$ 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 $\sup_{n \geq n_0}(f_n/g_n) \leq C_1$. Similarly, if f_{iT} and g_{iT} are positive double sequences of real numbers for $i = 1, 2, 3, \cdots$; and $T = 1, 2, 3, \cdots$, then $f_{iT} = \ominus(g_{iT})$ if there exist $T_0 \geq 1$ and positive constants C_0 and C_1, such that $\inf_{T \geq T_0}(f_{iT}/g_{iT}) \geq C_0$ and $\sup_{T \geq T_0}(f_{iT}/g_{iT}) \leq C_1$.

2 Model specification under parameter instability

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

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

with time-varying parameters, $\{\beta_{it},\ i = 1, 2, ..., k\}$, and an error term, u_t. Intercepts and other pre-selected variables can also be included.[2] Since the parameters are time-varying we refer to the covariate i as "*signal*" if its average marginal effect, $\bar{\beta}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(\beta_{it})$, is not equal to zero. The strength of the signal can be captured by the exponent coefficient ϑ_i in $\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})$. For $\vartheta_i = 0$, the signal is strong and the average marginal effect, $\bar{\beta}_{i,T}$, does not converge to zero. For $0 < \vartheta_i < 1/2$, the signal is semi-strong and the average marginal effect converges to zero, but not too fast. For $\vartheta_i \geq 1/2$, the average marginal effect tends to zero very fast, making it infeasible for the OCMT procedure to distinguish such weak signals from noise, unless weak signals are sufficiently correlated with at least one strong or semi-strong signal. In this paper, we do not impose any restrictions on the correlations among signals, and we focus

[2] See the working paper version of the paper available at `https://doi.org/10.24149/gwp394r2`.

only on the covariates with strong and semi-strong signals, where $0 \leq \vartheta_i < 1/2$. For simplicity of exposition, unless specified otherwise, we will refer to both strong and semi-strong signals simply as signals.

The identity of the k signals are unknown, and 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. The time variations in β_{it}, for $i = 1, 2, ..., k$, are assumed to be exogenous, in the sense that β_{it} are distributed independently of the covariates in the active set $\mathcal{S}_{Nt}$. This assumption rules out correlated time variations that can arise in non-linear regressions where y_t is a non-linear function of the signals. One important example is given by the bilinear model

$$y_t = \sum_{i=1}^{k} \beta_i(x_{it}) x_{it} + u_t,$$

where it is assumed that β_{it} systematically varies with x_{it}. Nevertheless, in the context of linear regressions, our assumptions about parameter instability includes many models of parameter instability studied in the literature. Specifically, our analysis accommodates cases where the coefficients 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, as

$$\beta_{it} = \beta_i^{(s)}, \text{ if } t \in [T_{s-1}, T_s) \text{ for } s = 1, 2, \cdots, S,$$

where $T_0 = 1$ and $T_S = T$.

In this paper we follow Chudik et al. (2018) and consider the application of the OCMT procedure for variable selection even when the parameters are time-varying, and provide theoretical arguments in favour of using the full sample of data available without down-weighting. We first recall that OCMT's variable selection is based on the net effect of x_{it} on y_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. The average net effect of the covariate x_{it} on y_t can be defined as

$$\bar{\theta}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it} y_t).$$

By substituting y_t from (1) we can further write $\bar{\theta}_{i,T}$ as (noting that β_{jt} and x_{it} are assumed to

be independently distributed)

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

where $\sigma_{ij,t} = \mathbb{E}(x_{it}x_{jt})$, and $\bar{\sigma}_{iu,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it}u_t)$. In what follows we allow for a mild degree of correlation between x_{it}, and u_t, by assuming that $\bar{\sigma}_{iu,T} = O(T^{-\epsilon_i})$, for some $\epsilon_i \geq 1/2$. In this case the average net effect of the i^{th} covariate simplifies to

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

In line with our assumption about the average 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 $\epsilon_i \geq 1/2$.

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 approaching one by OCMT.

3 Parameter instability and OCMT

The OCMT procedure begins with N separate regressions, 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:

$$y_t = \phi_{i,T} x_{it} + \eta_{it}, \text{ for } t = 1, 2, \cdots, T; \; i = 1, 2, ..., N, \tag{2}$$

where

$$\phi_{i,T} \equiv \left(T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it}^2) \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it} y_t) \right) = \left[\bar{\sigma}_{ii,T} \right]^{-1} \bar{\theta}_{i,T}, \tag{3}$$

with $\bar{\sigma}_{ii,T} = T^{-1} \sum_{t=1}^{T} \sigma_{ii,t}$. 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*	$\bar{\beta}_{i,T} = \ominus(T^{-\vartheta_i})^{\dagger}$ 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 $\bar{\theta}_{i,T} = \ominus(T^{-\epsilon_i})$

† The signals are assumed to be (semi) strong such that $0 \leq \vartheta_i < 1/2$.

for some $0 \leq \vartheta_i < 1/2$, and $\epsilon_i \geq 1/2$. To simplify the exposition, we consider the covariates x_{it}, for $i = 1, 2, \cdots, k$, as signals, and for $i = k+1, k+2, \cdots, k+k_T^*$, as pseudo-signals. The remaining covariates in the active set, $\{x_{it}$, for $i = k+k_T^*+1, k+k_T^*+2, \cdots, N\}$, are classified as (pure) noise variables. We assume that the number of signals, k, is a finite fixed integer but we allow the number of pseudo-signals, denoted by k_T^*, to grow with N and T. 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 $\epsilon_i \geq 1/2$). 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. To allow for hidden signals, Chudik et al. (2018) extend the OCMT method to have multiple stages.

The OCMT procedure

1. For $i = 1, 2, \cdots, N$, regress y_t on x_{it}; $y_t = \phi_{i,T} x_{it} + \eta_{it}$; and compute the t-ratio of $\phi_{i,T}$, given by

$$t_{i,T} = \frac{\hat{\phi}_{i,T}}{s.e.\left(\hat{\phi}_{i,T}\right)} = \frac{\sum_{t=1}^{T} x_{it} y_t}{\hat{\sigma}_i \sqrt{\sum_{t=1}^{T} x_{it}^2}}, \tag{4}$$

   where $\hat{\phi}_{i,T} = \left(\sum_{t=1}^{T} x_{it}^2\right)^{-1} \left(\sum_{t=1}^{T} x_{it} y_t\right)$ is the least squares estimator of $\phi_{i,T}$, $\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} \hat{\eta}_{it}^2$, and $\hat{\eta}_{it} = y_t - \hat{\phi}_{i,T} x_{it}$, is the regression residual.

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

$$c_p(N, \delta) = \Phi^{-1}\left(1 - \frac{p}{2N^{\delta}}\right), \tag{5}$$

   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

$$\hat{\mathcal{J}}_i = I\left[|t_{i,T}| > c_p(N, \delta)\right], \text{ for } i = 1, 2, \cdots, N. \tag{6}$$

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

OCMT uses the t-ratio of $\phi_{i,T}$, defined by (4), to select the signals (strong as well as semi-strong), $\{x_{it} : i = 1, 2, \cdots, k\}$, and none of the noise variables, $\{x_{it} : k+k_T^*+1, k+k_T^*+2, \cdots, N\}$. The selected model is referred to as an approximating model since it can include pseudo-signals, $\{x_{it} : k+1, k+2, \cdots, k+k_T^*\}$, that proxy for the true 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, to be discussed below in Section 5.

Before presenting our technical assumptions and theoretical results under parameter instability, it is instructive to discuss and compare the key conditions under which Lasso and OCMT lead to consistent model selection under parameter stability.

4 Lasso and OCMT under parameter stability

As formally established by Zhao and Yu (2006) and Meinshausen and Bühlmann (2006), three main conditions are required for the Lasso variable selection to be consistent. Here we follow Lahiri (2021) who also considers the convergence of Lasso estimated coefficients to their true values. The key condition is the "Irrepresentable Condition" (IRC) that places restrictions on the magnitudes of the sample correlations across the signals, $\mathbf{x}_{1t} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, and the rest of the covariates in the active set, namely $\mathbf{x}_{2t} = (x_{k+1,t}, x_{k+2,t}, \cdots, x_{Nt})'$. Let

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}$$

be the $N \times N$ matrix of sample correlations of the covariates in the active set, partitioned conformably to $\mathbf{x}_t = (\mathbf{x}_{1t}', \mathbf{x}_{2t}')'$. The IRC can be written as

$$\left\| \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \text{sign}\left(\boldsymbol{\beta}_0\right) \right\|_\infty \leq 1, \tag{7}$$

where $\|.\|_\infty$ is the ℓ_∞ norm of a vector, sign(.) is the sign function, and $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \cdots, \beta_{0k})'$ is the $k \times 1$ vector of the coefficients of the signals. The following example provides more intuition on how IRC imposes restrictions on the magnitudes of the sample correlations between the covariates in the active set.

Example 1 *Suppose the DGP for y_t contains only two signals, x_{1t} and x_{2t}. Denote the sample correlation coefficient between x_{1t} and x_{2t} by $\hat{\rho}$, and the sample correlation coefficients of x_{1t} and x_{2t} with the rest of the covariates in the active set, $x_{3,t}, x_{4,t}, \cdots, x_{Nt}$, by $\hat{\rho}_{i1}$ and $\hat{\rho}_{i2}$, for*

$i = 3, 4, \cdots, N$, *respectively. Then, after some algebra, the IRC given by (7) simplifies to*

$$\max_{i\in\{3,4,\cdots,N\}} |(\hat{\rho}_{i1} - \hat{\rho}\hat{\rho}_{i2})sign(\beta_{01}) + (\hat{\rho}_{i2} - \hat{\rho}\hat{\rho}_{i1})sign(\beta_{02})| \leq 1 - \hat{\rho}^2.$$

There are two cases: (A) $sign\,(\beta_{01}) = sign\,(\beta_{02})$ *and (B)* $sign\,(\beta_{01}) \neq sign\,(\beta_{02})$*. Under case (A) it follows that the IRC condition is met if*

$$\max_{i\in\{3,4,\cdots,N\}} |\hat{\rho}_{i1} + \hat{\rho}_{i2}| \leq 1 + \hat{\rho}.$$

Similarly under case (B) it is required that

$$\max_{i\in\{3,4,\cdots,N\}} |\hat{\rho}_{i1} - \hat{\rho}_{i2}| \leq 1 - \hat{\rho}.$$

From the above example, it is clear that IRC places restrictions on the magnitude of sample correlation among signals ($\hat{\rho}$ in the above example), as well as the magnitude of sample correlation between signals and pseudo-signals ($\hat{\rho}_{i1}$ and $\hat{\rho}_{i2}$). Notably, the IRC is met for noise variables but need not hold for pseudo-signals. OCMT also has no difficulty in dealing with noise variables, and is very effective at eliminating them. However, for consistent estimation of the approximate model, post OCMT selection, it is necessary to restrict the number of selected covariates relative to the sample size, T. To this end, OCMT assumes that the number of pseudo-signals, k_T^*, could grow at an order less than the square root of the number of observations, namely

$$k_T^* = \ominus(T^d) \text{ for some } 0 \leq d < \frac{1}{2}.$$

It is important to note that OCMT does not place any restrictions on the magnitude of correlations of signals and pseudo-signals. Instead, it limits the number of covariates that are correlated with the signals (k_T^*). Clearly, the IRC could be violated even when the number of pseudo-signals grows at an order less than $\sqrt{T}$. Hence the OCMT's requirement on the number of pseudo-signals allows for cases where the IRC does not hold, and *vice versa*.

The condition on the number of pseudo-signals (k_T^*) in the OCMT framework has been recently relaxed by Sharifvaghefi (2023). To illustrate how this is done, suppose there are no noise variables and hence the signals, $\mathbf{x}_{1t} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, are correlated with all the remaining covariates in the active set. In this case if $N \gg \sqrt{T}$, a straightforward application of OCMT will not be valid. But, we can model the correlation between the signals, $\mathbf{x}_{1t}$, and the remaining covariates, $\mathbf{x}_{2t}$, as

$$x_{it} = \sum_{j=1}^{k} \psi_{ij} x_{jt} + \xi_{it} = \boldsymbol{\psi}_i' \mathbf{x}_{1t} + \xi_{it}, \text{ for } i = k+1, k+2, \cdots, N.$$

The signals thus act as strong factors for the pseudo-signals. Given that the identity of signals and pseudo-signals are unknown and the number of pseudo-signals is large, it is reasonable to propose the existence of latent factors, $\mathbf{f}_t$, that are common across the covariates in the active set. This idea can be formally expressed as:

$$x_{it} = \boldsymbol{\psi}_i' \mathbf{f}_t + \varepsilon_{it} \quad \text{for } i = 1, 2, \cdots, N,$$

where $\boldsymbol{\psi}_i$ is vector of factor loadings, and ε_{it} refers to the idiosyncratic components that are weakly cross-correlated such that

$$\sup_j \sum_{i=1}^{N} |cov(\varepsilon_{it}, \varepsilon_{jt})| < C < \infty. \tag{8}$$

Substituting x_{it} into the DGP for y_t, given by (1), we obtain:

$$y_t = \boldsymbol{\delta}_0' \mathbf{f}_t + \sum_{i=1}^{k} \beta_{i0} \varepsilon_{it} + u_t,$$

with $\boldsymbol{\delta}_0 = \sum_{i=1}^{k} \beta_{i0} \boldsymbol{\psi}_i$. When the common factors, $\mathbf{f}_t$, and idiosyncratic components, ε_{it} , are known, this model would correspond to that presented in working paper version of our work, where common factors $\mathbf{f}_t$ can be used as preselected variables. Since $\mathbf{f}_t$ and ε_{it} are not known, Sharifvaghefi (2023) shows that when both N and T are large the OCMT selection can be carried out using the principal component estimators of $\mathbf{f}_t$ and ε_{it}, denoted by $\hat{\mathbf{f}}_t$ and $\hat{\varepsilon}_{it}$, using all the covariates in the active set. The large N is required for consistent estimation of the common factors. As a result, the OCMT condition on the number of pseudo-signals now relates to the number of ε_{it} for $i = k+1, k+2, ..., N$ that are correlated with ε_{it} for $i = 1, 2, ..., k$, which is bounded under condition (8).

For variable selection consistency of Lasso under parameter stability, the literature further requires the penalty term, λ_T, to grow at an order greater than $\sqrt{T}$ such that:

$$\lim_{T \to \infty} \Pr \left(\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{x}_{2t}^{\perp} u_t \right\|_{\infty} > \frac{\lambda_T}{\sqrt{T}} \right) = 0,$$

where $\mathbf{x}_{2t}^{\perp}$ is the part of variation in $\mathbf{x}_{2t}$ that is orthogonal to $\mathbf{x}_{1t}$ and u_t is the error term in the data generating process. The exact choice of λ_T in practice is often unclear, with practitioners typically relying on cross-validation methods.

A third condition required by Lasso for variable selection consistency is the beta-min condition:

$$\min_{j=1,2,\cdots,k} |\beta_{j0}| > (2T)^{-1} \lambda_T \left| \mathbf{R}_{11}^{-1} \text{sign}\left(\boldsymbol{\beta}_0\right) \right|_j$$

where $|.|_j$ denotes the absolute value of the j^{th} element of a vector. Given that λ_T must grow at an order greater than $\sqrt{T}$, we can conclude from the beta-min condition that $\beta_{i0} \gg \frac{1}{\sqrt{T}}$ for $i = 1, 2, \cdots, k$. For example, Lahiri (2021) assumes that $\beta_{i0} \gg \sqrt{\frac{k \log(T)}{T}}$. The OCMT's requirement on the strength of signals (under parameter stability) is given by $\beta_{i0} = \ominus(T^{-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$. This condition is essentially very similar to the Lasso's beta-min condition.

5 Asymptotic properties of OCMT under parameter instability

We establish the asymptotic properties of the OCMT procedure for variable selection assuming the time variations in β_{it} for $i = 1, 2, ..., k$ are distributed independently of the regressors in the active set. We also make additional assumptions that bound the degree of time variations in β_{it} and x_{it}, in addition to assuming the exponentially decaying tail probabilities for β_{it} and x_{it}. Our assumptions on x_{it}, $i = 1, 2, ..., k$ and their correlations with the other variables in the active set are in line with those assumed in the literature. A formal statement of these assumptions are set out in Section 5.1. Theorem 1 establishes that OCMT continues to asymptotically select an approximating model that includes all the signals and none of the noise variables. Additional assumptions are required for investigating the asymptotic properties of the least squares estimates of the post OCMT selected model. These assumptions and the related theorems are provided in Section 5.3. Theorem 2 establishes the rate at which the least squares estimates of the coefficients of the selected model converge to their true time averages. It is shown that the regular convergence rate of $\sqrt{T}$ is achieved only if k_T^* (the number of selected covariates) is fixed in T. Irregular convergence rates result when k_T^* rises in T. Theorem 3 shows that the sum of squared residuals of the estimated model converges in probability to its limiting value at the oracle rate of $\sqrt{T}$. The limiting value consists of two components: the first is the unavoidable uncertainty due to the unobserved error term, u_t, and the second is the cost (in terms of fit) of ignoring the time variations in the coefficients of the signals.

Suppose the target variable, y_t, is generated by (1) in terms of x_{it} for $i = 1, 2, ..., k$, and $\mathbf{x}_t = (x_{1t}, x_{2t}, \cdots, x_{kt}, x_{k+1,t}, ..., x_{Nt})'$ is the $N \times 1$ vector of covariates in the active set ($N \gg k$). Let $\bar{\beta}_{i,T} \equiv T^{-1} \sum_{t=1}^{T} \mathbb{E}(\beta_{it})$, for $i = 1, 2, ..., k$, and $\bar{\theta}_{i,T} = \sum_{j=1}^{k} \left(T^{-1} \sum_{t=1}^{T} \mathbb{E}(\beta_{jt}) \sigma_{ij,t} \right) + \bar{\sigma}_{iu,T}$, for $i = 1, 2, ..., N$, where $\sigma_{ij,t} = \mathbb{E}(x_{it} x_{jt})$, and $\bar{\sigma}_{iu,T} = T^{-1} \sum_{t=1}^{T} \mathbb{E}(x_{it} u_t)$. Define the filtrations $\mathcal{F}_t^u = \sigma(u_t, u_{t-1}, \cdots)$, $\mathcal{F}_t^x = \sigma(\mathbf{x}_t, \mathbf{x}_{t-1}, \cdots)$, and $\mathcal{F}_{jt}^{\beta} = \sigma(\beta_{jt}, \beta_{j,t-1}, \cdots)$, for $j = 1, 2, \cdots, k$. Set $\mathcal{F}_t^{\beta} = \cup_{j=1}^{k} \mathcal{F}_{jt}^{\beta}$ and $\mathcal{F}_t = \mathcal{F}_t^q \cup \mathcal{F}_t^{\mathrm{a}} \cup \mathcal{F}_t^{\beta} \cup \mathcal{F}_t^u$, and consider the following assumptions:

5.1 Assumptions

Assumption 1 (Coefficients of signals)

(a) The number of signals, k, is a finite fixed integer. (b) β_{jt}, $j = 1, 2, \cdots, k$, are distributed independently of $x_{it'}$, $i = 1, 2, \cdots, N$, and $u_{t'}$ for all t and t'. (c) The signals are (semi) strong in the sense that $\bar{\beta}_{j,T} = \ominus(T^{-\vartheta_j})$ for $0 \leq \vartheta_j < 1/2$, $j = 1, 2, ..., k$. (d) There are no hidden signals in the sense that $\bar{\theta}_{j,T} = \ominus(T^{-\vartheta_j})$, for $0 \leq \vartheta_j < 1/2$, $j = 1, 2, ..., k$.

Assumption 2 (Martingale difference processes)

For $i, i' = 1, 2, \cdots, N$, $j = 1, 2, \cdots, k$, and $t = 1, 2, \cdots, T$, (a) $\mathbb{E}\left[x_{it}x_{i't} - \mathbb{E}(x_{it}x_{i't})|\mathcal{F}_{t-1}\right] = 0$, (b) $\mathbb{E}\left[u_t^2 - \mathbb{E}\left(u_t^2\right)|\mathcal{F}_{t-1}\right] = 0$, (c) $\mathbb{E}\left[x_{it}u_t - \mathbb{E}(x_{it}u_t)|\mathcal{F}_{t-1}\right] = 0$, where $T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}u_t) = O(T^{-\epsilon_i})$, with $\epsilon_i \geq 1/2$, and (d) $\mathbb{E}\left[\beta_{jt} - \mathbb{E}(\beta_{jt})|\mathcal{F}_{t-1}\right] = 0$.

Assumption 3 (Exponential decaying probability tails)

There exist sufficiently large positive constants C_0 and C_1, and $s > 0$ such that for all $\alpha > 0$, (a) $\sup_{i,t}\Pr(|x_{it}| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, (b) $\sup_{i,t}\Pr(|\beta_{it}| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, and (c) $\sup_t \Pr(|u_t| > \alpha) \leq C_0\exp(-C_1\alpha^s)$.

Before presenting the theoretical results, we briefly discuss the rationale behind our assumptions and compare them with the assumptions typically made in the high-dimensional linear regressions and the parameter instability literature.

Assumption 1(a) posits that the number of signals is a fixed integer. This is crucial to ensure that the random variable y_t has a distribution with an exponentially decaying probability tail. Under the premise that the covariates x_{it} for all i and t are non-random and fixed, which is a common assumption in the penalized regression setting, it becomes permissible for the number of signals to grow with the sample size at an order slower that $\sqrt{T}$. Assumption 1(b) is common in the literature under parameter instability and restrict the distribution of time-varying parameters to be independent of the covariates. Assumption 1(c) is an identification assumption needed to distinguish signals from noise variables and is similar to the beta-min condition already discussed in Section 4. Finally, Assumption 1(d) ensures that there are no hidden signals. As discussed in Section 3, we make this assumption to simplify the theoretical derivations, and one can use the multi-stage OCMT procedure suggested by Chudik et al. (2018) to allow for hidden signals.

To establish that the OCMT procedure with the critical value function $c_p(N, \delta) = \Phi^{-1}\left(1 - \frac{p}{2N^\delta}\right)$ does not select any of the noise variables with a probability approaching one as N and T go to infinity, we need to show that the t-statistic given by (4) follows a distribution with exponentially decaying tails. We utilize the concentration inequality of an exponential decaying rate

to accomplish this goal. Assumptions 2 and 3 place constraints on the sequence of random variables, x_{it} for $i = 1, 2, \ldots, N$, β_{jt} for $j = 1, 2, \ldots, k$, and u_t such that they adhere to a martingale difference process and exhibit exponential decaying probability tails. These assumptions are sufficient to establish the exponential decaying concentration inequality, as provided in Lemma S-3.1 in the online theory supplement. Notably, these assumptions could be relaxed provided that the exponential decaying concentration inequality holds. For example, Theorem 1 of Merlevède et al. (2011) and Lemma D1 of the online theory supplement for Chudik et al. (2018) establishes that this inequality can be achieved while allowing for weak time-series dependence. In penalized regression literature, a commonly held assumption is that the covariates are non-random and fixed. Moreover, error terms $\{u_t\}_{t=1}^{T}$ are typically assumed to be serially independent. See, for example, see Zhao and Yu (2006), Javanmard and Montanari (2013), Lee et al. (2015), Belloni et al. (2014), Javanmard and Lee (2020), and Lahiri (2021). Additionally, in the Lasso literature it is often assumed that u_t possesses an exponentially decaying probability tail. See, for example, Javanmard and Montanari (2018), Hansen and Liao (2019), Fan et al. (2020), and Javanmard and Lee (2020).

5.2 Variable selection consistency

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 the assumptions set out in Section 5.1, 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

$$\mathcal{A}_0 = \left\{\sum_{i=1}^{k} \hat{\mathcal{J}}_i = k\right\} \cap \left\{\sum_{i=k+k_T^*+1}^{N} \hat{\mathcal{J}}_i = 0\right\}. \tag{9}$$

Note that 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 *Consider the DGP for y_t, $t = 1, 2, \cdots, T$, given by (1), and the set $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$ that contains k signals, k_T^* pseudo-signals, and $N - k - k_T^*$ noise variables. Suppose that Assumptions 1-3 hold and $N = \ominus(T^{\kappa})$ with $\kappa > 0$. Then, there exist finite positive constants C_0 and C_1 such that, for any $0 < \pi < 1$ and any null sequence $d_T > 0$, the probability of selecting the approximating model $\mathcal{A}_0$, as defined by (9), by the OCMT procedure with the*

critical value function $c_p(N,\delta)$ *given by (5), for some* $\delta > 0$*, is*

$$\Pr(\mathcal{A}_0) = 1 - O\left[T^{\kappa\left(1-\mathcal{X}_{NT}\left(\frac{1-\pi}{1+d_T}\right)^2\delta\right)}\right] - O\left[T^{\kappa}\exp\left(-C_0 T^{C_1}\right)\right], \tag{10}$$

where,

$$\mathcal{X}_{NT} = \inf_{i\in\{k+k^*+1,\cdots,N\}} \frac{\bar{\sigma}^2_{\eta_i,T}\bar{\sigma}^2_{x_i,T}}{\bar{\omega}^2_{iy,T}},$$

$\bar{\sigma}^2_{x_i,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}^2)$, $\bar{\omega}^2_{iy,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}^2 y_t^2|\mathcal{F}_{t-1})$, $\bar{\sigma}^2_{\eta_i,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\eta_{it}^2)$, $\eta_{it} = y_t - \phi_{i,T}x_{it}$*, and* $\phi_{i,T}$ *is defined by (3).*

This theorem shows that the probability of selecting the approximating model is unaffected by parameter instability, so long as the average net effects of the signals are non-zero or converge to zero sufficiently slowly in T, as defined formally by Assumption 1. The theorem also highlights the importance of an appropriate choice of δ for model selection consistency. Corollary S.1 in the online theory supplement shows that if the covariates in the active set are generated by a stationary process and the noise variables are independent of y_t then $\mathcal{X}_{NT} = 1$. As a result, for any $\delta > 1$, OCMT consistently selects the approximating model, $\mathcal{A}_0$. Notably, $c_p(N,\delta)$ is reasonably stable with respect to small increases in δ in the neighborhood of $\delta = 1$ and the extensive Monte Carlo studies in Chudik et al. (2018) also suggest that setting $\delta = 1$ performs well in practice.[3]

5.3 Properties of the post OCMT selected model

To investigate the asymptotic properties of the least squares estimates of the selected model (post OCMT) we require the following additional assumption:

Assumption 4 (Eigenvalues) *The eigenvalue condition*

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

holds, where $\mathbf{x}_{\tilde{k}_T,t}$*, for* $t = 1,2,...,T$ *are the* $\tilde{k}_T \times 1$ *vector of observations on signals (k) and pseudo-signals (*k_T^**) with* $\tilde{k}_T = k + k_T^*$*.*

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

[3]One could also use the heteroscedasticity and/or autocorrelation robust standard errors in computation of t-statistics given by (4) to ensure the consistent selection of the approximating model for any $\delta > 1$ in a more general setup.

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 (6). 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

$$y_t = \sum_{i=1}^{N} \hat{\mathcal{J}}_i x_{it} b_i + \eta_t = \sum_{\ell=1}^{\hat{k}_T} \gamma_\ell w_{\ell t} + \eta_t, \tag{11}$$

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

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\sum_{t=1}^{T} \mathbf{w}_t \mathbf{w}_t'\right)^{-1} \left(T^{-1}\sum_{t=1}^{T} \mathbf{w}_t y_t\right), \tag{12}$$

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 that 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 signals and pseudo-signals of the approximating model tend to their true mean values, defined formally below.

Theorem 2 *Let the DGP for y_t, $t = 1, 2, \cdots, T$, be given by (1) and write down the regression model selected by the OCMT procedure as (11). Suppose that Assumptions 1-4 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 least squares (LS) estimator of $\boldsymbol{\gamma}_T = \left(\gamma_1, \gamma_2, \cdots, \gamma_{\hat{k}_T}\right)'$, given by (12).*

(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

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

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

Remark 1 *The above theorem builds on Theorem 1 and establishes that in the post OCMT selected model estimated by LS only signals will end up having non-zero limiting values, as* N *and* $T \to \infty$*. This theorem also shows that the convergence rate of the LS estimators depends on* d*, defined by* $k_T^* = \ominus(T^d)$*, and the regular* $\sqrt{T}$ *rate of convergence is achieved only if* $d = 0$*. Similarly, Lahiri (2021) establishes that the Lasso procedure cannot achieve both variable selection consistency and* $\sqrt{T}$*-consistency in coefficient estimation.*

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 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-varying means, and still estimate* $\boldsymbol{\gamma}_T^*$ *consistently by LS.*

Lastly, we consider the fit of the post OCMT selected regression in terms of its residuals given by

$$\hat{\eta}_t = y_t - \sum_{\ell=1}^{\hat{k}_T} \hat{\gamma}_\ell w_{\ell t}, \text{ for } t = 1, 2, ..., T. \tag{13}$$

It is worth noting that even when all the signal variables are correctly selected, the forecasts based on the selected model will be biased due to parameter instability. The implications of parameter instability for the in-sample fit of the selected regression is derived in Proposition S.1 of the online theory supplement, abstracting from variable selection uncertainty. In what follows we derive the asymptotic properties of the sum of squared residuals (SSR) of the selected model, namely $\sum_{t=1}^{T} \hat{\eta}_t^2$, taking account of the costs associated with variable selection uncertainty and parameter instability. To this end we need the following assumption on the cross correlation of parameter heterogeneity.

Assumption 5 (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 5 is a high-level assumption. Here is an example of conditions under which this assumption holds. Suppose, Assumptions 1 and 2 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, $\sum_{t=1}^{T}\sum_{t'=1}^{T} cov(h_{ij,t}, h_{ij,t'}) = O(T)$. *See Lemma S-2.8 in the online theory supplement for a proof.*

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 (1) and write down the regression model selected by the OCMT procedure as (11). 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 residuals of the selected model, estimated by LS and given by (13).*

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

$$T^{-1}SSR = \bar{\sigma}_{u,T}^2 + \bar{\Delta}_{\beta,T} + O_p\left(T^{-\frac{1}{2}}\right) + O_p\left(T^{d-1}\right), \tag{14}$$

where $\bar{\sigma}_{u,T}^2 = T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(u_t^2\right)$, *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) Let $\tilde{k}_T = k + k_T^*$ *and suppose that* $\mathbb{E}\left(\mathbf{x}_{\tilde{k}_T,t}\mathbf{x}'_{\tilde{k}_T,t}\right)$ *is time-invariant (fixed). Then,*

$$T^{-1}SSR = \bar{\sigma}_{u,T}^2 + \bar{\Delta}_{\beta,T}^* + O_p\left(T^{-\frac{1}{2}}\right) + O_p\left(T^{d-1}\right), \tag{15}$$

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* $\sigma_{ijt,\beta}^* = \mathbb{E}\left[(\beta_{it} - \bar{\beta}_{i,T})(\beta_{jt} - \bar{\beta}_{j,T})\right]$.

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 (14) and (15), we can conclude that the average of squared 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).*

Results (14) and (15) in Theorem 3 show that the SSR of the selected model depends on (*i*) the unavoidable uncertainty due to the unobserved error term, u_t, given by the term $\bar{\sigma}_{u,T}^2$, (*ii*) the cost (in terms of fit) of ignoring the time variation in the coefficients of the signals, $\boldsymbol{\beta}_{it}$, $i = 1, 2, \cdots, k$, as given by the term $\bar{\Delta}_{\beta,T}$ and $\bar{\Delta}_{\beta,T}^*$, respectively, and (*iii*) the $O_p\left(T^{-1/2}\right)$ term

due to sampling uncertainty (which will be present even in the absence of variable selection uncertainty), and (iv) the $O_p\left(T^{d-1}\right)$ term which is due to variable selection uncertainty, and will be dominated by $O_p\left(T^{-1/2}\right)$ when $d < 1/2$. Therefore, the cost of variable selection can be controlled when using OCMT if the number of pseudo-signals, k_T^*, do not rise faster than $\sqrt{T}$. However, to reduce the cost associated with parameter instability more information about the nature of time variations in $\boldsymbol{\beta}_{it}$ and $\sigma_{ijt,x}$ are required. For example, $\bar{\Delta}_{\beta,T}$ (or $\bar{\Delta}^*_{\beta,T}$) could be lower if $\boldsymbol{\Omega}_{\beta,t}$ 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 squared forecast errors 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 of parameter instability at the expense of increased sampling uncertainty by appropriate weighting of the observations.

6 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 squared forecast error (MSFE) of the selected models. Sub-section 6.1 outlines the MC designs, sub-section 6.2 provides a summary of how the OCMT, Lasso, A-Lasso, and boosting procedures are implemented, and finally sub-section 6.3 presents the main MC findings. Details of Lasso, A-Lasso, and boosting procures and how they are implemented are provided in Section S-1 of the online Monte Carlo supplement.

6.1 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 $\mu_{jt} = \mathbb{E}\left(\tilde{x}_{jt}\right)$. 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

$$y_t = d_t + \rho_{y,t} y_{t-1} + \sum_{j=1}^{k} \beta_{jt} x_{jt} + \tau_u u_t, \tag{16}$$

where

$$d_t = c_t + \sum_{j=1}^{k} \beta_{jt} \mu_{jt}. \tag{17}$$

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 (16). 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 (16) 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 (17), 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,

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

using the starting values $\eta_{j,0} \sim IID\mathcal{N}(0,1)$, and $\rho_{\eta j} = 0.5$, for all j. $\{e_{\eta_j t}\}$ 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

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

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

$$b_{1t} = b_{2t} = \begin{cases} 2 & \text{if } t \in \{1, 2, \cdots, [T/3]\}, \\ 0 & \text{if } t \in \{[T/3]+1, [T/3]+2, \cdots, [2T/3]\}, \\ 1 & \text{if } t \in \{[2T/3]+1, [2T/3]+2, \cdots, T\}, \end{cases} \tag{18}$$

and

$$b_{3t} = b_{4t} = \begin{cases} 0.5 & \text{if } t \in \{1, 2, \cdots, [T/2]\}, \\ 1.5 & \text{if } t \in \{[T/2]+1, [T/2]+2, \cdots, T\}, \end{cases} \tag{19}$$

where [.] is the nearest integer function.

We also set $c_t = 0$ in (17) and generate the intercept as $d_t = \sum_{j=1}^{k} \beta_{jt}\mu_{jt}$, where

$$\mu_{1t} = \mu_{2t} = \begin{cases} 0.6 & \text{if } t \in \{1, 2, \cdots, [T/3]\}, \\ 1.5 & \text{if } t \in \{[T/3]+1, [T/3]+2, \cdots, [2T/3]\}, \\ 0.9 & \text{if } t \in \{[2T/3]+1, [2T/3]+2, \cdots, T\}, \end{cases} \tag{20}$$

and

$$\mu_{3t} = \mu_{4t} = \begin{cases} 0.9 & \text{if } t \in \{1, 2, \cdots, [T/2]\}, \\ 1.1 & \text{if } t \in \{[T/2]+1, [T/2]+2, \cdots, T\}. \end{cases} \tag{21}$$

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 and set r_t equal to 0.9 in the first half of the sample and 0.4 in the second half of the sample. 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

$$\rho_{y,t} = \begin{cases} 0.2 & \text{if } t \in \{1, 2, \cdots, [T/2]\}, \\ 0.4 & \text{if } t \in \{[T/2]+1, [T/2]+2, \cdots, T\}. \end{cases} \tag{22}$$

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.

6.2 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. 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). 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.

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

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.

6.3 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 most of the times selects fewer covariates relative to Lasso, A-Lasso, and boosting , while maintaining the TPR at a similar level. As a result, OCMT has mostly the lowest average FPR among the selection methods under consideration. 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, and A-Lasso. 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 almost always 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 parameter instabilities separately, on left and right panels of Table 3. Two observations stand out from this table. First, the reported average relative MSFEs are almost always greater than one for all the N and T choices, indicating that OCMT outperforms Lasso, A-Lasso, and boosting. Second, the degree to which OCMT outperforms Lasso and A-Lasso tends to increase with the degree of parameter 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.

7 Empirical applications

The rest of the paper considers 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:[4]

- 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 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.

[4]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.

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.

7.1 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. [5] 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.[6] 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

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

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

the selection stage dominates Lasso, A-Lasso and boosting in terms of MSFE and the differences are statistically highly significant.

Further, OCMT outperforms Lasso, A-Lasso and boosting 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 the direction of 56.057% of 8,659 forecasts, as compared to 55.769%, which we obtain for Lasso, A-Lasso and boosting 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, A-Lasso and boosting. As can be seen from the last column of Table 6, Lasso, A-Lasso and boosting on average select many more covariates than OCMT (1-15 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, A-Lasso and boosting in the majority of the stocks in terms of MSFE and MDFA. OCMT outperforms Lasso, A-Lasso and boosting in at least 22 out of 28 stocks in terms of MSFE, under no down-weighting, and almost universally when Lasso, A-Lasso and boosting 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.

7.2 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.[7] 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:

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

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

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

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 application, 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, A-Lasso and boosting 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 and boosting. Table 12 summarizes country-specific MSFE and DM findings for OCMT relative to Lasso, A-Lasso and boosting. The results show OCMT underperforms Lasso in more than half of the countries for one-year ahead horizon, but outperforms Lasso, A-Lasso and boosting 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, A-Lasso and boosting. Overall, OCMT has a slightly higher MDFA and hence predicts the direction of real output growth better than Lasso, A-Lasso and boosting 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. Boosting on average selects 2.7 variables from the active set for $h = 4$, and 1.4 variables on average for $h = 8$.

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.

8 Concluding remarks

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 feature 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. Our MC results as well as empirical applications uniformly suggest that OCMT without down-weighting at the selection stage outperforms, in terms of mean squared forecast errors, Lasso, Adaptive Lasso, boosting, as well as when OCMT is applied with down-weighted observations.

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
A. Without parameter instabilities									
	OCMT								
20	5.03	6.17	7.22	0.83	0.91	0.96	0.08	0.13	0.17
40	4.69	5.98	6.87	0.80	0.91	0.95	0.04	0.06	0.08
100	4.31	5.52	6.35	0.77	0.88	0.93	0.01	0.02	0.03
	Lasso								
20	6.82	7.00	7.20	0.84	0.89	0.93	0.17	0.17	0.17
40	8.26	8.57	8.74	0.82	0.89	0.92	0.12	0.13	0.13
100	10.76	11.00	10.51	0.79	0.87	0.90	0.08	0.08	0.07
	A-Lasso								
20	5.15	5.35	5.55	0.73	0.80	0.85	0.11	0.11	0.11
40	6.39	6.78	6.96	0.73	0.81	0.86	0.09	0.09	0.09
100	8.65	9.05	8.83	0.72	0.81	0.86	0.06	0.06	0.05
	Boosting								
20	4.59	4.63	4.70	0.77	0.83	0.88	0.08	0.07	0.06
40	6.04	5.79	5.69	0.76	0.83	0.87	0.07	0.06	0.05
100	11.36	9.27	8.43	0.75	0.82	0.86	0.08	0.06	0.05
B. With parameter instabilities									
	OCMT								
20	4.04	5.07	5.96	0.73	0.85	0.92	0.06	0.08	0.11
40	3.78	4.90	5.67	0.70	0.84	0.91	0.02	0.04	0.05
100	3.54	4.62	5.26	0.66	0.81	0.88	0.01	0.01	0.02
	Lasso								
20	7.28	7.76	8.17	0.76	0.82	0.87	0.21	0.22	0.23
40	9.80	10.60	11.13	0.74	0.82	0.86	0.17	0.18	0.19
100	13.68	14.83	15.56	0.70	0.79	0.83	0.11	0.12	0.12
	A-Lasso								
20	5.49	5.95	6.30	0.65	0.72	0.78	0.15	0.15	0.16
40	7.55	8.28	8.76	0.64	0.73	0.79	0.12	0.13	0.14
100	10.71	11.85	12.58	0.63	0.73	0.78	0.08	0.09	0.09
	Boosting								
20	4.59	4.66	4.75	0.68	0.74	0.79	0.09	0.08	0.08
40	6.52	6.35	6.21	0.68	0.75	0.80	0.10	0.08	0.08
100	12.70	10.73	10.03	0.67	0.74	0.78	0.10	0.08	0.07

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 6 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**			**150**			**200**	
A. Without parameter instabilities									
	OCMT(Down-weighting only at the estimation stage)								
20	**31.76**	32.66	34.20	**28.53**	29.33	31.06	**26.19**	27.17	28.75
40	**29.13**	29.56	30.51	**26.72**	27.24	28.52	**32.05**	34.00	36.29
100	**29.25**	29.56	30.49	**27.93**	28.97	30.69	**28.94**	29.64	31.48
	OCMT(Down-weighting at the variable selection and estimation stages)								
20	**31.76**	32.61	35.08	**28.53**	29.25	32.19	**26.19**	27.36	30.67
40	**29.13**	29.46	31.95	**26.72**	27.21	31.50	**32.05**	34.27	41.13
100	**29.25**	30.20	33.85	**27.93**	29.46	36.72	**28.94**	31.19	40.14
	Lasso								
20	**31.82**	33.35	35.49	**28.59**	29.49	31.61	**26.25**	27.22	29.08
40	**29.48**	30.91	34.16	**26.35**	28.00	32.31	**31.78**	33.75	37.80
100	**30.63**	33.29	37.05	**28.33**	30.90	35.16	**29.13**	31.43	35.10
	A-Lasso								
20	**33.24**	34.72	37.09	**29.47**	30.44	32.85	**27.01**	27.82	30.13
40	**31.66**	32.87	36.30	**27.98**	30.08	34.72	**33.03**	35.09	38.89
100	**35.29**	37.89	41.49	**30.91**	33.92	38.70	**31.52**	34.37	38.13
	Boosting								
20	**32.69**	35.51	41.25	**29.51**	31.98	38.09	**26.77**	29.22	35.82
40	**30.67**	34.22	42.31	**27.20**	31.66	41.76	**32.90**	40.16	51.66
100	**33.68**	42.00	48.44	**29.28**	38.67	46.82	**29.98**	39.84	47.17
B. With parameter instabilities									
	OCMT(Down-weighting only at the estimation stage)								
20	35.87	**34.94**	35.45	31.18	**30.17**	31.02	29.12	**27.82**	28.91
40	32.42	**31.42**	31.70	30.03	**28.76**	29.41	35.28	**34.78**	36.46
100	33.31	**32.55**	32.83	31.66	**30.55**	31.45	32.72	**30.99**	32.22
	OCMT(Down-weighting at the variable selection and estimation stages)								
20	35.87	**35.62**	37.29	31.18	**31.01**	33.74	29.12	**28.46**	31.59
40	32.42	**32.09**	34.41	30.03	**29.51**	33.94	**35.28**	35.75	43.09
100	**33.3**1	33.48	37.29	**31.66**	32.09	39.04	**32.72**	33.36	44.01
	Lasso								
20	**36.84**	37.04	38.27	31.71	**31.27**	33.02	29.80	**28.75**	30.35
40	**33.43**	33.81	36.39	**30.44**	30.47	34.40	35.61	**35.40**	39.15
100	**34.95**	36.48	39.61	**32.64**	34.22	37.81	33.77	**33.67**	37.14
	A-Lasso								
20	38.48	**38.26**	39.62	32.62	**31.93**	34.02	30.40	**29.12**	31.29
40	**35.64**	35.85	38.73	32.41	**32.39**	36.87	37.06	**36.64**	40.51
100	**39.82**	41.19	44.04	**35.67**	37.48	41.55	**36.78**	36.82	40.54
	Boosting								
20	**36.57**	38.08	43.26	**31.77**	33.65	39.96	**29.29**	30.45	37.32
40	**33.78**	37.36	45.32	**29.97**	34.37	44.59	**35.43**	41.28	52.66
100	**36.09**	44.69	51.61	**31.78**	40.73	49.02	**33.01**	42.10	49.64

Notes: The reported results are averaged across four experiments (with/without dynamics and low/high fit) for models with and without parameter instabilities. See Section 6 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**	**150**	**200**	**100**	**150**	**200**
	A. Without parameter instabilities			B. With parameter instabilities		
	Lasso					
20	1.023	1.011	0.994	1.067	1.045	1.027
40	1.061	1.035	0.998	1.094	1.087	1.036
100	1.129	1.074	1.056	1.132	1.129	1.098
	A-Lasso					
20	1.067	1.043	1.021	1.100	1.069	1.046
40	1.135	1.106	1.039	1.164	1.156	1.077
100	1.277	1.176	1.147	1.269	1.236	1.202
	Boosting					
20	1.114	1.122	1.109	1.116	1.143	1.127
40	1.202	1.200	1.194	1.225	1.233	1.204
100	1.382	1.299	1.289	1.331	1.299	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 6 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, A-Lasso and boosting 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			63.378			68.835		
A-Lasso	62.857			65.142			71.586		
Boosting	64.663			98.763			222.698		
	Selected pair-wise panel DM tests								
	No down-weighting			Light down-weighting			Heavy down-weighting		
	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting
OCMT	-0.764	-4.063	-7.343	-3.318	-5.600	-19.053	-7.653	-9.722	-30.078
Lasso	-	-6.192	-6.081	-	-8.297	-18.519	-	-8.947	-29.476
A-Lasso	-	-	-3.215	-	-	-18.084	-	-	-29.197

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, A-Lasso and boosting 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	55.769	1.497
	Light	54.348	2.120
	Heavy	53.447	3.758
A-Lasso	No	55.122	1.187
	Light	53.586	1.610
	Heavy	53.055	2.819
Boosting	No	54.221	1.723
	Light	50.872	8.108
	Heavy	49.244	14.565

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, A-Lasso and boosting 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	22	3	6	2
	Light†	27	8	1	0
	Heavy‡	27	17	1	0
A-Lasso	No	25	6	3	0
	Light	28	13	0	0
	Heavy	28	24	0	0
Boosting	No	28	13	0	0
	Light	28	28	0	0
	Heavy	28	28	0	0

		MDFA	
	Down-weighting	OCMT outperforms	OCMT underperforms
Lasso	No	11	10
	Light	20	8
	Heavy	18	9
A-Lasso	No	16	7
	Light	21	5
	Heavy	21	6
Boosting	No	19	6
	Light	21	3
	Heavy	21	3

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, A-Lasso and boosting 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.	Emes	All	Adv.	Emes
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
Boosting	11.353	6.988	18.068	10.886	6.401	17.787	11.868	7.060	19.264
	Pair-wise Panel DM tests (All countries)								
	No down-weighting			Light down-weighting			Heavy down-weighting		
	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting
OCMT	0.220	-1.079	-0.445	0.486	-1.007	-0.195	-1.799	-2.441	-1.920
Lasso	-	-2.625	-1.322	-	-3.626	-1.790	-	-3.157	-0.894
A-Lasso	-	-	1.837	-	-	2.714	-	-	2.271

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, A-Lasso and boosting 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.	Emes	All	Adv.	Emes
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
Boosting	10.182	7.600	14.154	9.818	7.231	13.796	11.040	8.213	15.391
	Pair-wise Panel DM tests (All countries)								
	No down-weighting			Light down-weighting			Heavy down-weighting		
	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting
OCMT	-2.684	-4.200	-2.681	-2.137	-4.015	-2.933	-3.606	-4.789	-4.923
Lasso	-	-5.000	-0.430	-	-4.950	-2.317	-	-4.969	-4.588
A-Lasso	-	-	3.778	-	-	3.661	-	-	-0.252

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, A-Lasso and boosting 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-years-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
Boosting	No	11	1	22	3
	Light	11	1	22	3
	Heavy	17	1	16	1
	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
Boosting	No	23	2	10	0
	Light	25	1	8	0
	Heavy	32	4	1	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, A-Lasso and boosting 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	86.2	86.7	85.3	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
Boosting	No	86.7	87.1	86.0	8.17	8.39	3.06
	Light	86.6	86.6	86.6	7.43	5.48	7.30
	Heavy	85.4	85.6	85.1	5.66	6.06	1.44

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, A-Lasso and boosting 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	87.2	86.2	88.7	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
Boosting	No	87.3	85.8	89.7	0.63	0.19	1.44
	Light	87.6	86.7	89.1	2.23	3.77	1.05
	Heavy	86.2	84.9	88.1	1.47	2.07	0.79

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, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.

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.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

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

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., Y. Ke, and K. Wang (2020). Factor-adjusted regularized model selection. *Journal of Econometrics 216*(1), 71–85.

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.

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.

Hansen, C. and Y. Liao (2019). The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications. *Econometric Theory 35*(3), 465–509.

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.

Javanmard, A. and J. D. Lee (2020). A flexible framework for hypothesis testing in high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(3), 685–718.

Javanmard, A. and A. Montanari (2013). Model selection for high-dimensional regression under the generalized irrepresentability condition. *Advances in neural information processing systems 26*.

Javanmard, A. and A. Montanari (2018). Debiasing the lasso: Optimal sample size for gaussian designs.

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.

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

Lahiri, S. N. (2021). Necessary and sufficient conditions for variable selection consistency of the lasso in high dimensions.

Lee, J. D., Y. Sun, and J. E. Taylor (2015). On model selection consistency of regularized m-estimators.

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.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso.

Merlevède, F., M. Peligrad, and E. Rio (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*(3-4), 435–474.

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. (2023). 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.

Su, L., T. T. Yang, Y. Zhang, and Q. Zhou (2023). A one-covariate-at-a-time multiple testing approach to variable selection in additive models. *arXiv preprint arXiv:2204.12023*.

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.

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

July 15, 2024

This online theory supplement has three sections. Section S-1 provides the proofs of Theorems 1 to 3, and additional propositions and corollaries. Section S-2 establishes the main lemmas needed for the proof of the theorems in Section S-1. Section S-3 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 $\sup_{n \geq n_0}(f_n/g_n) \leq C_1$. 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 Proof of the Theorems

This section provides the proofs of Theorems 1 to 3. The proofs are based on lemmas presented in Section S-2. Among these, Lemmas S-2.6 and S-2.7 are key. For each covariate

$i = 1, 2, \cdots, N$, Lemmas S-2.6 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 (1), Lemma S-2.7 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 section we consider the following events:

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

where $\{\hat{\mathcal{J}}_i \text{ for } i = 1, 2, \cdots, N\}$ are the selection indicators defined by (6). $\mathcal{A}_0$ is the event of selecting the approximating model, defined by (9). $\mathcal{H}$ 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.

S-1.1 Proof of Theorem 1

To establish result (10), first note that $\mathcal{A}_0^c = \mathcal{H}^c \cup \mathcal{G}^c$ and hence ($\mathcal{H}^c$ denotes the complement of $\mathcal{H}$)

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

where $\mathcal{H}$ and $\mathcal{G}$ are given by (S.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-2.6, we can conclude that there exist sufficiently large finite 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 1, the number of signals is finite we can further conclude that

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

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 $\epsilon_i \geq 1/2$ and hence by Lemma S-2.6, we can conclude that there exist sufficiently large finite positive constants C_0 and C_1 such that for any $0 < \pi < 1$, $\Pr(\hat{\mathcal{J}}_i = 1) \leq \exp\left[-\frac{(1-\pi)^2\bar{\sigma}^2_{\eta_i,T}\bar{\sigma}^2_{x_i,T}c_p^2(N,\delta)}{2\bar{\omega}^2_{iy,T}(1+d_T)^2}\right] + \exp(-C_0T^{C_1})$, in which $\bar{\sigma}^2_{x_i,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}^2)$, $\bar{\omega}^2_{iy,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}^2y_t^2|\mathcal{F}_{t-1})$, $\bar{\sigma}^2_{\eta_i,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\eta_{it}^2)$, $\eta_{it} = y_t - \phi_{i,T}x_{it}$, and $\phi_{i,T}$ is defined in (3). Therefore,

$$\Pr(\mathcal{G}^c) \leq N\exp\left[-\frac{\mathcal{X}_{NT}(1-\pi)^2c_p^2(N,\delta)}{2(1+d_T)^2}\right] + N\exp(-C_0T^{C_1}),$$

where $\mathcal{X}_{NT} = \inf_{i\in\{k+k^*+1,k+k^*+2,\cdots,N\}}\frac{\bar{\sigma}^2_{\eta_i,T}\bar{\sigma}^2_{x_i,T}}{\bar{\omega}^2_{iy,T}}$. By result (II) of Lemma S-3.2 in online theory supplement we can further conclude that for any $0 < \pi < 1$,

$$\Pr(\mathcal{G}^c) = O\left(N^{1-\mathcal{X}_{NT}\left(\frac{1-\pi}{1+d_T}\right)^2\delta}\right) + O\left[N\exp(-C_0T^{C_1})\right], \tag{S.4}$$

Using (S.3) and (S.4) in (S.2), we obtain $\Pr(\mathcal{A}_0^c) = O\left(N^{1-\mathcal{X}_{NT}\left(\frac{1-\pi}{1+d_T}\right)^2\delta}\right) + O\left[N\exp(-C_0T^{C_1})\right]$ and $\Pr(\mathcal{A}_0) = 1 - O\left(N^{1-\mathcal{X}_{NT}\left(\frac{1-\pi}{1+d_T}\right)^2\delta}\right) - O\left[N\exp(-C_0T^{C_1})\right]$, which completes the proof.

S-1.2 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-2.7 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, $\Pr\left(T^{\frac{1-d}{2}}\left\|\hat{\boldsymbol{\gamma}}_T-\boldsymbol{\gamma}_T^*\right\|>B_\varepsilon\right)<\varepsilon$ for all $T>T_\varepsilon$, and we conclude that

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

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 $\left\|\hat{\boldsymbol{\gamma}}_T-\boldsymbol{\gamma}_T^\diamond\right\|=O_P\left(T^{\frac{d-1}{2}}\right)$, which completes the proof.

S-1.3 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 $B>0$,

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

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

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

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-2.7 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 $\lim_{T\to\infty}\Pr\left(\mathcal{A}_0^c\right)=0$. Hence, for any $\varepsilon>0$, there exists $B_\varepsilon>0$ and $T_\varepsilon>0$ such that $\Pr\left(T^{\frac{1}{2}}\left|D_T\right|>B_\varepsilon|\mathcal{A}_0\right)+\Pr\left(\mathcal{A}_0^c\right)<\varepsilon$, for all $T>T_\varepsilon$. Therefore, $\Pr\left(T^{\frac{1}{2}}\left|D_T\right|>B_\varepsilon\right)<\varepsilon$ for all $T>T_\varepsilon$, and we conclude that

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

Furthermore, by Lemma S-2.7, $\bar{\Delta}_{\beta,T}$ is non-negative. 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,

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

with $\bar{\Delta}_{\beta,T}^*\geq 0$ which completes the proof.

S-1.4 Propositions and corollaries

Proposition S.1 *Suppose the target variable y_t is generated according to (1), and Assumptions 1-4 hold. Consider the following regression equation:*

$$y_t=\sum_{i=1}^k x_{it}\gamma_{iT}+\eta_t=\mathbf{x}'_{k,t}\boldsymbol{\gamma}_T+\eta_t,\ t=1,2....,T \tag{S.5}$$

where $\boldsymbol{\gamma}_T$ *is defined by*

$$\boldsymbol{\gamma}_T = \arg\min_{\mathbf{b}} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(y_t - \mathbf{x}'_{k,t}\mathbf{b}\right)^2. \tag{S.6}$$

Then there exists a positive constnt $\epsilon \geq 1/2$*, such that*

$$\boldsymbol{\gamma}_T = \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right)\right]^{-1} T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{k}\mathbb{E}\left(\mathbf{x}_{k,t}x_{it}\right)\mathbb{E}\left(\beta_{it}\right) + d_T\boldsymbol{\tau}_k,$$

where $d_T = O\left(T^{-\epsilon}\right)$*, and* $\boldsymbol{\tau}_k$ *is the* $k \times 1$ *vector of ones. Also, if the expected value of* β_{it} *for* $i = 1, 2, \cdots, k$ *are time-invariant, i.e.,* $\mathbb{E}\left(\beta_{it}\right) = \beta_i$*, then* $\gamma_{iT} = \beta_i + d_T$ *for* $i = 1, 2, \cdots, k$ *and there exists* $\varrho \geq 1$ *such that*

$$\mathbb{E}\left(\eta_t^2\right) = \Delta_{\beta,t} + \mathbb{E}(u_t^2) + e_T,$$

where $e_T = O\left(T^{-\varrho}\right)$

$$\Delta_{\beta,t} = \sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ijt,x}\sigma_{ij,\beta} = tr\left(\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\boldsymbol{\Omega}_{\beta,t}\right) \geq 0, \tag{S.7}$$

$\boldsymbol{\Sigma}_{\mathbf{x}_k,t} \equiv (\sigma_{ijt,x})$*,* $\boldsymbol{\Omega}_{\beta,t} \equiv (\sigma_{ijt,\beta})$*, for* $i, j = 1, 2, \cdots, k$*,* $\sigma_{ijt,x} = \mathbb{E}\left(x_{it}x_{jt}\right)$*, and* $\sigma_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)\right]$*. Alternatively, if the covariance matrix of the signals are time-invariant, i.e.,* $\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right) = \boldsymbol{\Sigma}_{x_k}$*, then* $\gamma_{iT} = \bar{\beta}_{iT} + d_T$ *for* $i = 1, 2, \cdots, k$*, where* $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\beta_{it}\right)$*, and there exists* $\varrho \geq 1$ *such that*

$$\mathbb{E}\left(\eta_t^2\right) = \Delta^*_{\beta,t} + \mathbb{E}(u_t^2) + e_T$$

where $e_T = O\left(T^{-\varrho}\right)$

$$\Delta^*_{\beta,t} = \sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij,x}\sigma^*_{ijt,\beta} = tr\left(\boldsymbol{\Omega}^*_{\beta,t}\boldsymbol{\Sigma}_{\mathbf{x}_k}\right) \geq 0, \tag{S.8}$$

$\boldsymbol{\Omega}^*_{\beta,t} \equiv \left(\sigma^*_{ijt,\beta}\right)$*, for* $i, j = 1, 2, \cdots, k$*, and* $\sigma^*_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \bar{\beta}_{i,T})(\beta_{jt} - \bar{\beta}_{j,T})\right]$*.*

Remark 5 *Proposition S.1 shows that in a linear regression model that does not consider parameter instability, the deviation of each coefficient from the simple time-average of the corresponding coefficients in the DGP approaches zero. Moreover,* $\Delta_{\beta,t} \geq 0$ *and* $\Delta^*_{\beta,t} \geq 0$ *represent the costs, in mean squared error sense, of neglecting parameter instability.*

Corollary S.1 *Let* y_t *for* $t = 1, 2, \cdots, T$ *be generated by (1), and consider the active set* $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$ *which contains* k *signals,* k^*_T *pseudo-signals, and* $N - k - k^*_T$ *noise variables. Suppose Assumptions 1-3 hold and the noise variables,* x_{it} $i = k + k^*_T + 1, k + k^*_T + 2, \cdots, N$*,*

are independent of the target, y_t, and have time-invaraint unconditional variances, $\mathbb{V}\left(x_{it}^2\right)=\sigma_i^2$ for $i=k+k_T^*+1,k+k_T^*+2,\cdots,N$, and $N=\ominus(T^\kappa)$ with $\kappa>0$. Then, there exist finite positive constants C_0 and C_1 such that, for any π in $(0,1)$ and any null sequence $d_T>0$, the probability of selecting the approximating model $\mathcal{A}_0$, defined by (9), by the OCMT procedure with the critical value function $c_p(N,\delta)$ given by (5), for some $\delta>0$, is given by

$$\Pr(\mathcal{A}_0)=1-O\left[T^{\kappa\left(1-\left(\frac{1-\pi}{1+d_T}\right)^2\delta\right)}\right]-O\left[T^\kappa\exp\left(-C_0T^{C_1}\right)\right]. \tag{S.9}$$

Remark 6 *Corollary S.1 shows that if we further assume that the noise variables are independent of y_t and their variance does not change over time, then for any $\delta>1$, the OCMT procedure consistently selects the approximating model.*

S-1.5 Proof of propositions and corollaries

Proof of Proposition S.1 *Since the objective function for $\boldsymbol{\gamma}_T$ is convex and, by Assumption 4, $T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}_{k,t}'\right)$ is invertible, then by the first-order condition of the minimization in (S.6) we have*

$$\boldsymbol{\gamma}_T=\left[T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}_{k,t}'\right)\right]^{-1}T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}y_t\right).$$

Substituting y_t from (1), now yields

$$\boldsymbol{\gamma}_T=\left[T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}_{k,t}'\right)\right]^{-1}T^{-1}\sum_{t=1}^T\sum_{i=1}^k\mathbb{E}\left(\mathbf{x}_{k,t}x_{it}\beta_{it}\right)+\left[T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}_{k,t}'\right)\right]^{-1}T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}u_{it}\right).$$

By part (c) of Assumptions 2, all the elements of the $k\times1$ vector $T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}u_{it}\right)$ are $O\left(T^{-\epsilon}\right)$ for some $\epsilon\geq1/2$. Moreover, by Assumptions 3 and 4, all the element of $k\times k$ matrix $\left[T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}_{k,t}'\right)\right]^{-1}$ are finite fixed numbers. Since, by Assumption 1, the number of signals, k, is a finite fixed number, we can further conclude that all the elements of $k\times1$ vector,

$$\left[T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}_{k,t}'\right)\right]^{-1}T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}u_{it}\right),$$

are $O\left(T^{-\epsilon}\right)$ for some $\epsilon\geq1/2$ and consequently we can write

$$\boldsymbol{\gamma}_T=\left[T^{-1}\sum_{t=1}^T\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}_{k,t}'\right)\right]^{-1}T^{-1}\sum_{t=1}^T\sum_{i=1}^k\mathbb{E}\left(\mathbf{x}_{k,t}x_{it}\beta_{it}\right)+d_T\boldsymbol{\tau}_k,$$

where $d_T=O\left(T^{-\epsilon}\right)$ for some $\epsilon\geq1/2$ and $\boldsymbol{\tau}_k$ is the $k\times1$ vector of ones. By Assumption

1, β_{it} is independent of x_{jt} for all $i, j = 1, 2, \cdots, k$, therefore,

$$\boldsymbol{\gamma}_T = \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \mathbf{x}'_{k,t} \right) \right]^{-1} T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{k} \mathbb{E}\left(\mathbf{x}_{k,t} x_{it} \right) \mathbb{E}\left(\beta_{it} \right) + d_T \boldsymbol{\tau}_k.$$

Consider first the case where $\mathbb{E}(\beta_{it})$ is time-invariant and set $\mathbb{E}(\beta_{it}) = \beta_i$. Then

$$\begin{aligned}
\boldsymbol{\gamma}_T &= \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \mathbf{x}'_{k,t} \right) \right]^{-1} T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{k} \mathbb{E}\left(\mathbf{x}_{k,t} x_{it} \right) \beta_i + d_T \boldsymbol{\tau}_k \\
&= \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \mathbf{x}'_{k,t} \right) \right]^{-1} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \sum_{i=1}^{k} x_{it} \beta_i \right) + d_T \boldsymbol{\tau}_k \\
&= \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \mathbf{x}'_{k,t} \right) \right]^{-1} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \mathbf{x}'_{k,t} \boldsymbol{\beta} \right) + d_T \boldsymbol{\tau}_k \\
&= \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \mathbf{x}'_{k,t} \right) \right]^{-1} \left[T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\mathbf{x}_{k,t} \mathbf{x}'_{k,t} \right) \right] \boldsymbol{\beta} + d_T \boldsymbol{\tau}_k = \boldsymbol{\beta} + d_T \boldsymbol{\tau}_k,
\end{aligned}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)'$. So, in this case γ_{iT} would converge to the expected value of β_{it} at $T \to \infty$. Moreover,

$$\eta_t = y_t - \sum_{i=1}^{k} x_{it} (\beta_i + d_T).$$

By substituting for y_t from (1), we have

$$\eta_t = \sum_{i=1}^{k} x_{it} \left(\beta_{it} - \beta_i \right) + u_t + d_T \sum_{i=1}^{k} x_{it}.$$

Therefore, by Assumptions 1 and 2,

$$\mathbb{E}\left(\eta_t^2 \right) = \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_{ijt,x} \sigma_{ijt,\beta} + \mathbb{E}(u_t^2) + e_T.$$

where $e_T = O\left(T^{-\varrho}\right)$ for some $\varrho \geq 1$, $\sigma_{ijt,x} = \mathbb{E}\left(x_{it} x_{jt}\right)$, $\sigma_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j) \right]$. We further have

$$\Delta_{\beta,t} = \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_{ijt,x} \sigma_{ijt,\beta} = tr\left(\boldsymbol{\Omega}_{\beta,t} \ \boldsymbol{\Sigma}_{\mathbf{x}_k,t} \right),$$

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

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

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 $\Delta_{\beta,t} \geq 0$.

Consider now a second case where $\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right)$ *is time-invariant and set* $\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right) = \mathbf{\Sigma}_{x_k}$. *Then*

$$\begin{aligned}
\boldsymbol{\gamma}_T &= \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right)\right]^{-1}\sum_{i=1}^{k}\mathbb{E}\left(\mathbf{x}_{k,t}x_{it}\right)\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\beta_{it}\right)\right] + d_T\boldsymbol{\tau}_k \\
&= \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right)\right]^{-1}\sum_{i=1}^{k}\mathbb{E}\left(\mathbf{x}_{k,t}x_{it}\bar{\beta}_{iT}\right) + d_T\boldsymbol{\tau}_k \\
&= \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right)\right]^{-1}\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\bar{\boldsymbol{\beta}}_T\right) + d_T\boldsymbol{\tau}_k \\
&= \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right)\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{x}_{k,t}\mathbf{x}'_{k,t}\right)\right]\bar{\boldsymbol{\beta}}_T + d_T\boldsymbol{\tau}_k = \bar{\boldsymbol{\beta}}_T + d_T\boldsymbol{\tau}_k
\end{aligned}$$

where $\bar{\boldsymbol{\beta}}_T = \left(\bar{\beta}_{1T}, \bar{\beta}_{2T}, \cdots, \bar{\beta}_{kT}\right)'$ *and* $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^{k}\mathbb{E}\left(\beta_{it}\right)$. *So, in this case* γ_{iT} *would converge to the simple average of expected value of* β_{it} *across time. Moreover,*

$$\eta_t = y_t - \sum_{i=1}^{k} x_{it}(\bar{\beta}_{iT} + d_T) \tag{S.10}$$

By substituting for y_t *from (1), we have*

$$\eta_t = \sum_{i=1}^{k} x_{it}\left(\beta_{it} - \bar{\beta}_{iT}\right) + u_t + d_T\sum_{i=1}^{k} x_{it}. \tag{S.11}$$

Therefore, by Assumptions 1 and 2,

$$\mathbb{E}\left(\eta_t\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij,x}\sigma^*_{ijt,\beta} + \mathbb{E}(u_t^2) + e_T, \tag{S.12}$$

where $e_T = O\left(T^{-\varrho}\right)$ *for some* $\varrho \geq 1$, $\sigma_{ij,x} = \mathbb{E}\left(x_{it}x_{jt}\right)$ *and* $\sigma^*_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \bar{\beta}_{i,T})(\beta_{jt} - \bar{\beta}_{j,T})\right]$. *We further have*

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

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

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

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* $\Delta^*_{\beta,t} \geq 0$.

Proof of Corollary S.1 *By Theorem 1, we have that under Assumptions 1-3, there exist finite positive constants C_0 and C_1 such that, for any $0 < \pi < 1$, the probability of selecting the approximating model $\mathcal{A}_0$, as defined by (9), is given by*

$$\Pr(\mathcal{A}_0) = 1 - O\left[T^{\kappa\left(1-\mathcal{X}_{NT}\left(\frac{1-\pi}{1+d_T}\right)^2\delta\right)}\right] - O\left[T^{\kappa}\exp\left(-C_0T^{C_1}\right)\right], \tag{S.13}$$

where

$$\mathcal{X}_{NT} = \inf_{i\in\{k+k^*+1,\cdots,N\}} \frac{\bar{\sigma}^2_{\eta_i,T}\bar{\sigma}^2_{x_i,T}}{\bar{\omega}^2_{iy,T}}.$$

with $\bar{\sigma}^2_{x_i,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}^2)$, $\bar{\omega}^2_{iy,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}^2y_t^2|\mathcal{F}_{t-1})$, $\bar{\sigma}^2_{\eta_i,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\eta_{it}^2)$, $\eta_{it} = y_t - \phi_{i,T}x_{it}$, and $\phi_{i,T}$ is defined in (3). Note that,

$$\begin{aligned}\bar{\sigma}^2_{\eta_i,T} &= T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(y_t^2\right) + \phi^2_{i,T}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(x_{it}^2\right)\right] - 2\phi_{i,T}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(x_{it}y_t\right)\right]\\ &= \bar{\sigma}^2_{y,T} + \phi^2_{i,T}\bar{\sigma}^2_{x_i,T} - 2\phi_{i,T}\bar{\theta}_{i,T} = \bar{\sigma}^2_{y,T} - \phi^2_{i,T}\bar{\sigma}^2_{x_i,T}.\end{aligned}$$

But, x_{it} for all $i \in \{k+k_T^*+1, k+k_T^*+2,\cdots,N_T\}$ are independent of y_t and hence $\phi_{i,T} = 0$. Consequently, $\bar{\sigma}^2_{\eta_i,T} = \bar{\sigma}^2_{y,T}$ for $i \in \{k+k_T^*+1, k+k_T^*+2,\cdots,N_T\}$. Moreover, since x_{it} for all $i \in \{k+k_T^*+1, k+k_T^*+2,\cdots,N_T\}$ are independent of y_t, we can write

$$\bar{\omega}^2_{iy,T} = T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(x_{it}^2|\mathcal{F}_{t-1}\right)\mathbb{E}\left(y_t^2|\mathcal{F}_{t-1}\right) = T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(x_{it}^2\right)\mathbb{E}\left(y_t^2\right),$$

for $i \in \{k+k_T^*+1, k+k_T^*+2,\cdots,N_T\}$. Therefore,

$$\mathcal{X}_{NT} = \inf_{i\in\{k+k^*+1,\cdots,N\}} \frac{\bar{\sigma}^2_{y,T}\bar{\sigma}^2_{x_i,T}}{T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(x_{it}^2\right)\mathbb{E}\left(y_t^2\right)},$$

In cases where $\mathbb{E}(x_{it}^2)$ for $i \in \{k+k_T^*+1, k+k_T^*+2,\cdots,N_T\}$ are time-invariant, we can conclude that $\mathcal{X}_{NT} = 1$ and hence the probability of selecting the approximating model is given by

$$\Pr(\mathcal{A}_0) = 1 - O\left[T^{\kappa\left(1-\left(\frac{1-\pi}{1+d_T}\right)^2\delta\right)}\right] - O\left[T^{\kappa}\exp\left(-C_0T^{C_1}\right)\right],$$

as required. Note that $d_T \to 0$ and $T \to \infty$ and π is an arbitrary constant between zero and one.

Since π is an arbitrary constant between zero and one, result (S.9) of Corollary S.1 implies that for any $\delta > 1$, we can select an approximating model with probability approaching one as N and T grows to infinity.

S-2 Main lemmas

Lemma S-2.1 *Let y_t be a target variable generated by equation (1), and x_{it} be a covariate in the active set $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. Under Assumptions 1, and 2, we have*

$$\mathbb{E}\left[y_t x_{it} - \mathbb{E}(y_t x_{it})|\mathcal{F}_{t-1}\right] = 0, \text{ for } i = 1, 2, \cdots, N,$$

and

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

Proof. For $i = 1, 2, \cdots, N$, we have

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

By Assumption 2, $\mathbb{E}(\beta_{jt}|\mathcal{F}_{t-1}) = \mathbb{E}(\beta_{jt})$, $\mathbb{E}(x_{jt}x_{it}|\mathcal{F}_{t-1}) = \mathbb{E}(x_{jt}x_{it})$, and $\mathbb{E}(u_t x_{it}|\mathcal{F}_{t-1}) = \mathbb{E}(u_t x_{it})$. Therefore,

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

Also to establish the last result, note that y_t can be written as

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

where $\mathbf{x}_{k,t} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, and $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \cdots, \beta_{kt})'$. Hence,

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

■

Lemma S-2.2 *Let y_t be a target variable generated by equation (1). Under Assumptions 3-1, 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

$$|y_t| \leq \textstyle\sum_{j=1}^{k} |\beta_{jt} x_{jt}| + |u_t|.$$

Therefore,

$$\Pr(|y_t| > \alpha) \leq \textstyle\sum_{j=1}^{k} |\beta_{jt} x_{jt}| + |u_t| > \alpha),$$

and by Lemma S-3.3 for any $0 < \pi_i < 1$, $i = 1, 2, \cdots, k+1$, with $\sum_{i=1}^{k+1} \pi_j = 1$, we can further write

$$\Pr(|y_t| > \alpha) \leq \sum_{j=1}^{k} \Pr(|\beta_{jt} x_{jt}| > \pi_j \alpha) + \Pr(|u_t| > \pi_{k+1} \alpha).$$

Moreover, by Lemma S-3.4, we have

$$\Pr(|\beta_{jt} x_{jt}| > \pi_j \alpha) \leq \Pr[|x_{jt}| > (\pi_j \alpha)^{1/2}] + \Pr[|\beta_{jt}| > (\pi_i \alpha)^{1/2}],$$

and hence

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

Therefore, under Assumptions 3-1, 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-2.3 *Let x_{it} be a covariate in the active set, $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. Suppose Assumptions 2-3 hold and $\zeta_T = \ominus(T^{\lambda})$ for some $\lambda > 0$. Then, if $0 < \lambda \leq (s+2)/(s+4)$, for any $0 < \pi < 1$,*

$$\Pr(|\mathbf{x}_i' \mathbf{x}_j - \mathbb{E}(\mathbf{x}_i' \mathbf{x}_j)| > \zeta_T) \leq \exp\left(-\frac{(1-\pi)^2 \zeta_T^2}{2T \bar{\omega}_{ij,T}^2} \right),$$

where, $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$ and $\bar{\omega}_{ij,T}^2 = T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(x_{it}^2 x_{jt}^2 | \mathcal{F}_{t-1} \right)$. Also, if $\lambda > (s+2)/(s+4)$, there exists a finite positive constant C_1,

$$\Pr(|\mathbf{x}_i' \mathbf{x}_j - \mathbb{E}(\mathbf{x}_i' \mathbf{x}_j)| > \zeta_T) \leq \exp\left(-C_1 \zeta_T^{s/(s+1)} \right),$$

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

Proof. Note that $[\mathbf{x}_i' \mathbf{x}_j - \mathbb{E}(\mathbf{x}_i' \mathbf{x}_j)] = \sum_{t=1}^{T} [x_{it} x_{jt} - \mathbb{E}(x_{it} x_{jt})]$, for all i and j. By Assumption 2 we have

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

for $i, j = 1, 2, \cdots, N$. Moreover, by Assumption 3, for all $i = 1, 2, \cdots, N$ and $\alpha > 0$, there exist some finite positive constants C_0 and C_1 such that,

$$\sup_t \Pr(|x_{it}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s).$$

Therefore, by Lemma S-3.7, for all $i, j = 1, 2, \cdots, N$ and $\alpha > 0$,

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

Hence, by Lemma S-3.1, if $0 < \lambda \leq (s+2)/(s+4)$, for any $0 < \pi < 1$,

$$\Pr(|\mathbf{x}_i'\mathbf{x}_j - \mathbb{E}(\mathbf{x}_i'\mathbf{x}_j)| > \zeta_T) \leq \exp\left(-\frac{(1-\pi)^2\zeta_T^2}{2T\bar{\omega}_{ij,T}^2}\right).$$

Moreover, if $\lambda > (s+2)/(s+4)$, then there exists a finite positive constant C_1, such that

$$\Pr(|\mathbf{x}_i'\mathbf{x}_j - \mathbb{E}(\mathbf{x}_i'\mathbf{x}_j)| > \zeta_T) \leq \exp\left(-C_1\zeta_T^{s/(s+1)}\right).$$

■

Lemma S-2.4 *Let y_t be a target variable generated by the DGP given by (1) and x_{it} be a covariate in the active set, $\{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. Suppose Assumptions 1-3 hold and $\zeta_T = \ominus(T^\lambda)$ for some $\lambda > 0$. Then, if $0 < \lambda \leq (s+4)/(s+8)$, for any $0 < \pi < 1$,*

$$\Pr(|\mathbf{x}_i'\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp\left(-\frac{(1-\pi)^2\zeta_T^2}{2T\bar{\omega}_{iy,T}^2}\right),$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$, $\theta_{i,T} = T\bar{\theta}_{i,T} = \mathbb{E}(\mathbf{x}_i'\mathbf{y})$ and $\bar{\omega}_{iy,T}^2 = T^{-1}\sum_{t=1}^T \mathbb{E}\left(x_{it}^2 y_t^2 | \mathcal{F}_{t-1}\right)$. Also, if $\lambda > (s+4)/(s+8)$, there exists a finite positive constant C_1 such that

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

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

Proof. Note that $[\mathbf{x}_i'\mathbf{y} - \theta_{i,T}] = \sum_{t=1}^T [x_{it}y_t - \mathbb{E}(x_{it}y_t)]$, for all i. By Lemma S-2.1

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

for $i = 1, 2, \cdots, N$. Moreover, by Assumption 3, for all $i = 1, 2, \cdots, N$ and $\alpha > 0$, there exist some finite positive constants C_0 and C_1 such that,

$$\sup_t \Pr(|x_{it}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s).$$

Also, by Lemma S-2.2, there exist some finite positive constants C_0 and C_1 such that,

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

Therefore, by Lemma S-3.7, for all $i = 1, 2, \cdots, N$ and $\alpha > 0$,

$$\sup_t \Pr(|x_{it} y_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^{s/4}).$$

Hence, by Lemma S-3.1, if $0 < \lambda \leq (s+4)/(s+8)$, for any $0 < \pi < 1$,

$$\Pr(|\mathbf{x}_i' \mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp\left(-\frac{(1-\pi)^2 \zeta_T^2}{2T\bar{\omega}_{iy,T}^2}\right).$$

Moreover, if $\lambda > (s+4)/(s+8)$, there exists a finite positive constant C_1,

$$\Pr(|\mathbf{x}_i' \mathbf{y}_j - \theta_{i,T}| > \zeta_T) \leq \exp\left(-C_1 \zeta_T^{s/(s+1)}\right).$$

■

Lemma S-2.5 *Let y_t be a target variable generated by equation (1) and x_{it} be a covariate in the active set, $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. Suppose Assumptions 1-3 hold and $\zeta_T = \ominus(T^\lambda)$ for some $\lambda > 0$. Consider the projection regression of y_t on x_{it} as*

$$y_t = \phi_{i,T} x_{it} + \eta_{it},$$

where the projection coefficient $\phi_{i,T}$ is given by (3). Then, if $0 < \lambda \leq (s+4)/(s+8)$, there exist sufficiently large positive constants C_0, C_1 and C_2 such that

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

where $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \cdots, \eta_{iT})'$ and $\mathbf{M}_{x_i} = \mathbf{I} - T^{-1}\mathbf{x}_i (T^{-1}\mathbf{x}_i'\mathbf{x}_i)^{-1}\mathbf{x}_i'$ with $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$. Also, if $\lambda > (s+4)/(s+8)$, there exist sufficiently large positive constants C_0, C_1 and C_2 such that

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

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

Proof. Note that $\boldsymbol{\eta}_i' \mathbf{M}_{x_i} \boldsymbol{\eta}_i = \mathbf{y}' \mathbf{M}_{x_i} \mathbf{y}$, where $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$. By Assumption 2, we have

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

for $i = 1, 2, \cdots, N$. By Lemma S-2.1, we also have

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

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

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

Moreover, by Assumption 3, for all $i = 1, 2, \cdots, N$ and $\alpha > 0$, there exist some finite positive constants C_0 and C_1 such that,

$$\sup_t \Pr(|x_{it}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s).$$

Also, by Lemma S-2.2, there exist some finite positive constants C_0 and C_1 such that,

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

Therefore by Lemma S-3.20, we can conclude that there exist sufficiently large positive constants C_0, C_1, and C_2 such that if $0 < \lambda \leq (s+4)/(s+8)$, then

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

and if $\lambda > (s+4)/(s+8)$, then

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

for all $i = 1, 2, \cdots, N$. ■

Lemma S-2.6 *Let y_t be a target variable generated by equation (1) and x_{it} be a covariate in the active set, $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. Suppose Assumptions 1-3 hold and consider the projection regression of y_t on x_{it} as*

$$y_t = \phi_{i,T} x_{it} + \eta_{it}, \tag{S.14}$$

where $\phi_{i,T}$ is given in (3). Define,

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

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$, $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \cdots, \eta_{iT})'$ and $\mathbf{M}_{x_i} = \mathbf{I} - T^{-1} \mathbf{x}_i (T^{-1} \mathbf{x}_i' \mathbf{x}_i)^{-1} \mathbf{x}_i'$. Then, there exist sufficiently large finite positive constants C_0 and C_1 such that for any $0 < \pi < 1$, any null sequence $d_T > 0$, and $\epsilon_i \geq \frac{1}{2}$

$$\Pr\left[|t_{i,T}| > c_p(N, \delta) | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-\frac{(1-\pi)^2 \bar{\sigma}_{\eta_i,T}^2 \bar{\sigma}_{x_i,T}^2 c_p^2(N,\delta)}{2\bar{\omega}_{iy,T}^2 (1+d_T)^2}\right] + \exp(-C_0 T^{C_1}),$$

where $c_p(N, \delta)$ is defined by (5), $\theta_{i,T} = T\bar{\theta}_{i,T} = \mathbb{E}(\mathbf{x}_i' \mathbf{y})$, $\bar{\sigma}_{\eta_i,T}^2 = T^{-1} \sum_{t=1}^T \mathbb{E}\left(\eta_{it}^2\right)$, $\bar{\sigma}_{x_i,T}^2 = T^{-1} \sum_{t=1}^T \mathbb{E}\left(x_{it}^2\right)$ and $\bar{\omega}_{iy,T}^2 = T^{-1} \sum_{t=1}^T \mathbb{E}\left(x_{it}^2 y_t^2 | \mathcal{F}_{t-1}\right)$. Also, if $c_p(N, \delta) = o(T^{1/2-\vartheta-c})$ for any $0 \leq \vartheta_i < 1/2$ and a finite positive constant c, then, there exist some finite positive constants C_0

and C_1 such that

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

Proof. We have $|t_{i,T}| = \mathcal{A}_{iT}\mathcal{B}_{iT}$, where,

$$\mathcal{A}_{iT} = \frac{|T^{-1/2}\mathbf{x}_i'\mathbf{y}|}{\bar{\sigma}_{\eta_{i,T}}\bar{\sigma}_{x_{i,T}}},$$

and

$$\mathcal{B}_{iT} = \frac{\bar{\sigma}_{\eta_i,T}\bar{\sigma}_{x_i,T}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_{x_i}\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{x}_i}}.$$

In the first case where $\theta_{i,T} = \ominus(T^{1-\epsilon_i})$ for some $\epsilon_i \geq 1/2$, by using Lemma S-3.4 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-3.6,

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

where $\mathcal{M}_{iT} = |(T^{-1}\mathbf{x}_i'\mathbf{x}_i)/\bar{\sigma}^2_{x_i,T} - 1|$ and $\mathcal{R}_{iT} = |(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_{x_i}\boldsymbol{\eta}_i)/\bar{\sigma}^2_{\eta_i,T} - 1|$. By using Lemmas S-3.3 and S-3.4 , 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

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

Therefore, by Lemma S-2.3, 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

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

Therefore, by Lemma S-2.5, 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

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

Note that since $\epsilon_i \geq 1/2$ and $c_p(N,\delta) \to \infty$ as N and consequently T goes to infinity, 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-2.4, for any $0 < \pi < 1$,

$$\Pr\left[|\mathbf{x}_i'\mathbf{y}| > \tfrac{\bar{\sigma}_{\eta_i,T}\bar{\sigma}_{x_i,T}}{1+d_T} T^{1/2} c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-\tfrac{(1-\pi)^2 \bar{\sigma}^2_{\eta_i,T}\bar{\sigma}^2_{x_i,T} c_p^2(N,\delta)}{2\bar{\omega}^2_{iy,T}(1+d_T)^2}\right].$$

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 and C_1 such that for any $0 < \pi < 1$

$$\Pr\left[|t_{i,T}| > c_p(N,\delta) | \theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-\frac{(1-\pi)^2 \bar{\sigma}^2_{\eta_i,T}\bar{\sigma}^2_{x_i,T} c_p^2(N,\delta)}{2\bar{\omega}^2_{iy,T}(1+d_T)^2}\right] + \exp(-C_0 T^{C_1}).$$

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-3.8,

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

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 $0 \leq \vartheta < 1/2$, $|\theta_{i,T}| - \bar{\sigma}_{\eta_i,T}\bar{\sigma}_{x_i,T}[(1+d_T)T]^{1/2}c_p(N,\delta) = \ominus(T^{1-\vartheta_i}) > 0$ and by Lemma S-3.5, we have

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

Therefore, by Lemma S-2.4, there exist some finite positive constants C_0 and C_1 such that,

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

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 $\mathcal{B}_{iT}$,

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

where $\mathcal{M}_{iT} = |(T^{-1}\mathbf{x}_i'\mathbf{x}_i)/\bar{\sigma}^2_{x_i,T} - 1|$ and $\mathcal{R}_{iT} = |(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_{x_i}\boldsymbol{\eta}_i)/\bar{\sigma}^2_{\eta_i,T} - 1|$. By using Lemmas S-3.3 and S-3.4 , 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

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

So, by Lemma S-2.3, we know that there exist some constants C_0 and C_1 such that,

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

Similarly,

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

Also note that

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

Therefore, by Lemma S-2.5, there exist some constants C_0 and C_1 such that,

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

Similarly,

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

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

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

■

Lemma S-2.7 *Suppose y_t are generated by*

$$y_t = \sum_{i=1}^{k} x_{it}\beta_{it} + u_t \text{ for } t = 1, 2, \cdots, T, \tag{S.15}$$

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*

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{y}\right), \tag{S.16}$$

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*

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{W}\hat{\boldsymbol{\gamma}}_T. \tag{S.17}$$

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 1-3 holds. Now,*

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

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

where $\boldsymbol{\gamma}_T^* = (\boldsymbol{\beta}', \ \mathbf{0}_{l_T}')'$ *and* $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)'$. *Under Assumption 5 we also have*

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

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

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

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

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 5 also holds, then*

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

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.15) 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,

$$\mathbf{y} = \mathbf{X}_k\boldsymbol{\beta} + \mathbf{R}\boldsymbol{\tau} + \mathbf{u}, \tag{S.18}$$

where $\mathbf{X}_k = (\mathbf{x}_{k1}, \mathbf{x}_{k2}, \cdots, \mathbf{x}_{kT})'$, $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \quad \mathbf{r}_T)'$ and $\mathbf{u} = (u_1, u_2, \cdots, u_T)'$. By substituting (S.18) into (S.16), 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

$$\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^* = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right). \tag{S.19}$$

We can further write

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

Hence, by the sub-additive property of norms and Lemma S-3.9, we have

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

By Assumption 2

$$\left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\| = \left\|T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t u_t)\right\| = O\left(T^{-\frac{2\epsilon-d}{2}}\right),$$

where $\epsilon \geq 1/2$.

Assumptions 2 and 3 imply that $\mathbf{W}$ and $\mathbf{u}$ satisfy condition (i) and (ii) of Lemma S-3.12, and by Lemmas S-3.12 and S-3.13,

$$\left\|T^{-1}\left[\mathbf{W}'\mathbf{u} - \mathbb{E}\left(\mathbf{W}'\ \mathbf{u}\right)\right]\right\| = O_p\left(T^{-\frac{1-d}{2}}\right).$$

Similarly,

$$\left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{W}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{W}\right)\right]\right\|_F = O_p\left(T^{-(1/2-d)}\right),$$

and since $l_T = \ominus(T^d)$ with $0 \leq d < 1/2$, by Lemma S-3.14,

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p\left(T^{-(1/2-d)}\right).$$

Now consider $\left\|T^{-1}\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau}\right\|$. 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, (recalling that $r_{it} = x_{it}\left(\beta_{it} - \beta_i\right)$)

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

Since, by Assumption 2, 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,

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

Since $\mathbf{W}$ satisfy condition (i) of Lemma S-3.12, we have $\sup_{i,\ell,t,t'}\left|\mathbb{E}(w_{\ell t}w_{\ell t'}x_{it}x_{it'})\right| < C < \infty$. Also, note that for any $t' < t$,

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

and by Assumption 2, $\mathbb{E}\left(\beta_{it}-\beta_i|\mathcal{F}_{t-1}\right)=0$. Therefore,

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

Since, by Assumption 1, k is also a finite fixed integer, we conclude that

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

and hence, by Lemma S-3.13,

$$\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\| = O_p\left(T^{-\frac{1-d}{2}}\right).$$

So, we can conclude that

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

as required.

In the next step, consider the mean squared errors of the model, $T^{-1}\hat{\boldsymbol{\eta}}_T'\hat{\boldsymbol{\eta}}_T$. By substituting y from (S.18) into equation (S.17) 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.19), we get

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

we can further write

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

Therefore,

$$\begin{aligned}
&T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} - T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) \leq \\
&\quad T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\ \ \mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] + T^{-1}\left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right] + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} + \\
&\quad \left\|T^{-1}\left[\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u} - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 + \left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 + \\
&\quad \left\|T^{-1}\left[\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u} - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F + \\
&\quad \left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F.
\end{aligned} \tag{S.20}$$

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 5, $T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(\tilde{r}_{ij,t}\tilde{r}_{ij,t'}\right) = O\left(T^{-1}\right)$, and hence, by Lemma S-3.13, 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 1, k is a finite fixed integer, we can further conclude that

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\tilde{r}_{ij,t}\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.21}$$

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 2, $\mathbb{E}\left[(\beta_{it} - \beta_i)|\mathcal{F}_{t-1}\right] = 0$ 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-3.13, $\left|T^{-1}\sum_{t=1}^{T} r_{it}u_t\right| = O_p\left(\frac{1}{\sqrt{T}}\right)$. Since, by Assumption 1, k is a finite fixed integer, we conclude that

$$T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T} r_{it}u_t\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.22}$$

By substituting (S.21) and (S.22) into (S.20), and noting that $\left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\|^2 = O\left(T^{-(2\epsilon-d)}\right)$,

for some $\epsilon \geq 1/2$,

$$\left\|T^{-1}\left[\mathbf{W}'\mathbf{R}\boldsymbol{\tau}+\mathbf{W}'\mathbf{u}-\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|^2 = O_p(T^{-(1-d)}),$$

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}-\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p(T^{-(1/2-d)}),$$

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

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

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 $\bar{\sigma}^2_{u,T}=T^{-1}\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)$. 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 $\boldsymbol{\Omega}_{\beta,t}\equiv(\sigma_{ijt,\beta})$ and $\boldsymbol{\Sigma}_{\mathbf{x}_k,t}\equiv(\sigma_{ijt,x})$ for $i,j=1,2,\cdots,k$. By result 9(b) on page 44 of Lütkepohl (1996), we can further write

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

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.

In the second scenario, where $\mathbb{E}\left(\mathbf{w}_t\mathbf{w}_t'\right)$ is time-invariant, we can write (S.15) 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.15) in matrix format as

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

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

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

We can further use the similar lines of arguments as in the first scenario and write

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

We know that $\left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\ \mathbf{u}\right)\right\| = O\left(T^{-\frac{2\epsilon-d}{2}}\right)$ for some $\epsilon \geq 1/2$. Also,

$$\left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\| = O_p\left(T^{-\frac{1-d}{2}}\right),$$

and

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p\left(T^{-(1/2-d)}\right).$$

Now consider $\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|$. 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 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,

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

Since, by Assumption 2, $\mathbb{E}\left[w_{\ell t}w_{\ell' t} - \mathbb{E}(w_{\ell t}w_{\ell' t})|\mathcal{F}_{t-1}\right] = 0$ 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 2, 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,

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

and hence, by Lemma S-3.13,

$$\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\| = O_p\left(T^{-\frac{1-d}{2}}\right).$$

So, we conclude that

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

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

$$\begin{aligned}
&T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}}-T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}+\mathbf{u}'\mathbf{u}\right)\leq \\
&\quad T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\ \ \mathbf{H}\boldsymbol{\tau}-\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right]+T^{-1}\left[\mathbf{u}'\mathbf{u}-\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right]+2T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u}+ \\
&\quad\left\|T^{-1}\left[\mathbf{W}'\mathbf{H}\boldsymbol{\tau}+\mathbf{W}'\mathbf{u}-\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2+\left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2+ \\
&\quad\left\|T^{-1}\left[\mathbf{W}'\mathbf{H}\boldsymbol{\tau}+\mathbf{W}'\mathbf{u}-\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}-\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F+ \\
&\quad\left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}-\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F
\end{aligned} \tag{S.23}$$

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 5, $T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(\tilde{h}_{ij,t}\tilde{h}_{ij,t'}\right) = O\left(T^{-1}\right)$, and hence, by Lemma S-3.13, 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 1, k is a finite fixed integer, we can further conclude that

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\tilde{h}_{ij,t}\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.24}$$

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

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

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

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

Hence, by Lemma S-3.13, $\left|T^{-1}\sum_{t=1}^{T}h_{it}u_t\right| = O_p\left(\frac{1}{\sqrt{T}}\right)$. Since, by Assumption 1, k is a finite fixed integer, we conclude that

$$T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u} = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T}h_{it}u_t\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.25}$$

By substituting (S.24) and (S.25) into (S.23), and noting that $\left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\|^2 = O\left(T^{-(2\epsilon-d)}\right)$, for some $\epsilon \geq 1/2$,

$$\left\|T^{-1}\left[\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u} - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|^2 = O_p(T^{-(1-d)}),$$

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p(T^{-(1/2-d)}),$$

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

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

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

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

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. ■

Lemma S-2.8 *Let y_t $t = 1,2,\cdots,T$ be generated by (1). Suppose Assumption 1 and 2 hold, and the cross products of coefficients of the signals in DGP for y_t 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, $\sum_{t=1}^{T}\sum_{t'=1}^{T} cov(h_{ij,t}, h_{ij,t'}) = O(T)$ where $h_{ij,t} = x_{it}x_{jt}(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})$.

Proof. To show this, let $\tilde{h}_{ij,t} = h_{ij,t} - \mathbb{E}\left(h_{ij,t}\right)$. We have

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

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 Lemma,

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

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

S-3 Supplementary lemmas

Lemma S-3.1 *Let z_t be a martingale difference process with respect to $\mathcal{F}_{t-1}^z = \sigma(z_{t-1}, z_{t-2}, \cdots)$, 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-3.2 *Let*

$$c_p(n,\delta) = \Phi^{-1}\left(1 - \frac{p}{2f(n,\delta)}\right), \tag{S.26}$$

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-3.3 *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-3.4 *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-3.5 *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-3.6 *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 A3 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S-3.7 *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-3.4, 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-3.8 *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-3.9 *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,

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

and hence

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

■

Lemma S-3.10 *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-3.11 *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-3.9,

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

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-3.12 *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) $\sup_{i,t} \mathbb{E}\left|x_{it}\right|^4 < C$*,* $\sup_{j,t} \mathbb{E}\left|y_{jt}\right|^4 < C$*, 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,

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

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

Lemma S-3.13 *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

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

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-3.14 *Let $\mathbf{\Sigma}_T$ be a positive definite matrix and $\hat{\mathbf{\Sigma}}_T$ be its corresponding estimator. Suppose that $\lambda_{\min}(\mathbf{\Sigma}_T) > c > 0$, 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-3.11,

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

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

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

and hence, for all $T > T_0$,

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

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

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

Notice that $a_T \to 0$ as $T \to \infty$, therefore for large enough T^* we have both $\frac{\varepsilon}{C\|\boldsymbol{\Sigma}_T^{-1}\|_2^2} - 2a_T$ and

$\frac{\varepsilon}{C\|\mathbf{\Sigma}_T^{-1}\|_2^2} - a_T$ being greater than zero for all $T > T^*$. Now, by setting $T_\varepsilon = \max\{T^*, T_0\}$ and

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

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

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

■

By using Lemma S-3.11 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-3.15 *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-3.16 *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-3.11 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-3.15, 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-3.17 *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-3.16. ■

Lemma S-3.18 *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.

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

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

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

■

Lemma S-3.19 *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

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

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

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

Now, note that

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

and hence by Lemma S-3.3, for any $0 < \pi_j < 1$, $j = 1, 2, \cdots, \ell + 1$ we have,

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

where $\psi_T^* = sup_{i,j}\{\psi_{ji,T}\}$ and $\pi^* = \inf_{j \in 1,2,\cdots,\ell+1}\{\pi_j\}$. Also, there exist sufficiently large positive constants C_0 and C_1, and $s > 0$ such that for all $\alpha > 0$,

$$\sup_{j,t} \Pr(|z_{jt}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s),$$

and

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

Therefore,

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

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

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

■

Lemma S-3.20 *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-3.3, 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-3.9,

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

and again by Lemma S-3.4, 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-3.3, we have

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

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-3.18 we have $\sum_{t=1}^{T}\mathbb{E}(\tilde{x}_{it}z_{\ell t}) = 0$ and hence we can further write

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

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,

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

Let

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

and $i = 1, 2, ..., N$. By Lemmas S-3.7, S-3.18, and S-3.19, 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-3.1, there exists a finite positive constant C such that

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

or

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

for $h = \lambda, \alpha$. Also, depending on the value of exponentially bounded probability tail parameter, from Lemmas S-3.16 and S-3.17 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, there exists a finite positive constant C such that

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

or,

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

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 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

July 15, 2024

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.[8] 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* $\tilde{\mathbf{X}}^*$*.*

[8] For the implementation of Lasso and A-Lasso, we use the Matlab Glmnet package available at `https://hastie.su.domains/glmnet_matlab/`.

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* $\mathbf{m} = (1, 2, \cdots, K, 1, 2, \cdots, K, \cdots)'$ *where* K *is the number of folds.*

2. *Let* $\mathbf{m}^* = (m_1^*, m_2^*, \cdots, m_T^*)'$ *be a* $T \times 1$ *vector generated by randomly permuting the elements of* $\mathbf{m}$.

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

$$g_\ell = \{t : t \in \{1, 2, \cdots, T\} \text{ and } m_t^* = \ell\} \text{ for } \ell = 1, 2, \cdots, K.$$

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

   (a) *Remove the observations related to fold* ℓ *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* ℓ*, denoted by* $\hat{y}_t^f(\varphi_j)$.

5. *Compute the mean squared forecast errors for a given value of* φ_j *by*

$$MSFE(\varphi_j) = \frac{1}{T} \sum_{t=1}^{T} \left[y_t - \hat{y}_t^f(\varphi_j) \right]^2.$$

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 mean squared forecast errors 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.

Boosting

We implement boosting algorithm using a BIC stopping criterion. We have also considered corrected AIC stopping criterion (Bühlmann (2006)) and these results are available in the working paper version of this paper.

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*

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

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

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

$$\hat{\sigma}^2 = \frac{1}{T}\left(\tilde{\mathbf{y}} - \mathcal{B}_m\tilde{\mathbf{y}}\right)'\left(\tilde{\mathbf{y}} - \mathcal{B}_m\tilde{\mathbf{y}}\right),$$

$$\mathcal{B}_m = I - \left(I - v\mathcal{H}^{(\hat{s}_m)}\right)\left(I - v\mathcal{H}^{(\hat{s}_{m-1})}\right)...\left(I - v\mathcal{H}^{(\hat{s}_1)}\right),$$

$$\mathcal{H}^{(j)} = \frac{\tilde{\mathbf{x}}^*_{j\circ}\tilde{\mathbf{x}}^{*\prime}_{j\circ}}{\tilde{\mathbf{x}}^{*\prime}_{j\circ}\tilde{\mathbf{x}}^*_{j\circ}}.$$

We set $m_{\max} = 500$ and $v = 0.5$.

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**			**150**			**200**	
A. Without parameter instability									
	OCMT (down-weighting at the selection stage)								
20	**0.83**	0.68	0.60	**0.91**	0.73	0.67	**0.96**	0.77	0.74
40	**0.80**	0.64	0.57	**0.91**	0.71	0.66	**0.95**	0.76	0.73
100	**0.77**	0.60	0.53	**0.88**	0.67	0.63	**0.93**	0.72	0.71
	Lasso								
20	**0.84**	0.78	0.73	**0.89**	0.80	0.74	**0.93**	0.82	0.75
40	**0.82**	0.76	0.73	**0.89**	0.79	0.74	**0.92**	0.81	0.75
100	**0.79**	0.74	0.70	**0.87**	0.78	0.75	**0.90**	0.79	0.76
	A-Lasso								
20	**0.73**	0.69	0.64	**0.80**	0.73	0.66	**0.85**	0.75	0.67
40	**0.73**	0.69	0.65	**0.81**	0.74	0.68	**0.86**	0.76	0.69
100	**0.72**	0.68	0.64	**0.81**	0.73	0.69	**0.86**	0.75	0.70
	Boosting								
20	0.77	0.76	**0.78**	**0.83**	0.81	0.83	**0.88**	0.85	0.86
40	0.76	0.77	**0.81**	0.83	0.83	**0.86**	0.87	0.86	**0.89**
100	0.75	0.79	**0.81**	0.82	0.85	**0.85**	0.86	**0.88**	0.87
B. With parameter instability									
	OCMT (down-weighting at the selection stage)								
20	**0.73**	0.59	0.57	**0.85**	0.69	0.68	**0.92**	0.76	0.75
40	**0.70**	0.56	0.54	**0.84**	0.67	0.66	**0.91**	0.75	0.75
100	**0.66**	0.51	0.50	**0.81**	0.63	0.63	**0.88**	0.70	0.72
	Lasso								
20	**0.76**	0.72	0.71	**0.82**	0.78	0.75	**0.87**	0.81	0.77
40	**0.74**	0.70	0.70	**0.82**	0.76	0.75	**0.86**	0.79	0.77
100	**0.70**	0.67	0.66	**0.79**	0.73	0.73	**0.83**	0.77	0.76
	A-Lasso								
20	**0.65**	0.63	0.62	**0.72**	0.69	0.66	**0.78**	0.74	0.69
40	**0.64**	0.62	0.62	**0.73**	0.70	0.68	**0.79**	0.74	0.71
100	**0.63**	0.61	0.59	**0.73**	0.68	0.67	**0.78**	0.72	0.71
	Boosting								
20	0.68	0.69	**0.74**	0.74	0.77	**0.81**	0.79	0.82	**0.85**
40	0.68	0.70	**0.77**	0.75	0.78	**0.84**	0.80	0.83	**0.88**
100	0.67	0.71	**0.76**	0.74	0.79	**0.82**	0.78	0.83	**0.85**

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 6 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**			**150**			**200**	
A. Without parameter instability									
	OCMT (down-weighting at the selection stage)								
20	0.08	**0.06**	0.10	0.13	**0.09**	0.17	0.17	**0.13**	0.23
40	0.04	**0.03**	0.09	**0.06**	0.06	0.16	**0.08**	0.10	0.22
100	**0.01**	0.02	0.07	**0.02**	0.04	0.14	**0.03**	0.07	0.21
	Lasso								
20	**0.17**	0.20	0.23	**0.17**	0.20	0.23	**0.17**	0.20	0.23
40	**0.12**	0.17	0.24	**0.13**	0.17	0.24	**0.13**	0.17	0.24
100	**0.08**	0.14	0.19	**0.08**	0.14	0.23	**0.07**	0.14	0.23
	A-Lasso								
20	**0.11**	0.14	0.17	**0.11**	0.14	0.16	**0.11**	0.14	0.16
40	**0.09**	0.13	0.18	**0.09**	0.13	0.18	**0.09**	0.13	0.18
100	**0.06**	0.11	0.15	**0.06**	0.11	0.18	**0.05**	0.12	0.18
	Boosting								
20	**0.08**	0.16	0.28	**0.07**	0.19	0.33	**0.06**	0.22	0.37
40	**0.07**	0.19	0.38	**0.06**	0.23	0.43	**0.05**	0.27	0.47
100	**0.08**	0.25	0.40	**0.06**	0.30	0.44	**0.05**	0.34	0.46
B. With parameter instability									
	OCMT (down-weighting at the selection stage)								
20	0.06	**0.05**	0.10	**0.08**	0.08	0.17	**0.11**	0.12	0.23
40	**0.02**	0.03	0.09	**0.04**	0.07	0.16	**0.05**	0.10	0.23
100	**0.01**	0.02	0.07	**0.01**	0.05	0.14	**0.02**	0.08	0.21
	Lasso								
20	**0.21**	0.23	0.26	**0.22**	0.23	0.26	**0.23**	0.24	0.26
40	**0.17**	0.20	0.26	**0.18**	0.21	0.27	**0.19**	0.21	0.27
100	**0.11**	0.16	0.21	**0.12**	0.17	0.24	**0.12**	0.17	0.25
	A-Lasso								
20	**0.15**	0.17	0.19	**0.15**	0.16	0.18	**0.16**	0.17	0.18
40	**0.12**	0.15	0.20	**0.13**	0.16	0.20	**0.14**	0.16	0.20
100	**0.08**	0.12	0.16	**0.09**	0.13	0.19	**0.09**	0.14	0.19
	Boosting								
20	**0.09**	0.17	0.28	**0.08**	0.20	0.34	**0.08**	0.23	0.38
40	**0.10**	0.20	0.38	**0.08**	0.23	0.43	**0.08**	0.27	0.48
100	**0.10**	0.25	0.40	**0.08**	0.30	0.44	**0.07**	0.34	0.46

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 6 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**			**150**			**200**	
A. Without parameter instability									
	OCMT (down-weighting at the selection stage)								
20	5.03	3.82	4.44	6.17	4.70	6.05	7.22	5.63	7.54
40	4.69	3.91	5.69	5.98	5.42	8.89	6.87	6.95	11.89
100	4.31	4.31	8.94	5.52	7.01	16.13	6.35	10.11	23.92
	Lasso								
20	6.82	7.09	7.60	7.00	7.17	7.55	7.20	7.31	7.62
40	8.26	9.78	12.49	8.57	10.01	12.58	8.74	10.10	12.68
100	10.76	16.51	22.19	11.00	17.58	26.09	10.51	17.59	26.14
	A-Lasso								
20	5.15	5.52	5.92	5.35	5.63	5.90	5.55	5.79	5.97
40	6.39	7.83	9.98	6.78	8.08	10.03	6.96	8.21	10.14
100	8.65	13.27	17.38	9.05	14.37	20.36	8.83	14.51	20.52
	Boosting								
20	4.59	6.20	8.66	4.63	7.00	9.92	4.70	7.76	10.87
40	6.04	10.58	18.30	5.79	12.30	20.67	5.69	14.16	22.54
100	11.36	28.60	43.23	9.27	33.19	47.00	8.43	37.53	49.32
B. With parameter instability									
	OCMT (down-weighting at the selection stage)								
20	4.04	3.39	4.32	5.07	4.45	6.03	5.96	5.45	7.52
40	3.78	3.62	5.71	4.90	5.28	8.96	5.67	6.92	12.00
100	3.54	4.37	9.28	4.62	7.25	16.44	5.26	10.38	24.22
	Lasso								
20	7.28	7.51	8.00	7.76	7.69	8.13	8.17	7.96	8.27
40	9.80	10.90	13.32	10.60	11.31	13.64	11.13	11.44	13.83
100	13.68	18.72	23.30	14.83	20.09	27.36	15.56	20.23	27.93
	A-Lasso								
20	5.49	5.83	6.22	5.95	6.04	6.35	6.30	6.27	6.47
40	7.55	8.63	10.58	8.28	9.03	10.81	8.76	9.23	11.03
100	10.71	14.82	18.14	11.85	16.20	21.27	12.58	16.50	21.86
	Boosting								
20	4.59	6.16	8.59	4.66	7.00	9.95	4.75	7.79	10.95
40	6.52	10.78	18.18	6.35	12.49	20.70	6.21	14.24	22.61
100	12.70	28.14	42.94	10.73	33.22	47.05	10.03	37.60	49.49

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 6 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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
A. Static									
	OCMT								
20	5.78	6.93	7.95	0.92	0.97	0.99	0.11	0.15	0.20
40	5.51	6.78	7.63	0.90	0.97	0.99	0.05	0.07	0.09
100	5.24	6.46	7.20	0.87	0.96	0.98	0.02	0.03	0.03
	Lasso								
20	7.51	7.87	8.08	0.87	0.92	0.95	0.20	0.21	0.21
40	9.43	10.12	10.59	0.86	0.92	0.95	0.15	0.16	0.17
100	12.32	13.47	13.84	0.83	0.90	0.93	0.09	0.10	0.10
	A-Lasso								
20	5.75	6.10	6.31	0.78	0.84	0.89	0.13	0.14	0.14
40	7.32	7.98	8.40	0.78	0.86	0.90	0.11	0.11	0.12
100	9.75	10.89	11.30	0.77	0.85	0.89	0.07	0.07	0.08
	Boosting								
20	5.00	5.11	5.14	0.81	0.87	0.91	0.09	0.08	0.08
40	6.67	6.56	6.45	0.81	0.87	0.91	0.09	0.08	0.07
100	11.71	10.32	9.74	0.79	0.86	0.90	0.09	0.07	0.06
B. Dynamic									
	OCMT								
20	3.29	4.31	5.23	0.65	0.80	0.89	0.03	0.06	0.08
40	2.96	4.10	4.91	0.60	0.78	0.87	0.01	0.02	0.04
100	2.61	3.69	4.41	0.55	0.73	0.83	0.00	0.01	0.01
	Lasso								
20	6.59	6.88	7.28	0.72	0.80	0.85	0.18	0.19	0.19
40	8.63	9.04	9.28	0.69	0.78	0.84	0.15	0.15	0.15
100	12.12	12.35	12.23	0.66	0.75	0.80	0.09	0.09	0.09
	A-Lasso								
20	4.89	5.20	5.54	0.60	0.68	0.74	0.12	0.12	0.13
40	6.63	7.08	7.31	0.60	0.69	0.75	0.11	0.11	0.11
100	9.61	10.01	10.11	0.58	0.69	0.75	0.07	0.07	0.07
	Boosting								
20	4.17	4.18	4.30	0.64	0.70	0.76	0.08	0.07	0.06
40	5.90	5.58	5.45	0.63	0.71	0.76	0.08	0.07	0.06
100	12.35	9.67	8.72	0.63	0.70	0.75	0.10	0.07	0.06

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-weighting only at the estimation stage)								
20	**17.03**	17.50	18.43	**15.59**	16.07	17.13	**14.44**	15.20	16.39
40	**15.83**	16.27	17.06	**14.71**	15.17	16.18	**16.87**	18.43	20.01
100	**16.00**	16.28	16.96	**15.03**	15.83	17.05	**15.81**	16.41	17.66
	OCMT(Down-weighting only at the variable selection and estimation stages)								
20	**17.03**	17.40	18.95	**15.59**	16.19	17.83	**14.44**	14.99	17.16
40	**15.83**	16.30	18.07	**14.71**	14.96	17.80	**16.87**	18.61	23.16
100	**16.00**	16.50	19.89	**15.03**	16.55	22.85	**15.81**	17.40	25.81
	Lasso								
20	**17.23**	17.77	18.89	**15.61**	16.15	17.16	**14.41**	14.76	15.71
40	**16.11**	17.00	18.33	**14.43**	15.21	17.06	**16.56**	17.97	19.91
100	**16.44**	17.99	20.58	**15.38**	16.76	18.84	**15.94**	17.26	18.90
	A-Lasso								
20	**18.08**	18.52	19.76	**16.05**	16.77	17.87	**14.75**	15.10	16.32
40	**17.25**	18.23	19.71	**15.25**	16.30	18.26	**17.24**	19.10	21.13
100	**18.56**	20.33	22.88	**16.55**	18.29	20.87	**16.93**	18.87	20.57
	Boosting								
20	**17.56**	18.43	21.17	**15.94**	16.94	19.68	**14.49**	15.63	18.94
40	**16.62**	17.87	21.69	**14.82**	16.52	21.52	**17.02**	20.22	26.76
100	**17.45**	21.81	25.46	**15.93**	19.98	24.11	**16.26**	21.13	25.42
B. With parameter instabilities									
	OCMT(Down-weighting only at the estimation stage)								
20	20.09	**19.32**	19.47	17.80	**17.09**	17.58	16.82	**15.82**	16.62
40	18.46	**17.94**	18.20	17.32	**16.38**	16.84	19.43	**19.18**	20.34
100	19.01	**18.55**	18.67	17.97	**17.19**	17.76	18.75	**17.46**	18.21
	OCMT(Down-weighting only at the variable selection and estimation stages)								
20	20.09	**19.95**	20.91	17.80	**17.71**	19.32	16.82	**15.92**	17.97
40	**18.46**	18.65	20.43	17.32	**17.01**	20.09	**19.43**	20.20	25.09
100	**19.01**	19.11	22.92	**17.97**	19.09	25.28	**18.75**	19.66	28.60
	Lasso								
20	20.93	**20.63**	21.04	18.29	**17.76**	18.43	17.12	**16.01**	16.70
40	**19.23**	19.47	20.40	17.70	**17.47**	19.09	**19.68**	19.68	21.40
100	**19.95**	20.64	22.69	**18.87**	19.35	20.84	19.50	**19.18**	20.70
	A-Lasso								
20	21.69	**21.16**	21.64	18.85	**18.18**	18.93	17.41	**16.12**	17.26
40	**20.32**	20.56	21.84	18.74	**18.52**	20.44	**20.52**	20.63	22.66
100	**22.24**	22.91	24.98	**20.58**	20.93	23.00	21.21	**21.00**	22.66
	Boosting								
20	21.02	**20.61**	22.66	18.36	**18.34**	20.97	17.05	**16.72**	20.02
40	**19.52**	20.23	23.87	**17.44**	18.57	23.42	**19.74**	21.43	27.85
100	**19.98**	23.60	27.42	**18.69**	21.53	25.57	**19.11**	22.61	26.85

Notes: The reported results are averaged over two experiments (low fit and high fit) for models without and with parameter instabilities. See Section 6 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-weighting only at the estimation stage)								
20	**46.49**	47.83	49.98	**41.46**	42.60	44.99	**37.93**	39.14	41.11
40	**42.43**	42.85	43.97	**38.73**	39.30	40.86	**47.24**	49.56	52.57
100	**42.50**	42.85	44.01	**40.84**	42.11	44.33	**42.07**	42.88	45.30
	OCMT(Down-weighting only at the variable selection and estimation stages)								
20	**46.49**	47.82	51.20	**41.46**	42.32	46.54	**37.93**	39.72	44.19
40	**42.43**	42.62	45.84	**38.73**	39.45	45.19	**47.24**	49.92	59.09
100	**42.50**	43.90	47.82	**40.84**	42.37	50.59	**42.07**	44.99	54.47
	Lasso								
20	**46.41**	48.93	52.09	**41.58**	42.83	46.06	**38.09**	39.68	42.44
40	**42.85**	44.83	49.99	**38.27**	40.80	47.56	**47.00**	49.53	55.69
100	**44.82**	48.60	53.52	**41.28**	45.04	51.48	**42.33**	45.60	51.30
	A-Lasso								
20	**48.41**	50.92	54.42	**42.90**	44.11	47.83	**39.27**	40.54	43.94
40	**46.08**	47.51	52.88	**40.71**	43.86	51.19	**48.83**	51.08	56.65
100	**52.01**	55.44	60.11	**45.27**	49.56	56.53	**46.12**	49.87	55.69
	Boosting								
20	**47.82**	52.59	61.33	**43.09**	47.03	56.49	**39.04**	42.81	52.70
40	**44.73**	50.57	62.92	**39.59**	46.80	62.00	**48.77**	60.10	76.55
100	**49.91**	62.19	71.42	**42.63**	57.36	69.52	**43.70**	58.55	68.93
B. With parameter instabilities									
	OCMT(Down-weighting only at the estimation stage)								
20	51.65	**50.55**	51.44	44.55	**43.26**	44.45	41.43	**39.82**	41.21
40	46.37	**44.89**	45.21	42.74	**41.14**	41.98	51.12	**50.38**	52.58
100	47.60	**46.54**	46.98	45.35	**43.91**	45.14	46.68	**44.52**	46.22
	OCMT(Down-weighting only at the variable selection and estimation stages)								
20	51.65	**51.29**	53.67	44.55	**44.30**	48.16	41.43	**41.00**	45.22
40	46.37	**45.53**	48.40	42.74	**42.01**	47.79	51.12	**51.31**	61.08
100	**47.60**	47.85	51.66	45.35	**45.09**	52.79	46.68	**47.05**	59.42
	Lasso								
20	**52.76**	53.46	55.50	45.12	**44.79**	47.61	42.48	**41.48**	44.01
40	**47.64**	48.16	52.38	**43.17**	43.46	49.71	51.54	**51.12**	56.91
100	**49.94**	52.33	56.53	**46.40**	49.09	54.78	**48.04**	48.17	53.57
	A-Lasso								
20	**55.26**	55.36	57.59	46.38	**45.68**	49.11	43.39	**42.12**	45.31
40	**50.97**	51.14	55.62	**46.07**	46.25	53.31	53.60	**52.64**	58.36
100	**57.39**	59.47	63.11	**50.76**	54.02	60.09	**52.35**	52.63	58.42
	Boosting								
20	**52.12**	55.54	63.86	**45.19**	48.95	58.94	**41.52**	44.18	54.61
40	**48.03**	54.50	66.77	**42.50**	50.16	65.76	**51.12**	61.14	77.47
100	**52.19**	65.79	75.81	**44.87**	59.92	72.47	**46.92**	61.59	72.44

Notes: The reported results are averaged across two experiments (low fit and high fit) for models without and with parameter instabilities. See Section 6 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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	25.46	23.49	21.77	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	23.81	21.86	25.33	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	24.11	22.98	23.93	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	26.27	24.00	22.26	5.22	6.37	7.38	0.90	0.96	0.99	0.08	0.13	0.17
40	24.39	22.60	25.97	4.87	6.18	7.03	0.88	0.97	0.99	0.03	0.06	0.08
100	24.73	23.14	24.40	4.49	5.72	6.48	0.84	0.95	0.98	0.01	0.02	0.03
	LASSO											
20	26.50	24.14	22.28	6.71	7.03	7.22	0.85	0.91	0.95	0.17	0.17	0.17
40	24.76	22.28	25.55	7.91	8.41	8.73	0.82	0.90	0.94	0.12	0.12	0.12
100	25.34	23.77	24.66	9.86	10.56	10.23	0.79	0.88	0.92	0.07	0.07	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	28.54	25.06	22.83	6.71	7.03	7.22	0.85	0.91	0.95	0.17	0.17	0.17
40	27.07	24.05	26.92	7.91	8.41	8.73	0.82	0.90	0.94	0.12	0.12	0.12
100	30.00	26.12	26.82	9.86	10.56	10.23	0.79	0.88	0.92	0.07	0.07	0.07
	A-LASSO											
20	27.85	24.90	22.84	5.01	5.31	5.50	0.72	0.80	0.86	0.11	0.11	0.10
40	26.46	23.55	26.66	6.04	6.58	6.91	0.72	0.82	0.87	0.08	0.08	0.09
100	28.71	25.63	26.23	7.86	8.65	8.54	0.71	0.82	0.87	0.05	0.05	0.05
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	28.54	25.26	22.96	5.01	5.31	5.50	0.72	0.80	0.86	0.11	0.11	0.10
40	27.10	24.08	27.06	6.04	6.58	6.91	0.72	0.82	0.87	0.08	0.08	0.09
100	29.79	26.08	26.70	7.86	8.65	8.54	0.71	0.82	0.87	0.05	0.05	0.05
	Boosting											
20	26.71	24.37	22.23	4.50	4.66	4.77	0.76	0.84	0.90	0.07	0.06	0.06
40	25.22	22.67	26.03	5.82	5.79	5.73	0.76	0.84	0.89	0.07	0.06	0.05
100	26.55	24.36	25.00	10.31	8.92	8.31	0.74	0.83	0.88	0.07	0.06	0.05
	Boosting for variable selection only. LS for estimation/forecasting.											
20	27.27	24.77	22.52	4.50	4.66	4.77	0.76	0.84	0.90	0.07	0.06	0.06
40	26.44	23.50	26.54	5.82	5.79	5.73	0.76	0.84	0.89	0.07	0.06	0.05
100	31.07	26.81	27.25	10.31	8.92	8.31	0.74	0.83	0.88	0.07	0.06	0.05

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 6 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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	26.03	23.94	22.22	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	24.61	22.22	26.83	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	24.50	23.74	24.32	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	26.96	24.63	23.43	5.22	6.37	7.38	0.90	0.96	0.99	0.08	0.13	0.17
40	24.98	23.25	28.17	4.87	6.18	7.03	0.88	0.97	0.99	0.03	0.06	0.08
100	25.07	24.38	25.24	4.49	5.72	6.48	0.84	0.95	0.98	0.01	0.02	0.03
	LASSO											
20	29.17	26.43	23.52	6.71	7.03	7.22	0.85	0.91	0.95	0.17	0.17	0.17
40	27.76	24.94	29.54	7.91	8.41	8.73	0.82	0.90	0.94	0.12	0.12	0.12
100	30.72	28.34	27.99	9.86	10.56	10.23	0.79	0.88	0.92	0.07	0.07	0.07
	A-LASSO											
20	28.95	26.29	23.55	5.01	5.31	5.50	0.72	0.80	0.86	0.11	0.11	0.10
40	27.67	24.46	29.55	6.04	6.58	6.91	0.72	0.82	0.87	0.08	0.08	0.09
100	30.18	27.96	27.62	7.86	8.65	8.54	0.71	0.82	0.87	0.05	0.05	0.05
	Boosting											
20	27.77	25.76	23.25	4.50	4.66	4.77	0.76	0.84	0.90	0.07	0.06	0.06
40	27.14	24.13	28.70	5.82	5.79	5.73	0.76	0.84	0.89	0.07	0.06	0.05
100	31.86	28.51	28.00	10.31	8.92	8.31	0.74	0.83	0.88	0.07	0.06	0.05
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	26.77	24.91	23.15	4.02	5.01	6.09	0.72	0.77	0.81	0.06	0.10	0.14
40	25.04	23.06	28.66	4.16	5.95	7.73	0.68	0.75	0.80	0.04	0.07	0.11
100	25.38	25.31	26.52	4.72	8.01	11.67	0.63	0.72	0.76	0.02	0.05	0.09
	LASSO											
20	27.27	24.78	22.76	6.79	6.94	7.04	0.77	0.80	0.82	0.18	0.19	0.19
40	26.08	23.41	27.64	8.88	9.14	9.32	0.75	0.79	0.80	0.15	0.15	0.15
100	27.62	25.79	26.52	14.00	14.54	14.37	0.71	0.76	0.77	0.11	0.12	0.11
	A-LASSO											
20	28.46	25.79	23.33	5.21	5.39	5.54	0.66	0.71	0.74	0.13	0.13	0.13
40	27.96	25.11	29.45	7.06	7.35	7.55	0.67	0.72	0.74	0.11	0.11	0.11
100	31.28	28.11	28.98	11.26	11.91	11.91	0.65	0.71	0.72	0.09	0.09	0.09
	Boosting											
20	28.12	25.93	23.88	6.01	6.86	7.63	0.75	0.81	0.85	0.15	0.18	0.21
40	27.19	25.31	31.04	10.08	11.92	13.77	0.76	0.82	0.86	0.18	0.22	0.26
100	33.31	30.66	32.46	27.16	32.12	36.55	0.77	0.84	0.87	0.24	0.29	0.33

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	27.19	25.06	23.13	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	25.87	23.33	28.02	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	25.46	24.97	25.48	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	28.34	26.13	25.21	5.22	6.37	7.38	0.90	0.96	0.99	0.08	0.13	0.17
40	26.08	24.62	30.46	4.87	6.18	7.03	0.88	0.97	0.99	0.03	0.06	0.08
100	25.92	26.21	27.04	4.49	5.72	6.48	0.84	0.95	0.98	0.01	0.02	0.03
	LASSO											
20	30.43	28.22	24.94	6.71	7.03	7.22	0.85	0.91	0.95	0.17	0.17	0.17
40	29.11	26.49	31.76	7.91	8.41	8.73	0.82	0.90	0.94	0.12	0.12	0.12
100	32.33	31.25	30.51	9.86	10.56	10.23	0.79	0.88	0.92	0.07	0.07	0.07
	A-LASSO											
20	29.95	27.62	24.58	5.01	5.31	5.50	0.72	0.80	0.86	0.11	0.11	0.10
40	28.69	25.49	31.33	6.04	6.58	6.91	0.72	0.82	0.87	0.08	0.08	0.09
100	31.01	30.40	29.63	7.86	8.65	8.54	0.71	0.82	0.87	0.05	0.05	0.05
	Boosting											
20	28.76	26.95	24.44	4.50	4.66	4.77	0.76	0.84	0.90	0.07	0.06	0.06
40	28.33	25.36	30.09	5.82	5.79	5.73	0.76	0.84	0.89	0.07	0.06	0.05
100	33.44	30.59	29.83	10.31	8.92	8.31	0.74	0.83	0.88	0.07	0.06	0.05
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	29.10	27.31	26.41	4.80	6.53	8.15	0.63	0.70	0.76	0.11	0.19	0.26
40	27.52	27.22	35.47	6.33	9.90	13.05	0.60	0.69	0.76	0.10	0.18	0.25
100	29.94	34.28	38.83	10.34	18.38	26.56	0.56	0.66	0.73	0.08	0.16	0.24
	LASSO											
20	28.91	26.21	24.14	6.99	7.03	6.94	0.70	0.71	0.72	0.21	0.21	0.20
40	27.96	26.15	30.41	10.97	10.81	10.96	0.69	0.70	0.71	0.21	0.20	0.20
100	31.60	28.86	28.86	19.87	20.77	20.96	0.66	0.69	0.70	0.17	0.18	0.18
	A-LASSO											
20	30.32	27.39	25.11	5.39	5.43	5.41	0.60	0.62	0.63	0.15	0.15	0.14
40	30.06	28.02	32.39	8.71	8.61	8.74	0.61	0.63	0.65	0.16	0.15	0.15
100	35.21	31.96	31.41	15.44	16.29	16.52	0.59	0.63	0.65	0.13	0.14	0.14
	Boosting											
20	32.30	30.00	28.83	8.41	9.73	10.69	0.76	0.82	0.85	0.27	0.32	0.36
40	33.00	32.93	41.08	17.77	20.32	22.23	0.80	0.85	0.88	0.36	0.42	0.47
100	38.83	36.94	38.92	42.41	46.33	48.74	0.79	0.84	0.86	0.39	0.43	0.45

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	7.42	6.84	6.34	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	6.94	6.37	7.38	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	7.03	6.69	6.97	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	7.79	7.19	6.62	7.34	8.61	9.82	1.00	1.00	1.00	0.17	0.23	0.29
40	7.26	6.82	7.76	7.07	8.47	9.48	0.99	1.00	1.00	0.08	0.11	0.14
100	7.26	6.91	7.22	6.72	8.05	8.98	0.99	1.00	1.00	0.03	0.04	0.05
	LASSO											
20	7.96	7.07	6.55	7.52	7.56	7.53	0.98	0.99	1.00	0.18	0.18	0.18
40	7.46	6.59	7.56	8.88	8.97	9.10	0.97	0.99	1.00	0.13	0.13	0.13
100	7.54	7.00	7.21	11.05	11.28	10.78	0.96	0.99	1.00	0.07	0.07	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	8.59	7.32	6.72	7.52	7.56	7.53	0.98	0.99	1.00	0.18	0.18	0.18
40	8.09	7.17	8.01	8.88	8.97	9.10	0.97	0.99	1.00	0.13	0.13	0.13
100	8.72	7.69	7.80	11.05	11.28	10.78	0.96	0.99	1.00	0.07	0.07	0.07
	A-LASSO											
20	8.30	7.19	6.67	5.93	5.99	6.02	0.93	0.96	0.98	0.11	0.11	0.10
40	8.03	6.94	7.82	7.09	7.30	7.38	0.93	0.97	0.99	0.08	0.09	0.09
100	8.41	7.46	7.63	9.04	9.46	9.14	0.92	0.97	0.99	0.05	0.06	0.05
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	8.47	7.26	6.74	5.93	5.99	6.02	0.93	0.96	0.98	0.11	0.11	0.10
40	8.16	7.11	7.93	7.09	7.30	7.38	0.93	0.97	0.99	0.08	0.09	0.09
100	8.66	7.57	7.74	9.04	9.46	9.14	0.92	0.97	0.99	0.05	0.06	0.05
	Boosting											
20	8.42	7.50	6.75	5.41	5.31	5.23	0.96	0.98	0.99	0.08	0.07	0.06
40	8.01	6.97	8.01	6.79	6.45	6.23	0.95	0.98	0.99	0.08	0.06	0.06
100	8.36	7.50	7.52	11.46	9.78	8.93	0.94	0.98	0.99	0.08	0.06	0.05
	Boosting for variable selection only. LS for estimation/forecasting.											
20	8.12	7.14	6.57	5.41	5.31	5.23	0.96	0.98	0.99	0.08	0.07	0.06
40	7.92	6.93	7.89	6.79	6.45	6.23	0.95	0.98	0.99	0.08	0.06	0.06
100	9.41	7.92	7.80	11.46	9.78	8.93	0.94	0.98	0.99	0.08	0.06	0.05

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	7.58	6.98	6.47	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	7.17	6.47	7.82	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	7.14	6.92	7.08	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	8.05	7.50	6.96	7.34	8.61	9.82	1.00	1.00	1.00	0.17	0.23	0.29
40	7.55	7.10	8.69	7.07	8.47	9.48	0.99	1.00	1.00	0.08	0.11	0.14
100	7.49	7.28	7.57	6.72	8.05	8.98	0.99	1.00	1.00	0.03	0.04	0.05
	LASSO											
20	8.84	7.67	6.95	7.52	7.56	7.53	0.98	0.99	1.00	0.18	0.18	0.18
40	8.33	7.41	8.84	8.88	8.97	9.10	0.97	0.99	1.00	0.13	0.13	0.13
100	9.02	8.36	8.20	11.05	11.28	10.78	0.96	0.99	1.00	0.07	0.07	0.07
	A-LASSO											
20	8.63	7.55	7.00	5.93	5.99	6.02	0.93	0.96	0.98	0.11	0.11	0.10
40	8.34	7.22	8.67	7.09	7.30	7.38	0.93	0.97	0.99	0.08	0.09	0.09
100	8.93	8.13	8.06	9.04	9.46	9.14	0.92	0.97	0.99	0.05	0.06	0.05
	Boosting											
20	8.28	7.42	6.82	5.41	5.31	5.23	0.96	0.98	0.99	0.08	0.07	0.06
40	8.15	7.16	8.66	6.79	6.45	6.23	0.95	0.98	0.99	0.08	0.06	0.06
100	9.77	8.48	8.10	11.46	9.78	8.93	0.94	0.98	0.99	0.08	0.06	0.05
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	8.04	7.47	6.83	5.77	6.78	7.86	0.90	0.90	0.91	0.11	0.16	0.21
40	7.55	6.87	8.56	6.25	8.23	10.17	0.89	0.90	0.91	0.07	0.12	0.16
100	7.62	7.80	8.28	7.48	11.43	15.72	0.87	0.88	0.90	0.04	0.08	0.12
	LASSO											
20	8.27	7.53	6.75	8.00	8.10	8.14	0.95	0.96	0.97	0.21	0.21	0.21
40	7.92	7.00	8.29	10.47	10.65	10.87	0.94	0.96	0.96	0.17	0.17	0.18
100	8.35	7.72	8.00	16.25	17.03	16.87	0.93	0.95	0.95	0.13	0.13	0.13
	A-LASSO											
20	8.58	7.75	6.88	6.39	6.50	6.58	0.90	0.92	0.94	0.14	0.14	0.14
40	8.50	7.48	8.75	8.53	8.74	8.92	0.90	0.93	0.94	0.12	0.13	0.13
100	9.38	8.46	8.75	13.22	14.07	14.05	0.89	0.93	0.94	0.10	0.10	0.10
	Boosting											
20	8.74	7.95	7.37	6.97	7.67	8.33	0.94	0.96	0.97	0.16	0.19	0.22
40	8.54	7.74	9.39	11.13	12.79	14.57	0.94	0.97	0.98	0.18	0.22	0.27
100	10.30	9.31	9.80	28.66	33.21	37.40	0.94	0.97	0.98	0.25	0.29	0.33

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	7.92	7.30	6.74	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	7.54	6.80	8.16	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	7.42	7.27	7.42	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	8.51	8.13	7.58	7.34	8.61	9.82	1.00	1.00	1.00	0.17	0.23	0.29
40	8.04	7.73	9.56	7.07	8.47	9.48	0.99	1.00	1.00	0.08	0.11	0.14
100	8.01	7.90	8.28	6.72	8.05	8.98	0.99	1.00	1.00	0.03	0.04	0.05
	LASSO											
20	9.24	8.21	7.38	7.52	7.56	7.53	0.98	0.99	1.00	0.18	0.18	0.18
40	8.74	7.93	9.59	8.88	8.97	9.10	0.97	0.99	1.00	0.13	0.13	0.13
100	9.56	9.26	8.97	11.05	11.28	10.78	0.96	0.99	1.00	0.07	0.07	0.07
	A-LASSO											
20	8.97	7.98	7.37	5.93	5.99	6.02	0.93	0.96	0.98	0.11	0.11	0.10
40	8.70	7.57	9.28	7.09	7.30	7.38	0.93	0.97	0.99	0.08	0.09	0.09
100	9.36	8.91	8.67	9.04	9.46	9.14	0.92	0.97	0.99	0.05	0.06	0.05
	Boosting											
20	8.63	7.82	7.21	5.41	5.31	5.23	0.96	0.98	0.99	0.08	0.07	0.06
40	8.53	7.54	9.22	6.79	6.45	6.23	0.95	0.98	0.99	0.08	0.06	0.06
100	10.41	9.20	8.76	11.46	9.78	8.93	0.94	0.98	0.99	0.08	0.06	0.05
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	8.81	8.35	7.91	6.52	8.25	9.82	0.80	0.83	0.87	0.17	0.25	0.32
40	8.63	8.38	10.85	8.72	12.41	15.59	0.79	0.83	0.87	0.14	0.23	0.30
100	9.83	11.42	12.78	14.24	22.78	31.22	0.76	0.82	0.86	0.11	0.19	0.28
	LASSO											
20	8.88	8.11	7.27	8.56	8.56	8.53	0.90	0.91	0.91	0.25	0.25	0.24
40	8.70	7.96	9.40	13.08	12.99	13.18	0.90	0.91	0.91	0.24	0.23	0.24
100	9.57	8.82	8.95	22.24	23.59	23.96	0.89	0.90	0.91	0.19	0.20	0.20
	A-LASSO											
20	9.20	8.35	7.54	6.78	6.78	6.78	0.84	0.86	0.87	0.17	0.17	0.17
40	9.37	8.49	9.88	10.53	10.43	10.59	0.85	0.87	0.87	0.18	0.17	0.18
100	10.54	9.78	9.74	17.29	18.52	18.88	0.85	0.87	0.88	0.14	0.15	0.15
	Boosting											
20	10.05	9.37	9.05	9.31	10.45	11.31	0.93	0.95	0.96	0.28	0.33	0.37
40	10.39	10.10	12.44	18.75	21.01	22.80	0.94	0.96	0.97	0.37	0.43	0.47
100	12.09	11.28	11.92	43.33	47.00	49.27	0.94	0.96	0.96	0.40	0.43	0.45

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	69.43	62.57	58.00	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	63.86	58.55	71.68	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	65.21	61.36	63.98	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	71.51	64.00	58.39	2.49	3.58	4.57	0.52	0.71	0.86	0.02	0.04	0.06
40	65.69	59.71	72.73	2.11	3.37	4.26	0.45	0.69	0.83	0.01	0.02	0.02
100	65.53	63.15	64.85	1.74	2.89	3.72	0.38	0.61	0.77	0.00	0.00	0.01
	LASSO											
20	71.15	63.81	58.62	5.86	6.17	6.55	0.65	0.74	0.80	0.16	0.16	0.17
40	65.64	58.78	72.31	7.37	7.82	8.05	0.61	0.72	0.79	0.12	0.12	0.12
100	68.64	63.49	65.05	10.23	10.36	9.88	0.57	0.68	0.75	0.08	0.08	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	76.37	66.19	60.83	5.86	6.17	6.55	0.65	0.74	0.80	0.16	0.16	0.17
40	71.62	64.23	76.14	7.37	7.82	8.05	0.61	0.72	0.79	0.12	0.12	0.12
100	82.05	71.49	71.90	10.23	10.36	9.88	0.57	0.68	0.75	0.08	0.08	0.07
	A-LASSO											
20	74.27	65.89	60.46	4.26	4.51	4.82	0.52	0.60	0.67	0.11	0.11	0.11
40	70.61	62.33	75.17	5.59	6.00	6.19	0.51	0.61	0.68	0.09	0.09	0.09
100	79.63	69.76	70.80	8.11	8.37	8.17	0.49	0.60	0.67	0.06	0.06	0.05
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	75.96	66.61	61.08	4.26	4.51	4.82	0.52	0.60	0.67	0.11	0.11	0.11
40	72.24	63.79	76.04	5.59	6.00	6.19	0.51	0.61	0.68	0.09	0.09	0.09
100	81.59	71.57	71.96	8.11	8.37	8.17	0.49	0.60	0.67	0.06	0.06	0.05
	Boosting											
20	72.22	65.02	59.44	3.65	3.71	3.86	0.55	0.62	0.69	0.07	0.06	0.05
40	67.34	60.07	74.17	5.19	4.89	4.84	0.54	0.62	0.69	0.08	0.06	0.05
100	75.45	64.80	66.42	11.45	8.63	7.69	0.54	0.61	0.67	0.09	0.06	0.05
	Boosting for variable selection only. LS for estimation/forecasting.											
20	73.51	64.83	59.76	3.65	3.71	3.86	0.55	0.62	0.69	0.07	0.06	0.05
40	70.55	62.88	74.42	5.19	4.89	4.84	0.54	0.62	0.69	0.08	0.06	0.05
100	89.64	73.08	71.84	11.45	8.63	7.69	0.54	0.61	0.67	0.09	0.06	0.05

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	72.07	63.71	59.72	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	65.77	59.95	74.18	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	67.21	63.44	65.55	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	73.41	65.83	59.88	2.49	3.58	4.57	0.52	0.71	0.86	0.02	0.04	0.06
40	66.08	60.20	75.93	2.11	3.37	4.26	0.45	0.69	0.83	0.01	0.02	0.02
100	65.96	64.78	65.76	1.74	2.89	3.72	0.38	0.61	0.77	0.00	0.00	0.01
	LASSO											
20	79.09	68.02	63.50	5.86	6.17	6.55	0.65	0.74	0.80	0.16	0.16	0.17
40	72.99	67.27	81.07	7.37	7.82	8.05	0.61	0.72	0.79	0.12	0.12	0.12
100	84.04	76.52	74.54	10.23	10.36	9.88	0.57	0.68	0.75	0.08	0.08	0.07
	A-LASSO											
20	77.97	68.21	63.69	4.26	4.51	4.82	0.52	0.60	0.67	0.11	0.11	0.11
40	73.29	64.85	80.55	5.59	6.00	6.19	0.51	0.61	0.68	0.09	0.09	0.09
100	83.30	75.81	73.69	8.11	8.37	8.17	0.49	0.60	0.67	0.06	0.06	0.05
	Boosting											
20	75.68	66.22	61.82	3.65	3.71	3.86	0.55	0.62	0.69	0.07	0.06	0.05
40	72.16	64.47	78.63	5.19	4.89	4.84	0.54	0.62	0.69	0.08	0.06	0.05
100	91.72	77.53	73.52	11.45	8.63	7.69	0.54	0.61	0.67	0.09	0.06	0.05
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	73.46	65.02	61.18	1.84	2.69	3.51	0.36	0.48	0.56	0.02	0.04	0.06
40	65.33	60.75	76.85	1.71	2.89	4.15	0.32	0.45	0.54	0.01	0.03	0.05
100	67.44	65.23	69.24	1.66	3.45	5.70	0.26	0.38	0.48	0.01	0.02	0.04
	LASSO											
20	74.83	65.53	60.95	5.97	5.97	6.14	0.58	0.61	0.63	0.18	0.18	0.18
40	68.47	62.43	75.81	8.88	9.07	9.05	0.55	0.59	0.61	0.17	0.17	0.17
100	74.39	68.78	69.63	16.97	18.11	18.08	0.52	0.57	0.59	0.15	0.16	0.16
	A-LASSO											
20	77.98	67.62	62.31	4.58	4.61	4.75	0.48	0.51	0.53	0.13	0.13	0.13
40	72.52	67.09	78.02	7.02	7.22	7.24	0.47	0.51	0.54	0.13	0.13	0.13
100	84.64	75.49	76.04	13.48	14.65	14.77	0.45	0.51	0.54	0.12	0.13	0.13
	Boosting											
20	79.52	70.90	64.85	5.38	6.23	7.07	0.56	0.63	0.68	0.16	0.19	0.22
40	76.15	70.73	91.25	10.04	11.76	13.69	0.58	0.65	0.70	0.19	0.23	0.27
100	94.31	86.72	88.62	28.88	33.31	37.76	0.61	0.69	0.74	0.26	0.31	0.35

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	75.92	67.24	62.37	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	69.18	63.33	77.13	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	70.89	67.40	69.56	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	76.38	69.40	62.44	2.49	3.58	4.57	0.52	0.71	0.86	0.02	0.04	0.06
40	67.43	62.11	79.85	2.11	3.37	4.26	0.45	0.69	0.83	0.01	0.02	0.02
100	67.53	67.62	68.91	1.74	2.89	3.72	0.38	0.61	0.77	0.00	0.00	0.01
	LASSO											
20	82.79	72.47	67.13	5.86	6.17	6.55	0.65	0.74	0.80	0.16	0.16	0.17
40	76.17	71.85	86.14	7.37	7.82	8.05	0.61	0.72	0.79	0.12	0.12	0.12
100	88.80	82.86	80.86	10.23	10.36	9.88	0.57	0.68	0.75	0.08	0.08	0.07
	A-LASSO											
20	80.90	71.95	66.56	4.26	4.51	4.82	0.52	0.60	0.67	0.11	0.11	0.11
40	75.93	67.65	85.09	5.59	6.00	6.19	0.51	0.61	0.68	0.09	0.09	0.09
100	87.08	80.45	78.65	8.11	8.37	8.17	0.49	0.60	0.67	0.06	0.06	0.05
	Boosting											
20	78.85	69.68	63.98	3.65	3.71	3.86	0.55	0.62	0.69	0.07	0.06	0.05
40	75.00	66.92	82.17	5.19	4.89	4.84	0.54	0.62	0.69	0.08	0.06	0.05
100	96.04	83.14	78.06	11.45	8.63	7.69	0.54	0.61	0.67	0.09	0.06	0.05
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	78.41	71.21	68.06	2.51	4.05	5.46	0.35	0.47	0.56	0.06	0.11	0.16
40	69.92	69.21	90.54	3.15	5.93	8.79	0.31	0.45	0.55	0.05	0.10	0.16
100	72.73	77.44	83.26	4.95	10.98	18.27	0.26	0.40	0.51	0.04	0.09	0.16
	LASSO											
20	79.47	70.28	64.99	6.59	6.42	6.62	0.54	0.55	0.56	0.22	0.21	0.22
40	76.33	72.52	84.94	11.99	12.23	12.29	0.54	0.57	0.58	0.25	0.25	0.25
100	81.70	78.72	78.47	22.61	28.99	28.66	0.51	0.59	0.60	0.21	0.27	0.26
	A-LASSO											
20	83.16	73.12	67.33	5.10	4.99	5.12	0.45	0.46	0.47	0.17	0.16	0.16
40	80.63	78.23	86.17	9.51	9.67	9.77	0.46	0.49	0.51	0.19	0.19	0.19
100	91.52	86.31	85.21	17.78	22.51	22.38	0.44	0.51	0.52	0.16	0.20	0.20
	Boosting											
20	92.64	85.03	79.72	8.04	9.36	10.41	0.62	0.69	0.73	0.28	0.33	0.37
40	95.06	93.72	116.17	17.98	20.35	22.26	0.67	0.73	0.78	0.38	0.44	0.48
100	108.30	105.00	104.27	43.29	47.07	49.39	0.67	0.73	0.75	0.41	0.44	0.46

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	20.51	18.61	17.18	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	18.89	17.33	21.16	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	19.24	18.17	18.95	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	21.46	18.93	17.47	5.07	6.11	7.12	0.92	0.97	0.99	0.07	0.11	0.16
40	19.16	17.75	21.75	4.73	5.90	6.73	0.90	0.98	0.99	0.03	0.05	0.07
100	19.48	18.53	19.28	4.31	5.43	6.21	0.86	0.96	0.99	0.01	0.02	0.02
	LASSO											
20	21.67	19.35	17.57	7.18	7.24	7.48	0.88	0.93	0.96	0.18	0.18	0.18
40	20.05	17.76	21.70	8.89	9.09	9.09	0.86	0.93	0.96	0.14	0.13	0.13
100	21.00	19.07	19.61	11.89	11.79	11.16	0.83	0.92	0.95	0.09	0.08	0.07
	LASSO for variable selection only. LS for estimation/forecasting.											
20	23.15	19.92	18.22	7.18	7.24	7.48	0.88	0.93	0.96	0.18	0.18	0.18
40	22.09	19.43	22.84	8.89	9.09	9.09	0.86	0.93	0.96	0.14	0.13	0.13
100	25.04	21.43	21.70	11.89	11.79	11.16	0.83	0.92	0.95	0.09	0.08	0.07
	A-LASSO											
20	22.55	19.91	18.09	5.39	5.60	5.87	0.77	0.84	0.90	0.12	0.11	0.11
40	21.55	19.09	22.49	6.86	7.25	7.36	0.77	0.86	0.91	0.09	0.10	0.09
100	24.39	20.79	21.43	9.59	9.73	9.46	0.76	0.86	0.91	0.07	0.06	0.06
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	23.09	20.05	18.25	5.39	5.60	5.87	0.77	0.84	0.90	0.12	0.11	0.11
40	22.24	19.45	22.81	6.86	7.25	7.36	0.77	0.86	0.91	0.09	0.10	0.09
100	25.35	21.27	21.84	9.59	9.73	9.46	0.76	0.86	0.91	0.07	0.06	0.06
	Boosting											
20	23.41	21.17	18.64	4.79	4.85	4.93	0.81	0.88	0.93	0.08	0.07	0.06
40	22.11	19.11	23.38	6.37	6.03	5.94	0.81	0.88	0.93	0.08	0.06	0.06
100	24.38	20.46	20.99	12.21	9.73	8.78	0.79	0.87	0.91	0.09	0.06	0.05
	Boosting for variable selection only. LS for estimation/forecasting.											
20	22.29	19.45	17.77	4.79	4.85	4.93	0.81	0.88	0.93	0.08	0.07	0.06
40	21.55	18.78	22.26	6.37	6.03	5.94	0.81	0.88	0.93	0.08	0.06	0.06
100	26.87	21.52	21.66	12.21	9.73	8.78	0.79	0.87	0.91	0.09	0.06	0.05

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	21.30	18.94	17.68	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	19.46	17.78	21.97	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	19.82	18.78	19.37	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	22.24	19.36	18.40	5.07	6.11	7.12	0.92	0.97	0.99	0.07	0.11	0.16
40	19.61	18.40	23.20	4.73	5.90	6.73	0.90	0.98	0.99	0.03	0.05	0.07
100	19.73	19.43	20.00	4.31	5.43	6.21	0.86	0.96	0.99	0.01	0.02	0.02
	LASSO											
20	24.16	20.75	19.11	7.18	7.24	7.48	0.88	0.93	0.96	0.18	0.18	0.18
40	22.69	20.21	24.37	8.89	9.09	9.09	0.86	0.93	0.96	0.14	0.13	0.13
100	25.80	23.17	22.61	11.89	11.79	11.16	0.83	0.92	0.95	0.09	0.08	0.07
	A-LASSO											
20	23.77	20.66	19.03	5.39	5.60	5.87	0.77	0.84	0.90	0.12	0.11	0.11
40	22.71	20.05	24.26	6.86	7.25	7.36	0.77	0.86	0.91	0.09	0.10	0.09
100	25.87	22.66	22.72	9.59	9.73	9.46	0.76	0.86	0.91	0.07	0.06	0.06
	Boosting											
20	23.13	20.07	18.57	4.79	4.85	4.93	0.81	0.88	0.93	0.08	0.07	0.06
40	22.01	19.35	23.72	6.37	6.03	5.94	0.81	0.88	0.93	0.08	0.06	0.06
100	27.93	22.95	22.21	12.21	9.73	8.78	0.79	0.87	0.91	0.09	0.06	0.05
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	22.19	19.61	18.26	3.65	4.33	5.08	0.72	0.77	0.81	0.04	0.06	0.09
40	19.91	18.15	22.99	3.52	4.59	5.77	0.69	0.75	0.80	0.02	0.04	0.06
100	20.36	19.51	20.74	3.40	5.17	7.35	0.63	0.71	0.76	0.01	0.02	0.04
	LASSO											
20	23.03	20.14	18.42	7.61	7.67	7.92	0.83	0.85	0.87	0.22	0.21	0.22
40	21.19	19.17	23.25	10.90	11.18	11.17	0.81	0.85	0.86	0.19	0.19	0.19
100	22.81	21.30	21.56	18.84	20.65	21.04	0.78	0.84	0.85	0.16	0.17	0.18
	A-LASSO											
20	23.86	20.60	18.77	5.90	6.04	6.29	0.73	0.77	0.80	0.15	0.15	0.15
40	22.51	20.63	24.14	8.72	9.02	9.11	0.73	0.78	0.81	0.14	0.15	0.15
100	26.24	23.62	23.70	15.13	16.87	17.33	0.72	0.79	0.81	0.12	0.14	0.14
	Boosting											
20	25.67	23.16	20.77	6.43	7.25	8.02	0.80	0.85	0.89	0.16	0.19	0.22
40	24.98	22.87	28.94	11.07	12.74	14.61	0.81	0.87	0.90	0.20	0.23	0.28
100	30.06	28.00	28.48	29.72	34.12	38.43	0.83	0.89	0.91	0.26	0.31	0.35

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	22.47	19.96	18.48	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	20.47	18.82	22.89	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	20.90	19.92	20.53	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	23.57	20.58	19.77	5.07	6.11	7.12	0.92	0.97	0.99	0.07	0.11	0.16
40	20.51	19.61	25.29	4.73	5.90	6.73	0.90	0.98	0.99	0.03	0.05	0.07
100	20.49	21.04	21.68	4.31	5.43	6.21	0.86	0.96	0.99	0.01	0.02	0.02
	LASSO											
20	25.55	22.37	20.44	7.18	7.24	7.48	0.88	0.93	0.96	0.18	0.18	0.18
40	24.11	21.71	26.07	8.89	9.09	9.09	0.86	0.93	0.96	0.14	0.13	0.13
100	27.58	25.23	24.80	11.89	11.79	11.16	0.83	0.92	0.95	0.09	0.08	0.07
	A-LASSO											
20	24.86	22.00	20.11	5.39	5.60	5.87	0.77	0.84	0.90	0.12	0.11	0.11
40	23.78	21.23	25.86	6.86	7.25	7.36	0.77	0.86	0.91	0.09	0.10	0.09
100	27.08	24.37	24.49	9.59	9.73	9.46	0.76	0.86	0.91	0.07	0.06	0.06
	Boosting											
20	24.35	21.32	19.57	4.79	4.85	4.93	0.81	0.88	0.93	0.08	0.07	0.06
40	23.01	20.22	25.14	6.37	6.03	5.94	0.81	0.88	0.93	0.08	0.06	0.06
100	29.65	24.69	23.65	12.21	9.73	8.78	0.79	0.87	0.91	0.09	0.06	0.05
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	24.00	21.87	20.31	3.92	5.37	6.72	0.62	0.69	0.75	0.07	0.13	0.19
40	21.75	21.18	27.64	4.58	7.31	10.14	0.59	0.68	0.75	0.06	0.11	0.18
100	22.91	23.74	25.67	6.25	12.37	19.62	0.54	0.65	0.73	0.04	0.10	0.17
	LASSO											
20	24.70	21.84	19.90	8.25	8.17	8.38	0.77	0.78	0.79	0.26	0.25	0.26
40	23.65	22.59	26.43	13.93	14.29	14.30	0.78	0.80	0.80	0.27	0.28	0.28
100	25.34	24.24	24.13	24.04	31.00	30.99	0.75	0.80	0.82	0.21	0.28	0.28
	A-LASSO											
20	25.68	22.53	20.54	6.40	6.39	6.58	0.67	0.70	0.72	0.18	0.18	0.19
40	25.13	24.15	27.13	11.15	11.41	11.48	0.69	0.72	0.74	0.21	0.21	0.21
100	28.70	26.75	26.17	19.00	24.14	24.29	0.68	0.74	0.76	0.16	0.21	0.21
	Boosting											
20	30.02	27.95	25.67	8.87	10.13	11.08	0.80	0.86	0.89	0.28	0.34	0.38
40	30.78	30.28	36.93	18.71	21.02	22.86	0.84	0.89	0.92	0.38	0.44	0.48
100	34.53	34.04	33.58	43.88	47.60	49.86	0.84	0.88	0.90	0.41	0.44	0.46

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	28.87	25.69	24.13	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	26.29	24.34	28.05	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	27.06	25.75	26.63	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	29.25	26.04	24.51	4.42	5.51	6.39	0.81	0.93	0.97	0.06	0.09	0.12
40	27.08	25.16	28.35	4.13	5.34	6.11	0.78	0.92	0.97	0.03	0.04	0.06
100	27.69	26.10	27.31	3.89	5.07	5.68	0.73	0.90	0.95	0.01	0.01	0.02
	LASSO											
20	30.31	26.73	24.93	7.04	7.47	7.91	0.78	0.85	0.89	0.20	0.20	0.22
40	27.91	25.57	28.69	8.99	9.90	10.38	0.75	0.84	0.89	0.15	0.16	0.17
100	29.03	27.52	28.46	11.90	13.27	13.63	0.71	0.80	0.86	0.09	0.10	0.10
	LASSO for variable selection only. LS for estimation/forecasting.											
20	32.21	28.09	25.63	7.04	7.47	7.91	0.78	0.85	0.89	0.20	0.20	0.22
40	30.04	27.62	30.41	8.99	9.90	10.38	0.75	0.84	0.89	0.15	0.16	0.17
100	33.67	30.47	31.39	11.90	13.27	13.63	0.71	0.80	0.86	0.09	0.10	0.10
	A-LASSO											
20	31.61	27.63	25.38	5.28	5.67	6.07	0.65	0.73	0.80	0.13	0.14	0.14
40	29.30	27.00	29.88	6.89	7.67	8.11	0.65	0.75	0.81	0.11	0.12	0.12
100	32.45	29.75	30.64	9.27	10.63	11.02	0.63	0.74	0.80	0.07	0.08	0.08
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	32.13	28.02	25.75	5.28	5.67	6.07	0.65	0.73	0.80	0.13	0.14	0.14
40	30.08	27.66	30.48	6.89	7.67	8.11	0.65	0.75	0.81	0.11	0.12	0.12
100	33.64	30.53	31.26	9.27	10.63	11.02	0.63	0.74	0.80	0.07	0.08	0.08
	Boosting											
20	30.49	26.72	24.84	4.46	4.67	4.76	0.69	0.76	0.81	0.08	0.08	0.08
40	28.15	25.28	28.83	6.20	6.15	6.07	0.69	0.77	0.82	0.09	0.08	0.07
100	28.99	27.22	27.94	11.25	10.05	9.50	0.67	0.75	0.80	0.09	0.07	0.06
	Boosting for variable selection only. LS for estimation/forecasting.											
20	31.45	27.05	24.96	4.46	4.67	4.76	0.69	0.76	0.81	0.08	0.08	0.08
40	30.04	26.83	29.50	6.20	6.15	6.07	0.69	0.77	0.82	0.09	0.08	0.07
100	34.66	30.62	30.65	11.25	10.05	9.50	0.67	0.75	0.80	0.09	0.07	0.06

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 (18)-(19), intercepts are given by $d_t = \sum_{j=1}^{k} \beta_{jt} \mu_{jt}$ where parameter instabilities in μ_{jt} is given by (20)-(21), and $\rho_{y,t}$ is zero in experiments without dynamics, and given by (22) in experiments with dynamics. u_t is given by a GARCH(1,1). See Section 6 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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	28.11	24.94	22.95	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	25.99	23.34	27.74	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	26.55	24.96	25.34	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	28.67	25.57	23.76	4.42	5.51	6.39	0.81	0.93	0.97	0.06	0.09	0.12
40	26.72	24.36	28.71	4.13	5.34	6.11	0.78	0.92	0.97	0.03	0.04	0.06
100	27.33	25.69	26.22	3.89	5.07	5.68	0.73	0.90	0.95	0.01	0.01	0.02
	LASSO											
20	32.00	28.08	25.13	7.04	7.47	7.91	0.78	0.85	0.89	0.20	0.20	0.22
40	30.08	27.28	31.95	8.99	9.90	10.38	0.75	0.84	0.89	0.15	0.16	0.17
100	33.99	31.95	31.58	11.90	13.27	13.63	0.71	0.80	0.86	0.09	0.10	0.10
	A-LASSO											
20	31.89	27.80	25.16	5.28	5.67	6.07	0.65	0.73	0.80	0.13	0.14	0.14
40	30.03	27.05	31.40	6.89	7.67	8.11	0.65	0.75	0.81	0.11	0.12	0.12
100	34.01	31.34	30.79	9.27	10.63	11.02	0.63	0.74	0.80	0.07	0.08	0.08
	Boosting											
20	30.93	26.95	24.37	4.46	4.67	4.76	0.69	0.76	0.81	0.08	0.08	0.08
40	30.18	26.66	29.99	6.20	6.15	6.07	0.69	0.77	0.82	0.09	0.08	0.07
100	35.62	31.41	30.43	11.25	10.05	9.50	0.67	0.75	0.80	0.09	0.07	0.06
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	29.58	26.42	23.95	3.78	4.98	6.13	0.65	0.74	0.81	0.06	0.10	0.15
40	27.52	25.17	30.26	4.11	6.15	8.05	0.61	0.72	0.80	0.04	0.08	0.12
100	28.03	28.08	29.18	5.18	8.87	12.67	0.56	0.68	0.75	0.03	0.06	0.10
	LASSO											
20	30.41	26.42	24.12	7.13	7.42	7.63	0.73	0.78	0.82	0.21	0.21	0.22
40	28.59	25.80	29.48	9.64	10.28	10.44	0.70	0.76	0.80	0.17	0.18	0.18
100	30.30	28.65	28.61	15.48	16.59	16.50	0.65	0.72	0.76	0.13	0.14	0.13
	A-LASSO											
20	31.33	27.15	24.38	5.50	5.80	5.98	0.63	0.69	0.74	0.15	0.15	0.15
40	30.14	27.38	31.00	7.62	8.18	8.42	0.61	0.69	0.74	0.13	0.14	0.14
100	33.71	30.96	31.28	12.30	13.41	13.50	0.59	0.66	0.71	0.10	0.11	0.11
	Boosting											
20	30.42	27.30	25.01	5.99	6.92	7.72	0.69	0.77	0.83	0.16	0.19	0.22
40	29.67	27.54	32.25	10.24	12.10	13.92	0.70	0.78	0.83	0.19	0.22	0.26
100	35.02	32.18	33.88	26.74	32.14	36.59	0.70	0.78	0.83	0.24	0.29	0.33

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	28.48	25.64	23.58	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	26.72	24.10	28.65	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	26.88	25.69	26.26	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	29.24	26.36	24.98	4.42	5.51	6.39	0.81	0.93	0.97	0.06	0.09	0.12
40	27.33	25.16	30.63	4.13	5.34	6.11	0.78	0.92	0.97	0.03	0.04	0.06
100	27.70	26.79	27.47	3.89	5.07	5.68	0.73	0.90	0.95	0.01	0.01	0.02
	LASSO											
20	32.57	29.32	26.67	7.04	7.47	7.91	0.78	0.85	0.89	0.20	0.20	0.22
40	31.02	28.49	34.06	8.99	9.90	10.38	0.75	0.84	0.89	0.15	0.16	0.17
100	35.75	35.32	34.44	11.90	13.27	13.63	0.71	0.80	0.86	0.09	0.10	0.10
	A-LASSO											
20	32.43	28.56	26.21	5.28	5.67	6.07	0.65	0.73	0.80	0.13	0.14	0.14
40	30.72	27.76	32.81	6.89	7.67	8.11	0.65	0.75	0.81	0.11	0.12	0.12
100	35.32	33.70	32.74	9.27	10.63	11.02	0.63	0.74	0.80	0.07	0.08	0.08
	Boosting											
20	31.16	27.72	25.59	4.46	4.67	4.76	0.69	0.76	0.81	0.08	0.08	0.08
40	30.88	27.55	31.20	6.20	6.15	6.07	0.69	0.77	0.82	0.09	0.08	0.07
100	37.56	33.47	32.40	11.25	10.05	9.50	0.67	0.75	0.80	0.09	0.07	0.06
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	31.14	29.06	27.11	4.87	6.71	8.31	0.62	0.71	0.78	0.12	0.19	0.26
40	30.05	29.87	37.71	6.64	10.28	13.46	0.58	0.70	0.78	0.11	0.19	0.26
100	33.39	36.66	42.02	11.31	19.36	27.60	0.53	0.67	0.75	0.09	0.17	0.25
	LASSO											
20	31.31	27.55	25.21	7.42	7.61	7.63	0.69	0.74	0.75	0.23	0.23	0.23
40	30.17	28.45	32.23	11.69	11.92	12.11	0.67	0.72	0.75	0.22	0.23	0.23
100	33.59	31.12	30.97	20.62	22.23	22.78	0.63	0.69	0.73	0.18	0.19	0.20
	A-LASSO											
20	32.42	28.40	26.15	5.74	5.91	5.95	0.60	0.64	0.67	0.17	0.17	0.16
40	32.30	30.51	34.28	9.26	9.42	9.65	0.59	0.65	0.68	0.17	0.17	0.17
100	37.03	34.44	33.85	15.96	17.35	17.89	0.56	0.63	0.67	0.14	0.15	0.15
	Boosting											
20	33.77	31.31	30.02	8.38	9.81	10.83	0.73	0.81	0.85	0.27	0.33	0.37
40	35.26	35.00	42.12	17.72	20.39	22.32	0.76	0.84	0.88	0.37	0.43	0.47
100	40.68	38.33	40.39	42.16	46.42	48.87	0.75	0.82	0.85	0.39	0.43	0.45

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	10.59	9.37	8.83	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	9.61	9.03	10.18	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	9.91	9.48	9.70	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	10.93	9.56	9.13	6.13	7.22	8.21	0.96	0.99	1.00	0.11	0.16	0.21
40	9.84	9.48	10.52	5.97	7.12	7.91	0.96	0.99	1.00	0.05	0.08	0.10
100	10.33	9.85	10.19	5.87	7.00	7.65	0.94	0.99	1.00	0.02	0.03	0.04
	LASSO											
20	11.54	9.85	9.31	8.78	9.42	9.67	0.90	0.94	0.96	0.26	0.28	0.29
40	10.54	9.83	10.66	11.95	13.22	14.14	0.89	0.94	0.96	0.21	0.24	0.26
100	10.88	10.23	10.55	16.46	18.79	20.73	0.87	0.92	0.95	0.13	0.15	0.17
	LASSO for variable selection only. LS for estimation/forecasting.											
20	12.10	10.35	9.57	8.78	9.42	9.67	0.90	0.94	0.96	0.26	0.28	0.29
40	11.72	10.93	11.44	11.95	13.22	14.14	0.89	0.94	0.96	0.21	0.24	0.26
100	12.95	11.99	12.51	16.46	18.79	20.73	0.87	0.92	0.95	0.13	0.15	0.17
	A-LASSO											
20	11.77	10.07	9.43	6.79	7.41	7.62	0.81	0.87	0.91	0.18	0.20	0.20
40	11.34	10.49	11.16	9.25	10.39	11.20	0.81	0.89	0.92	0.15	0.17	0.19
100	12.02	11.42	11.79	12.84	14.83	16.51	0.81	0.88	0.92	0.10	0.11	0.13
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	12.01	10.24	9.54	6.79	7.41	7.62	0.81	0.87	0.91	0.18	0.20	0.20
40	11.70	10.94	11.45	9.25	10.39	11.20	0.81	0.89	0.92	0.15	0.17	0.19
100	12.68	11.91	12.30	12.84	14.83	16.51	0.81	0.88	0.92	0.10	0.11	0.13
	Boosting											
20	11.56	10.00	9.26	5.66	5.79	5.82	0.84	0.90	0.92	0.12	0.11	0.11
40	10.89	9.60	10.66	7.88	7.84	7.74	0.84	0.90	0.93	0.11	0.11	0.10
100	10.97	10.16	10.28	13.81	12.55	12.22	0.83	0.89	0.92	0.10	0.09	0.09
	Boosting for variable selection only. LS for estimation/forecasting.											
20	11.66	9.86	9.13	5.66	5.79	5.82	0.84	0.90	0.92	0.12	0.11	0.11
40	11.27	10.07	10.90	7.88	7.84	7.74	0.84	0.90	0.93	0.11	0.11	0.10
100	12.98	11.26	11.25	13.81	12.55	12.22	0.83	0.89	0.92	0.10	0.09	0.09

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	9.53	8.16	7.32	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	8.75	7.71	8.82	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	9.06	8.19	8.11	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	9.97	8.60	7.89	6.13	7.22	8.21	0.96	0.99	1.00	0.11	0.16	0.21
40	9.17	8.41	9.64	5.97	7.12	7.91	0.96	0.99	1.00	0.05	0.08	0.10
100	9.77	8.69	8.70	5.87	7.00	7.65	0.94	0.99	1.00	0.02	0.03	0.04
	LASSO											
20	11.36	9.51	8.39	8.78	9.42	9.67	0.90	0.94	0.96	0.26	0.28	0.29
40	11.19	10.09	10.75	11.95	13.22	14.14	0.89	0.94	0.96	0.21	0.24	0.26
100	12.83	11.73	11.82	16.46	18.79	20.73	0.87	0.92	0.95	0.13	0.15	0.17
	A-LASSO											
20	11.30	9.38	8.30	6.79	7.41	7.62	0.81	0.87	0.91	0.18	0.20	0.20
40	11.11	9.94	10.62	9.25	10.39	11.20	0.81	0.89	0.92	0.15	0.17	0.19
100	12.44	11.52	11.57	12.84	14.83	16.51	0.81	0.88	0.92	0.10	0.11	0.13
	Boosting											
20	10.83	9.06	7.85	5.66	5.79	5.82	0.84	0.90	0.92	0.12	0.11	0.11
40	10.86	9.16	10.11	7.88	7.84	7.74	0.84	0.90	0.93	0.11	0.11	0.10
100	12.79	10.90	10.55	13.81	12.55	12.22	0.83	0.89	0.92	0.10	0.09	0.09
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	10.32	9.00	7.89	5.30	6.49	7.63	0.82	0.87	0.90	0.10	0.15	0.20
40	9.78	8.86	10.14	6.04	8.16	10.19	0.80	0.86	0.90	0.07	0.12	0.16
100	10.18	10.10	10.14	7.95	12.09	16.33	0.77	0.84	0.88	0.05	0.09	0.13
	LASSO											
20	10.84	9.10	7.91	8.95	9.22	9.37	0.87	0.92	0.95	0.27	0.28	0.28
40	10.34	9.13	9.88	12.68	13.26	13.37	0.86	0.91	0.93	0.23	0.24	0.24
100	10.98	10.05	9.75	19.68	21.18	21.59	0.83	0.89	0.92	0.16	0.18	0.18
	A-LASSO											
20	11.00	9.21	7.87	7.00	7.32	7.48	0.79	0.86	0.90	0.19	0.19	0.19
40	10.98	9.67	10.25	10.08	10.59	10.80	0.80	0.87	0.90	0.17	0.18	0.18
100	12.11	10.90	10.71	15.50	17.01	17.51	0.78	0.85	0.89	0.12	0.14	0.14
	Boosting											
20	10.81	9.39	8.43	7.02	7.82	8.53	0.84	0.90	0.94	0.18	0.21	0.24
40	10.79	9.61	10.60	11.52	13.25	14.85	0.85	0.91	0.94	0.20	0.24	0.28
100	12.18	10.88	11.35	27.20	32.73	37.25	0.84	0.90	0.93	0.24	0.29	0.34

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	9.13	8.04	7.30	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	8.52	7.61	8.86	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	8.71	8.06	8.18	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	9.69	8.79	8.25	6.13	7.22	8.21	0.96	0.99	1.00	0.11	0.16	0.21
40	9.07	8.52	10.06	5.97	7.12	7.91	0.96	0.99	1.00	0.05	0.08	0.10
100	9.65	8.72	8.95	5.87	7.00	7.65	0.94	0.99	1.00	0.02	0.03	0.04
	LASSO											
20	11.11	9.68	8.81	8.78	9.42	9.67	0.90	0.94	0.96	0.26	0.28	0.29
40	11.23	10.32	11.21	11.95	13.22	14.14	0.89	0.94	0.96	0.21	0.24	0.26
100	13.44	12.66	12.90	16.46	18.79	20.73	0.87	0.92	0.95	0.13	0.15	0.17
	A-LASSO											
20	11.02	9.41	8.52	6.79	7.41	7.62	0.81	0.87	0.91	0.18	0.20	0.20
40	10.96	10.04	10.97	9.25	10.39	11.20	0.81	0.89	0.92	0.15	0.17	0.19
100	12.64	12.23	12.32	12.84	14.83	16.51	0.81	0.88	0.92	0.10	0.11	0.13
	Boosting											
20	10.44	9.04	8.01	5.66	5.79	5.82	0.84	0.90	0.92	0.12	0.11	0.11
40	10.76	9.21	10.37	7.88	7.84	7.74	0.84	0.90	0.93	0.11	0.11	0.10
100	13.12	11.48	11.11	13.81	12.55	12.22	0.83	0.89	0.92	0.10	0.09	0.09
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	10.68	9.57	8.82	6.35	8.16	9.74	0.77	0.83	0.87	0.16	0.24	0.31
40	10.80	10.32	12.48	8.78	12.47	15.65	0.75	0.82	0.88	0.14	0.23	0.30
100	12.45	13.91	15.18	14.85	23.23	31.55	0.72	0.81	0.86	0.12	0.20	0.28
	LASSO											
20	10.76	9.31	8.19	9.23	9.46	9.48	0.86	0.90	0.92	0.29	0.29	0.29
40	10.62	9.73	10.56	14.34	14.72	15.02	0.85	0.90	0.91	0.27	0.28	0.28
100	11.80	10.56	10.43	23.37	25.38	26.74	0.81	0.88	0.90	0.20	0.22	0.23
	A-LASSO											
20	10.87	9.46	8.37	7.21	7.44	7.51	0.78	0.84	0.87	0.20	0.20	0.20
40	11.39	10.36	11.04	11.41	11.68	12.03	0.79	0.85	0.87	0.21	0.21	0.21
100	12.93	11.56	11.46	17.98	19.74	20.84	0.76	0.83	0.87	0.15	0.16	0.17
	Boosting											
20	11.55	10.64	10.03	9.21	10.53	11.47	0.87	0.92	0.95	0.29	0.34	0.38
40	12.49	11.84	13.59	18.46	21.10	23.01	0.88	0.93	0.95	0.37	0.43	0.48
100	14.16	12.82	13.31	42.35	46.88	49.42	0.87	0.92	0.94	0.39	0.43	0.46

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	75.71	66.08	61.78	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	67.94	62.43	75.57	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	70.50	66.03	68.84	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	77.31	66.80	61.94	1.78	2.74	3.63	0.38	0.58	0.74	0.01	0.02	0.03
40	69.50	64.16	77.07	1.49	2.52	3.34	0.33	0.55	0.71	0.00	0.01	0.01
100	71.43	68.54	70.27	1.23	2.13	2.82	0.27	0.47	0.62	0.00	0.00	0.00
	LASSO											
20	78.26	67.40	63.50	5.86	6.27	6.82	0.59	0.67	0.74	0.17	0.18	0.19
40	71.00	64.28	77.18	7.94	8.49	8.73	0.55	0.65	0.72	0.14	0.15	0.15
100	74.43	69.24	71.63	11.54	11.95	11.84	0.50	0.61	0.67	0.10	0.10	0.09
	LASSO for variable selection only. LS for estimation/forecasting.											
20	83.48	70.28	65.14	5.86	6.27	6.82	0.59	0.67	0.74	0.17	0.18	0.19
40	78.58	70.50	81.00	7.94	8.49	8.73	0.55	0.65	0.72	0.14	0.15	0.15
100	89.38	78.67	79.53	11.54	11.95	11.84	0.50	0.61	0.67	0.10	0.10	0.09
	A-LASSO											
20	82.14	69.42	65.04	4.28	4.71	5.09	0.47	0.54	0.61	0.12	0.13	0.13
40	76.05	68.80	80.34	6.12	6.56	6.86	0.45	0.55	0.62	0.11	0.11	0.11
100	85.41	75.67	77.67	9.11	9.64	9.77	0.43	0.53	0.60	0.07	0.08	0.07
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	83.73	70.57	65.77	4.28	4.71	5.09	0.47	0.54	0.61	0.12	0.13	0.13
40	78.42	70.50	81.82	6.12	6.56	6.86	0.45	0.55	0.62	0.11	0.11	0.11
100	89.02	77.88	79.20	9.11	9.64	9.77	0.43	0.53	0.60	0.07	0.08	0.07
	Boosting											
20	77.31	67.12	62.13	3.59	3.57	3.69	0.50	0.56	0.62	0.08	0.07	0.06
40	70.98	63.07	76.55	5.34	5.04	4.86	0.49	0.56	0.62	0.08	0.07	0.06
100	77.35	67.07	70.40	12.09	9.35	8.33	0.49	0.56	0.60	0.10	0.07	0.06
	Boosting for variable selection only. LS for estimation/forecasting.											
20	80.27	67.66	63.58	3.59	3.57	3.69	0.50	0.56	0.62	0.08	0.07	0.06
40	75.81	67.93	78.29	5.34	5.04	4.86	0.49	0.56	0.62	0.08	0.07	0.06
100	94.18	77.52	79.66	12.09	9.35	8.33	0.49	0.56	0.60	0.10	0.07	0.06

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	75.04	64.94	60.47	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	67.36	61.25	74.44	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	70.12	65.02	66.76	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	76.46	65.65	60.52	1.78	2.74	3.63	0.38	0.58	0.74	0.01	0.02	0.03
40	67.80	62.56	77.03	1.49	2.52	3.34	0.33	0.55	0.71	0.00	0.01	0.01
100	70.41	67.10	67.92	1.23	2.13	2.82	0.27	0.47	0.62	0.00	0.00	0.00
	LASSO											
20	83.16	69.84	65.40	5.86	6.27	6.82	0.59	0.67	0.74	0.17	0.18	0.19
40	79.01	70.18	83.74	7.94	8.49	8.73	0.55	0.65	0.72	0.14	0.15	0.15
100	89.07	80.66	79.92	11.54	11.95	11.84	0.50	0.61	0.67	0.10	0.10	0.09
	A-LASSO											
20	83.11	69.88	65.95	4.28	4.71	5.09	0.47	0.54	0.61	0.12	0.13	0.13
40	78.09	68.87	83.49	6.12	6.56	6.86	0.45	0.55	0.62	0.11	0.11	0.11
100	88.15	78.50	79.01	9.11	9.64	9.77	0.43	0.53	0.60	0.07	0.08	0.07
	Boosting											
20	80.01	67.10	63.77	3.59	3.57	3.69	0.50	0.56	0.62	0.08	0.07	0.06
40	76.17	67.13	79.47	5.34	5.04	4.86	0.49	0.56	0.62	0.08	0.07	0.06
100	95.21	78.89	78.69	12.09	9.35	8.33	0.49	0.56	0.60	0.10	0.07	0.06
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	77.30	67.14	62.56	1.54	2.46	3.36	0.30	0.44	0.55	0.02	0.03	0.06
40	68.17	63.51	78.15	1.48	2.68	4.02	0.25	0.40	0.52	0.01	0.03	0.05
100	71.87	68.17	71.51	1.53	3.31	5.57	0.21	0.34	0.45	0.01	0.02	0.04
	LASSO											
20	80.00	67.40	62.93	6.20	6.21	6.52	0.55	0.60	0.64	0.20	0.19	0.20
40	72.11	65.23	77.42	9.52	9.74	9.83	0.52	0.58	0.62	0.19	0.19	0.18
100	78.28	73.66	72.59	18.48	19.90	19.93	0.49	0.55	0.59	0.17	0.18	0.18
	A-LASSO											
20	83.11	68.96	64.09	4.79	4.82	5.09	0.46	0.50	0.55	0.15	0.14	0.15
40	76.73	69.71	79.78	7.53	7.76	7.89	0.44	0.50	0.55	0.14	0.14	0.14
100	88.96	81.22	79.27	14.68	16.02	16.24	0.43	0.49	0.53	0.13	0.14	0.14
	Boosting											
20	82.60	72.71	66.02	5.41	6.23	7.08	0.53	0.61	0.67	0.16	0.19	0.22
40	80.49	74.27	92.06	10.25	11.94	13.79	0.54	0.63	0.69	0.20	0.24	0.28
100	97.54	89.07	91.91	29.16	33.91	38.20	0.58	0.66	0.71	0.27	0.31	0.35

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	77.37	67.65	62.80	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	69.77	64.11	76.89	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	72.54	68.20	70.34	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	78.00	67.66	62.51	1.78	2.74	3.63	0.38	0.58	0.74	0.01	0.02	0.03
40	68.28	63.77	80.08	1.49	2.52	3.34	0.33	0.55	0.71	0.00	0.01	0.01
100	71.14	68.87	70.25	1.23	2.13	2.82	0.27	0.47	0.62	0.00	0.00	0.00
	LASSO											
20	85.84	73.54	69.28	5.86	6.27	6.82	0.59	0.67	0.74	0.17	0.18	0.19
40	82.61	73.75	89.00	7.94	8.49	8.73	0.55	0.65	0.72	0.14	0.15	0.15
100	92.80	87.68	86.51	11.54	11.95	11.84	0.50	0.61	0.67	0.10	0.10	0.09
	A-LASSO											
20	85.01	72.33	68.99	4.28	4.71	5.09	0.47	0.54	0.61	0.12	0.13	0.13
40	80.01	71.15	87.53	6.12	6.56	6.86	0.45	0.55	0.62	0.11	0.11	0.11
100	89.74	83.20	84.42	9.11	9.64	9.77	0.43	0.53	0.60	0.07	0.08	0.07
	Boosting											
20	81.73	69.53	66.63	3.59	3.57	3.69	0.50	0.56	0.62	0.08	0.07	0.06
40	78.95	69.26	82.99	5.34	5.04	4.86	0.49	0.56	0.62	0.08	0.07	0.06
100	100.44	83.45	82.31	12.09	9.35	8.33	0.49	0.56	0.60	0.10	0.07	0.06
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	81.17	73.13	68.83	2.42	4.00	5.42	0.33	0.47	0.57	0.05	0.11	0.16
40	72.34	72.03	94.38	3.10	5.89	8.79	0.29	0.44	0.56	0.05	0.10	0.16
100	76.96	79.30	90.15	4.94	10.95	18.17	0.25	0.40	0.52	0.04	0.09	0.16
	LASSO											
20	83.35	71.86	66.78	6.89	6.89	7.12	0.55	0.58	0.60	0.23	0.23	0.24
40	78.82	74.92	86.24	12.68	12.95	13.04	0.54	0.59	0.61	0.26	0.27	0.26
100	84.81	82.65	81.18	23.76	30.22	30.01	0.50	0.58	0.61	0.22	0.28	0.28
	A-LASSO											
20	86.88	74.40	68.83	5.35	5.35	5.51	0.46	0.48	0.51	0.18	0.17	0.17
40	83.60	80.66	88.42	10.07	10.21	10.35	0.46	0.50	0.53	0.21	0.21	0.21
100	94.40	90.83	88.68	18.65	23.45	23.48	0.43	0.51	0.54	0.17	0.21	0.21
	Boosting											
20	95.04	87.48	81.60	8.09	9.41	10.46	0.61	0.68	0.74	0.28	0.33	0.38
40	98.94	97.85	116.43	18.06	20.44	22.35	0.65	0.73	0.78	0.39	0.44	0.48
100	112.52	108.00	108.11	43.50	47.37	49.77	0.65	0.72	0.75	0.41	0.45	0.47

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
	Oracle											
20	25.58	22.02	20.60	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	22.54	20.90	24.85	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	23.50	21.92	22.73	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	25.99	22.31	20.92	3.81	4.81	5.59	0.78	0.91	0.97	0.04	0.06	0.09
40	23.23	21.32	25.18	3.53	4.62	5.33	0.74	0.90	0.96	0.01	0.03	0.04
100	23.77	22.17	23.10	3.16	4.29	4.88	0.68	0.87	0.94	0.00	0.01	0.01
	LASSO											
20	27.25	22.84	21.46	7.44	7.86	8.27	0.78	0.84	0.89	0.22	0.22	0.24
40	24.27	22.07	25.90	10.31	10.78	11.25	0.76	0.83	0.88	0.18	0.19	0.19
100	25.44	23.55	24.44	14.82	15.29	16.04	0.72	0.81	0.85	0.12	0.12	0.13
	LASSO for variable selection only. LS for estimation/forecasting.											
20	29.07	23.73	21.98	7.44	7.86	8.27	0.78	0.84	0.89	0.22	0.22	0.24
40	26.88	24.34	27.30	10.31	10.78	11.25	0.76	0.83	0.88	0.18	0.19	0.19
100	31.67	27.37	27.90	14.82	15.29	16.04	0.72	0.81	0.85	0.12	0.12	0.13
	A-LASSO											
20	28.39	23.34	21.74	5.62	6.00	6.40	0.66	0.73	0.79	0.15	0.15	0.16
40	25.89	23.34	26.86	7.95	8.49	8.85	0.66	0.75	0.81	0.13	0.14	0.14
100	29.38	25.86	27.03	11.63	12.29	13.03	0.65	0.75	0.80	0.09	0.09	0.10
	A-LASSO for variable selection only. LS for estimation/forecasting.											
20	29.04	23.73	21.98	5.62	6.00	6.40	0.66	0.73	0.79	0.15	0.15	0.16
40	26.92	24.16	27.53	7.95	8.49	8.85	0.66	0.75	0.81	0.13	0.14	0.14
100	31.03	26.76	28.10	11.63	12.29	13.03	0.65	0.75	0.80	0.09	0.09	0.10
	Boosting											
20	26.93	23.25	20.92	4.64	4.60	4.73	0.69	0.75	0.81	0.09	0.08	0.07
40	25.08	21.93	25.69	6.68	6.36	6.15	0.69	0.77	0.82	0.10	0.08	0.07
100	27.04	22.67	23.43	13.67	10.97	10.07	0.69	0.76	0.80	0.11	0.08	0.07
	Boosting for variable selection only. LS for estimation/forecasting.											
20	27.66	22.92	21.24	4.64	4.60	4.73	0.69	0.75	0.81	0.09	0.08	0.07
40	25.99	23.44	26.04	6.68	6.36	6.15	0.69	0.77	0.82	0.10	0.08	0.07
100	32.44	26.25	26.55	13.67	10.97	10.07	0.69	0.76	0.80	0.11	0.08	0.07

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	24.08	20.43	18.65	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	21.27	19.26	22.93	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	22.39	20.27	20.52	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	24.65	20.87	19.12	3.81	4.81	5.59	0.78	0.91	0.97	0.04	0.06	0.09
40	21.99	19.72	23.74	3.53	4.62	5.33	0.74	0.90	0.96	0.01	0.03	0.04
100	22.68	20.72	21.13	3.16	4.29	4.88	0.68	0.87	0.94	0.00	0.01	0.01
	LASSO											
20	28.32	22.59	20.37	7.44	7.86	8.27	0.78	0.84	0.89	0.22	0.22	0.24
40	26.12	23.33	27.08	10.31	10.78	11.25	0.76	0.83	0.88	0.18	0.19	0.19
100	31.41	27.03	26.92	14.82	15.29	16.04	0.72	0.81	0.85	0.12	0.12	0.13
	A-LASSO											
20	28.11	22.74	20.49	5.62	6.00	6.40	0.66	0.73	0.79	0.15	0.15	0.16
40	25.90	22.88	27.02	7.95	8.49	8.85	0.66	0.75	0.81	0.13	0.14	0.14
100	30.62	26.33	27.09	11.63	12.29	13.03	0.65	0.75	0.80	0.09	0.09	0.10
	Boosting											
20	26.24	22.00	19.70	4.64	4.60	4.73	0.69	0.75	0.81	0.09	0.08	0.07
40	25.47	22.50	25.02	6.68	6.36	6.15	0.69	0.77	0.82	0.10	0.08	0.07
100	32.79	25.84	25.02	13.67	10.97	10.07	0.69	0.76	0.80	0.11	0.08	0.07
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	25.28	21.47	19.45	2.94	3.87	4.69	0.60	0.72	0.79	0.03	0.05	0.08
40	22.89	20.51	24.46	2.85	4.11	5.42	0.56	0.69	0.78	0.02	0.03	0.06
100	23.84	22.00	22.60	2.80	4.74	6.96	0.50	0.64	0.73	0.01	0.02	0.04
	LASSO											
20	26.92	22.17	20.03	7.75	7.92	8.29	0.74	0.81	0.85	0.24	0.23	0.24
40	24.21	21.70	24.81	11.75	11.97	12.12	0.73	0.80	0.83	0.22	0.22	0.22
100	26.38	24.51	23.75	21.23	22.70	22.92	0.70	0.76	0.81	0.18	0.20	0.20
	A-LASSO											
20	27.60	22.40	20.14	6.04	6.22	6.54	0.65	0.72	0.77	0.17	0.17	0.17
40	25.56	22.79	25.50	9.30	9.61	9.79	0.64	0.73	0.77	0.17	0.17	0.17
100	29.97	26.82	26.00	16.80	18.34	18.73	0.63	0.71	0.76	0.14	0.16	0.16
	Boosting											
20	28.49	25.18	22.34	6.21	7.05	7.85	0.71	0.79	0.84	0.17	0.19	0.22
40	28.51	26.06	30.23	11.12	12.68	14.39	0.72	0.80	0.85	0.21	0.24	0.27
100	34.04	30.77	31.27	29.46	34.10	38.36	0.73	0.81	0.85	0.27	0.31	0.35

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**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
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	24.13	20.79	19.13	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
40	21.47	19.80	23.45	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
100	22.57	20.80	21.32	4.00	4.00	4.00	1.00	1.00	1.00	0.00	0.00	0.00
	OCMT											
20	24.87	21.25	19.91	3.81	4.81	5.59	0.78	0.91	0.97	0.04	0.06	0.09
40	22.13	20.19	25.07	3.53	4.62	5.33	0.74	0.90	0.96	0.01	0.03	0.04
100	22.82	21.40	22.20	3.16	4.29	4.88	0.68	0.87	0.94	0.00	0.01	0.01
	LASSO											
20	28.82	23.33	21.38	7.44	7.86	8.27	0.78	0.84	0.89	0.22	0.22	0.24
40	26.91	24.49	29.04	10.31	10.78	11.25	0.76	0.83	0.88	0.18	0.19	0.19
100	32.69	29.08	29.69	14.82	15.29	16.04	0.72	0.81	0.85	0.12	0.12	0.13
	A-LASSO											
20	28.36	23.36	21.30	5.62	6.00	6.40	0.66	0.73	0.79	0.15	0.15	0.16
40	26.14	23.70	28.68	7.95	8.49	8.85	0.66	0.75	0.81	0.13	0.14	0.14
100	31.35	27.64	29.03	11.63	12.29	13.03	0.65	0.75	0.80	0.09	0.09	0.10
	Boosting											
20	26.09	22.48	20.53	4.64	4.60	4.73	0.69	0.75	0.81	0.09	0.08	0.07
40	26.12	23.19	26.15	6.68	6.36	6.15	0.69	0.77	0.82	0.10	0.08	0.07
100	34.44	27.12	26.22	13.67	10.97	10.07	0.69	0.76	0.80	0.11	0.08	0.07
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	26.17	23.19	21.60	3.63	5.24	6.62	0.58	0.70	0.78	0.07	0.12	0.18
40	24.46	23.54	27.79	4.32	7.20	10.10	0.54	0.68	0.77	0.05	0.11	0.18
100	26.35	26.29	28.69	6.02	12.21	19.55	0.49	0.64	0.74	0.04	0.10	0.17
	LASSO											
20	27.66	23.35	21.23	8.45	8.56	8.84	0.74	0.78	0.81	0.28	0.27	0.28
40	25.94	24.49	27.57	14.58	14.99	15.15	0.73	0.79	0.81	0.29	0.30	0.30
100	28.26	26.92	25.96	25.44	31.60	32.20	0.69	0.77	0.81	0.23	0.29	0.29
	A-LASSO											
20	28.30	23.82	21.80	6.59	6.70	6.91	0.64	0.69	0.73	0.20	0.20	0.20
40	27.65	25.96	28.30	11.58	11.91	12.10	0.65	0.72	0.75	0.22	0.23	0.23
100	31.82	29.35	28.16	19.98	24.54	25.22	0.62	0.70	0.75	0.18	0.22	0.22
	Boosting											
20	32.68	30.40	27.62	8.68	10.06	11.04	0.75	0.83	0.87	0.28	0.34	0.38
40	34.60	33.67	38.51	18.48	20.88	22.76	0.78	0.86	0.89	0.38	0.44	0.48
100	39.09	36.94	36.77	43.75	47.53	49.90	0.78	0.84	0.87	0.41	0.44	0.46

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.

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

July 15, 2024

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 $\mathfrak{g}_i$,

[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.

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:

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

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

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

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

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

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

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

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

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

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

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

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}, $\bar{r}^g_{it}$, $r_{sp,t}$, RV_{it}, RV^g_{it}, $RV_{sp,t}$, 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^s_{it} for $s = \{3, 6, 12\}$, mav^s_{it} for $s = \{3, 6, 12\}$, gr^s_{it} for $s = \{3, 6, 12\}$, gp^s_{it} for $s = \{3, 6, 12\}$, RVG^s_{it} for $s = \{3, 6, 12\}$, RSI^s_{it} for $s = \{3, 6, 12\}$, ADX^s_{it} for $s = \{3, 6, 12\}$, $WILLR^s_{it}$ 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 for forecasting output growth

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 growth 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 growth 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 (2023) 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-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, A-Lasso and boosting. The results indicate that the OCMT procedure outperforms Lasso, A-Lasso and boosting 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. Boosting selects 11.6 forecasters on average.

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

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

Table S.3: Mean square forecast error (MSFE) and panel DM test of OCMT of one-year ahead euro area 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, A-Lasso and boosting of one-year ahead euro area 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
Boosting	4.830	5.071	5.439

	Pair-wise Panel DM tests (All countries)								
	No down-weighting			Light down-weighting			Heavy down-weighting		
	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting	Lasso	A-Lasso	Boosting
OCMT	-1.413	-1.544	-0.934	-0.990	-1.265	-0.938	-1.070	-1.267	-1.155
Lasso	-	-1.484	0.819	-	-1.589	0.144	-	-1.527	-0.417
A-Lasso	-	-	2.005	-	-	1.707	-	-	1.402

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, $\lambda = \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$.