A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models*

A. Chudik
Federal Reserve Bank of Dallas

G. Kapetanios
King's College, London

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

July 3, 2018

Abstract

This paper provides an alternative approach to penalised regression for model selection in the context of high dimensional linear regressions where the number of covariates is large, often much larger than the number of available observations. We consider the statistical significance of individual covariates one at a time, whilst taking full account of the multiple testing nature of the inferential problem involved. We refer to the proposed method as One Covariate at a Time Multiple Testing (OCMT) procedure, and use ideas from the multiple testing literature to control the probability of selecting the approximating model, the false positive rate and the false discovery rate. OCMT is easy to interpret, relates to classical statistical analysis, is valid under general assumptions, is faster to compute, and performs well in small samples. The usefulness of OCMT is also illustrated by an empirical application to forecasting U.S. output growth and inflation.

Keywords: One covariate at a time, multiple testing, model selection, high dimensionality, penalised regressions, boosting, Monte Carlo experiments
JEL Classifications: C52, C55

*We are grateful to the Editor and three anonymous referees for constructive comments on earlier versions of the paper. We have also benefited from helpful comments by Jinchi Lv, Yingying Fan, Essie Maasoumi, Mahrad Sharifvaghefi, Ron Smith, and seminar participants at the Bank of England, Federal Reserve Bank of Dallas, Third IAAE Conference at the University of Milano-Bicocca, Michigan State University, University of Michigan (Department of Statistics), University of California, Irvine, University of Kent, University of Messina, University of Southern California, and University of Warwick. The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

1 Introduction

This paper contributes to the literature by proposing a multiple testing procedure to model selection in high dimensional regression settings. The goal of the proposed procedure is to select an approximating model that encompasses the true model, and does not contain any noise variables that are uncorrelated with signal (true) variables. We use ideas from the multiple testing literature to control the probability of selecting the approximating model, the false positive rate and the false discovery rate. We refer to the proposed method as One Covariate at a Time Multiple Testing (OCMT) procedure. OCMT is computationally simple and fast even for extremely large data sets.

Our approach is to be contrasted to penalised regressions where the vector of regression coefficients, $\boldsymbol{\beta}$, of a regression of y_t on $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$, known as the active set, is estimated by $\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}} = \text{argmin}_{\beta}[\sum_{t=1}^{T}(y_t - \boldsymbol{x}'_{nt}\boldsymbol{\beta})^2 + P_{\boldsymbol{\lambda}}(\boldsymbol{\beta})]$. $P_{\boldsymbol{\lambda}}(\boldsymbol{\beta})$ is a penalty function that penalises $\boldsymbol{\beta}$, while $\boldsymbol{\lambda}$ is a vector of tuning parameters to be set by the researcher. A variety of penalty functions have been considered, yielding a wide range of penalised regression methods. Chief among them is Lasso, where $P_{\boldsymbol{\lambda}}(\boldsymbol{\beta})$ is chosen to be proportional to the L_1 norm of $\boldsymbol{\beta}$. This has subsequently been generalised to penalty functions involving L_q, $0 \leq q \leq 2$, norms. While these techniques have found considerable use in econometrics,[1] their theoretical properties have been mainly analysed in the statistical literature starting with the seminal work of Tibshirani (1996) and followed up with important contributions by Fan and Li (2001), Antoniadis and Fan (2001), Efron et al. (2004), Zhou and Hastie (2005), Candes and Tao (2007), Lv and Fan (2009), Bickel et al. (2009), Zhang (2010), Fan and Lv (2013) and Fan and Tang (2013). Despite considerable advances made in the theory and practice of penalised regression, there are still a number of open questions. These include the choice of the penalty function and tuning parameters. A number of contributions, notably by Fan and Li (2001) and Zhang (2010), have considered the use of nonconvex penalty functions with some success.[2]

Like penalised regressions, OCMT is valid when the underlying regression model is sparse. Further, it does not require the $\boldsymbol{x}_{nt}$ to have a sparse covariance matrix, and is applicable even if the covariance matrix of the noise variables, to be defined below, is not sparse. Of course, since OCMT is a model selection device, well known impossibility results for the uniform validity of post-selection estimators, such as those obtained in Leeb and Pötscher (2006) and Leeb and Pötscher (2008), apply. The main idea is to test the statistical significance of the net contribution of all n available potential covariates in explaining y_t individually, whilst taking full account of the multiple testing nature of the problem under consideration. All covariates

[1] A general discussion of high-dimensional data and their use in microeconomic analysis can be found in Belloni et al. (2014a).

[2] As an alternative to penalized regression, a number of procedures developed in the machine learning literature such as boosting, regression trees, and step-wise regressions are also widely used. See, for example, Friedman et al. (2000), Friedman (2001), Buhlmann (2006) and Fan and Lv (2008).

with statistically significant net contributions are then selected *jointly* to form an initial model specification for y_t. Unlike boosting and other greedy algorithms, our procedure is not sequential and selects in a single step all covariates whose t-ratios exceed a given threshold. A second stage will be needed only if there exist hidden signals, in the sense that there are covariates whose net contribution to y_t is zero, despite the fact that they belong to the true model for y_t. To allow for the possibility of hidden signals, we propose a multi-stage version, where OCMT is repeated by testing the statistical contribution of the remaining covariates, not selected in the first stage, again one at a time, to the unexplained part of y_t. We will show that this multi-stage process converges in a finite number of steps, since the number of hidden signals cannot rise with n. In a final step all statistically significant covariates, from all stages, are included as joint determinants of y_t in a multiple regression setting. Whilst the initial regressions of our procedure are common to boosting (see Buhlmann (2006)) and to the screening approach discussed in Fan and Lv (2008), Huang et al. (2008), Fan et al. (2009) and Fan and Song (2010), OCMT provides an inferentially motivated stopping rule without resorting to the use of information criteria, or penalised regression after the initial stage.

Related sequential model selection approaches have been proposed, among others, by Fithian et al. (2014), Tibshirani et al. (2014) and Fithian et al. (2015). In the context of linear regression, these methods build regression models by selecting variables from active sets, based on a sequence of tests. The use of multiple testing, implies that the choice of critical values, used at every testing step in the sequence, is crucial and there have been a number of important contributions, in this respect, including Li and Barber (2015) and G'Sell et al. (2016).

We provide theoretical results for the proposed OCMT procedure under relatively mild assumptions. In particular, we do not assume either a fixed design or time series independence for $\boldsymbol{x}_{nt}$ but consider a martingale difference condition for the cross-products $x_{it}x_{jt}$ and $\boldsymbol{x}_{nt}u_t$, where u_t is the error term of the true model. While these martingale difference conditions are our maintained assumption, we also provide theoretical arguments that allow the covariates to follow mixing processes. We establish theoretical results on the true positive rate, the false positive rate, the false discovery rate, and the norms of the coefficient estimate as well as the regression error.

We investigate the small sample properties of the proposed estimator and compare its performance with a number of penalised regressions (including Lasso and Adaptive Lasso), and boosting techniques. We consider data generating processes with and without lagged values of y_t, and carry out a large number of experiments. Although no method uniformly dominates, the results clearly show that OCMT does well across a number of dimensions. In particular, OCMT is very successful at eliminating noise variables, whereas it is still quite powerful at picking up the signals. It is outperformed by Lasso and Adaptive Lasso for a small fraction of experiments only. The relative performance of OCMT is also illustrated in an empirical application to forecasting U.S. output growth and inflation.

The paper is structured as follows: Section 2 explains the basic idea behind the OCMT method and introduces the concepts of the true and approximating models. Section 3 provides a formal description of the OCMT method and derives its asymptotic properties. Sections 4 presents a number of extensions. Section 5 gives the details of the Monte Carlo experiments and the summary of the simulation results. Section 6 presents the empirical application, and Section 7 concludes. Online supplement, organized in three parts, provide additional theoretical results and proofs, a complete set of Monte Carlo results for all the experiments conducted, and additional empirical findings.

Notations: Generic positive finite constants are denoted by C_i for $i = 0, 1, 2, \ldots$. They can take different values at different instances. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequences of positive real numbers, then $f_n = O(g_n)$, if there exists a positive finite constant C_0 such that $|f_n|/g_n \leq C_0$ for all n. $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $f_n = \ominus(g_n)$ if there exists $N_0 \geq 1$ and positive finite constants C_0 and C_1, such that $\inf_{n \geq N_0}(f_n/g_n) \geq C_0$, and $\sup_{n \geq N_0}(f_n/g_n) \leq C_1$. $\to_p$ denotes convergence in probability as $n, T \to \infty$.

2 True and Approximating Models and OCMT

Consider the data generating process (DGP),

$$y_t = \mathbf{a}'\boldsymbol{z}_t + \textstyle\sum_{i=1}^{k} \beta_i x_{it} + u_t, \tag{1}$$

where $\boldsymbol{z}_t$ is a known vector of pre-selected variables, $x_{1t}, x_{2t}, \ldots, x_{kt}$ are the k unknown *true* or *signal* variables, $0 < |\beta_i| \leq C < \infty$, for $i = 1, 2, \ldots, k$, and u_t is an error term. It is assumed that $\boldsymbol{z}_t$ and x_{it}, $i = 1, 2, \ldots, k$, are uncorrelated with u_t at time t. $\boldsymbol{z}_t$ may include deterministic terms such as a constant, linear trend and dummy variables, and/or stochastic variables, possibly including common factors and lagged values of y_t, that are considered crucial for the modelling of y_t, and are selected based possibly on *a priori* theoretical grounds.

Further suppose that the k signals are contained in a set $\mathcal{S}_{nt} = \{x_{it}, i = 1, 2, \ldots, n\}$, with n being potentially larger than T, which we refer to as the *active* set.[3] In addition to the k signals, the active set is comprised of *noise* variables that have *zero* correlations with the signals once the effects of $\boldsymbol{z}_t$ are filtered out, and a remaining set of variables that, net of $\boldsymbol{z}_t$, are correlated with the signals. We refer to the latter as *pseudo-signals* or *proxy* variables, since they can be falsely viewed as signals.

[3] We assume that the signal variables are contained in the active set. Nevertheless, OCMT can be applied even if the active set does not contain all of the signal variables. It is clear that in such a setting the true model or a model that contains the true model cannot be identified. However, OCMT will still weed out the variables that are uncorrelated with the signals. In support of this, we provide Monte Carlo evidence in Section 5 of the online MC supplement, based on a Monte Carlo experiment suggested to us by a referee.

The OCMT procedure considers the least squares (LS) regression of y_t on $\boldsymbol{z}_t$ and the regressors in the active set *one at the time*. Let t_i be the t-ratio of x_{it} in the regression of y_t on $\boldsymbol{z}_t$ and x_{it}, for $i = 1, 2, ..., n$,

$$t_i = \frac{T^{-1/2}\boldsymbol{x}_i'\mathbf{M}_z\boldsymbol{y}}{\hat{\sigma}_i\sqrt{T^{-1}\boldsymbol{x}_i'\mathbf{M}_z\boldsymbol{x}_i}} = \frac{T^{-1/2}\boldsymbol{x}_i'\mathbf{M}_z\boldsymbol{\mu}}{\hat{\sigma}_i\sqrt{T^{-1}\boldsymbol{x}_i'\mathbf{M}_z\boldsymbol{x}_i}} + \frac{T^{-1/2}\boldsymbol{x}_i'\mathbf{M}_z\boldsymbol{u}}{\hat{\sigma}_i\sqrt{T^{-1}\boldsymbol{x}_i'\mathbf{M}_z\boldsymbol{x}_i}} = t_{i,\mu} + t_{i,u}, \tag{2}$$

where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$ and $\boldsymbol{y} = (y_1, y_2, ..., y_T)'$ are $T \times 1$ vectors of observations on x_{it} and y_t, respectively, $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_T)'$, $\mu_t = \sum_{i=1}^{k}\beta_i x_{it}$, $\boldsymbol{u} = (u_1, u_2, ..., u_T)'$, $\mathbf{M}_z = \boldsymbol{I}_T - \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'$, $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_T)'$ is the matrix of observations on $\boldsymbol{z}_t$, and $\hat{\sigma}_i$ is the standard error of the regression of y_t on $\boldsymbol{z}_t$ and x_{it}.

Consider first $t_{i,u}$, defined by (2), which plays a key role in the workings of the OCMT. As $n, T \to \infty$, we rely on $t_{i,u}$ to remain bounded in probability sufficiently sharply so as to allow for multiple testing over very large values of n. We obtain such bounds under a variety of relatively mild assumptions on u_t and x_{it}. For example, we allow u_t to be a martingale difference process and require x_{it} to be uncorrelated with u_t. We do not require x_{it} to be strictly exogenous.

Regarding $t_{i,\mu}$ in (2), we distinguish between the cases where $t_{i,\mu}$ is bounded in probability sufficiently sharply as $n, T \to \infty$ and when it is not. The latter case is of special interest and suggests that x_{it} has power in explaining y_t, net of the pre-selected variables, $\boldsymbol{z}_t$. In such a case, we select x_{it}, and we distinguish between the signal variables, that are contained in μ_t, and pseudo-signal variables, which are not in μ_t but are nevertheless correlated with it. We show that OCMT identifies all such covariates with probability approaching one.

In the former case where $t_{i,\mu}$ is bounded in probability sufficiently sharply as $n, T \to \infty$, we characterise x_{it} as a noise covariate if it is not contained in μ_t, and a hidden signal if it is contained in μ_t. We show that all hidden signals will be selected by the application of one or more additional stages of OCMT.

It is clear from the above exposition that our variable selection approach focusses on the net impact of x_{it} on y_t conditional on the vector of pre-selected variables $\boldsymbol{z}_t$, rather than the marginal effects defined by β_i. The conditional *net impact* coefficient of x_{it} on y_t generalizes the mean net impact coefficient considered by Pesaran and Smith (2014), and it is given by

$$\theta_{i,T}\left(\boldsymbol{z}\right) = \sum_{j=1}^{k}\beta_j \sigma_{ij,T}\left(\boldsymbol{z}\right), \tag{3}$$

where $\sigma_{ij,T}(\boldsymbol{z}) = E\left(T^{-1}\boldsymbol{x}_i'\boldsymbol{M}_z\boldsymbol{x}_j\right)$. To simplify the exposition, we suppress the T subscript and use $\theta_i\left(\boldsymbol{z}\right)$ and $\sigma_{ij}\left(\boldsymbol{z}\right)$ below.

$\theta_i(\boldsymbol{z})$ plays a crucial role in our proposed approach, as it determines whether $t_{i,\mu}$ in (2) is bounded in probability sufficiently sharply as $n, T \to \infty$. Ideally, we would like to be able to base our selection decision directly on β_i and its estimate. But when n is large such a strategy is not feasible. Instead, we propose to base variable selection on $\theta_i(\boldsymbol{z})$. It is important to stress that knowing $\theta_i(\boldsymbol{z})$ does not imply we can determine β_i. Due to the correlation between variables, nonzero $\theta_i(\boldsymbol{z})$ does not necessarily imply nonzero β_i and we have the following four possibilities:

	$\theta_i(\boldsymbol{z}) \neq 0$	$\theta_i(\boldsymbol{z}) = 0$
$\beta_i \neq 0$	(I) Signals with nonzero net effect	(II) Hidden signals
$\beta_i = 0$	(III) Pseudo-signals	(IV) Noise variables

.

The first and the last case, where $\theta_i(\boldsymbol{z}) \neq 0$ if and only if $\beta_i \neq 0$, is the most straightforward case to be considered. But there is also a possibility of case II where $\theta_i(\boldsymbol{z}) = 0$ and $\beta_i \neq 0$ and case III where $\theta_i(\boldsymbol{z}) \neq 0$ and $\beta_i = 0$. These cases will also be considered in our analysis. Case II is likely to be rare in practice since it requires an *exact* equality between the coefficients of the true model, namely $\beta_i = -\sum_{j=1,j\neq i}^{k} \beta_j \sigma_{ii}^{-1}(\boldsymbol{z})\, \sigma_{ij}(\boldsymbol{z})$. However, the presence of pseudo-signals (case III) is quite likely, and will be an important consideration in our model selection strategy.

We shall refer to the model that contains only the signals as the *true model*, and to the model that contains the signals as well as one or more of the pseudo-signals, but none of the noise variables, as an *approximating model*. We assume that there are k^* pseudo-signal variables ordered to follow the k signal variables, so that the first $k + k^*$ variables in $\mathcal{S}_{nt}$ are signals and pseudo-signals, although this is not known to the investigator. The remaining $n - k - k^*$ variables are the noise variables. We assume that k is an unknown fixed constant, but allow k^* to rise with n such that $k^*/n \to 0$, and $k^*/T \to 0$, at a sufficiently slow rate. Specifically, we allow $k^* = \ominus(n^{\epsilon})$ for some appropriately bounded $\epsilon \geq 0$. We expect ϵ to be small when the correlation between the signals and the remaining covariates is sparse.

Our secondary maintained assumptions are somewhat more general and, accordingly, lead to fewer and weaker results. A first specification assumes that there exists an ordering (possibly unknown) such that

$$\theta_i(\boldsymbol{z}) = C_i \varrho^i, \text{ for } i = 1, 2, ..., n, \text{ and } |\varrho| < 1, \tag{4}$$

for a given set of constants, C_i. A second specification modifies the decay rate and assumes that

$$\theta_i(\boldsymbol{z}) = C_i i^{-\gamma}, \text{for } i = 1, 2, ..., n, \text{ and for some } \gamma > 0. \tag{5}$$

In both specifications $\max_{1 \leq i \leq n} |C_i| < C < \infty$. These specifications allow for various rates of decay in the way covariates are correlated with the signals. These cases are of technical interest and cover the autoregressive type designs considered in the literature in order to model the correlations across the covariates. See, for example, Zhang (2010) and Belloni et al. (2014b).

3 The Multiple Testing Approach

OCMT is inspired by the multiple testing literature, although the focus of OCMT is on controlling the probability of selecting an approximating model and the false discovery rate, rather than controlling the size of the union of the multiple tests that are being carried out. To simplify the exposition below, we assume that the vector of pre-selected variables, $\boldsymbol{z}_t$, contains only an intercept, in which case, the DGP (1) simplifies to

$$y_t = a + \sum_{i=1}^{k} \beta_i x_{it} + u_t, \text{ for } t = 1, 2, ..., T. \tag{6}$$

In matrix notation, we have

$$\boldsymbol{y} = a\boldsymbol{\tau}_T + \boldsymbol{X}_k \boldsymbol{\beta}_k + \boldsymbol{u}, \tag{7}$$

where $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of ones, $\boldsymbol{X}_k = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$ is the $T \times k$ matrix of observations on signal variables, $\boldsymbol{\beta}_k = (\beta_1, \beta_2, ..., \beta_k)'$ is the $k \times 1$ vector of associated slope coefficients and $\boldsymbol{u} = (u_1, u_2, ..., u_T)'$ is $T \times 1$ vector of errors. In addition, the conditional net impact coefficient $\theta_i(\boldsymbol{z})$ simplifies, for $\boldsymbol{z}_t = 1$, to

$$\theta_i = \sum_{j=1}^{k} \beta_j \sigma_{ij}, \tag{8}$$

where (we again suppress the subscript T), $\sigma_{ij} = E\left(T^{-1}\boldsymbol{x}_i'\boldsymbol{M}_\tau \boldsymbol{x}_j\right)$, and $\boldsymbol{M}_\tau = \boldsymbol{I}_T - \boldsymbol{\tau}_T\boldsymbol{\tau}_T'/T$. We consider the following assumptions:

Assumption 1 *Let $\boldsymbol{X}_{k,k^*} = (\boldsymbol{X}_k, \boldsymbol{X}_{k^*}^*)$, where $\boldsymbol{X}_k = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$, and $\boldsymbol{X}_{k^*}^* = (\boldsymbol{x}_{k+1}, \boldsymbol{x}_{k+2}, ..., \boldsymbol{x}_{k+k^*})$ are $T \times k$ and $T \times k^*$ observation matrices on signals and pseudo-signals, and suppose that there exists T_0 such that for all $T > T_0$, $\left(T^{-1}\boldsymbol{X}_{k,k^*}'\boldsymbol{X}_{k,k^*}\right)^{-1}$ is nonsingular with its smallest eigenvalue uniformly bounded away from 0, and $\boldsymbol{\Sigma}_{k,k^*} = E\left(T^{-1}\boldsymbol{X}_{k,k^*}'\boldsymbol{X}_{k,k^*}\right)$ is nonsingular for all T.*

Assumption 2 *The error term, u_t, in DGP (6) is a martingale difference process with respect to $\mathcal{F}_{t-1}^u = \sigma(u_{t-1}, u_{t-2}, ...,)$, with a zero mean and a constant variance, $0 < \sigma^2 < C < \infty$.*

Assumption 3 *Let $\mathcal{F}_{it}^x = \sigma(x_{it}, x_{i,t-1}, ...)$, where x_{it}, for $i = 1, 2, ..., n$, is the i-th covariate in the active set $\mathcal{S}_{nt}$. Define $\mathcal{F}_t^{xn} = \cup_{j=k+k^*+1}^{n} \mathcal{F}_{jt}^x$, $\mathcal{F}_t^{xo} = \cup_{i=1}^{k+k^*} \mathcal{F}_{jt}^x$, and $\mathcal{F}_t^x = \mathcal{F}_t^{xn} \cup \mathcal{F}_t^{xo}$. Then, x_{it} is independent of $x_{jt'}$ for $i = 1, 2, ..., k + k^*$, $j = k + k^* + 1, k + k^* + 2, ..., n$, and for all t and t', and $E\left[x_{it}x_{jt} - E(x_{it}x_{jt}) \left|\mathcal{F}_{t-1}^x\right.\right] = 0$, for $i, j = 1, 2, ..., n$, and all t. Finally, $E\left(x_{it}u_t \left|\mathcal{F}_{t-1}\right.\right) = 0$, for $i = 1, 2, ..., n$, and all t, where $\mathcal{F}_t = \mathcal{F}_t^x \cup \mathcal{F}_t^u$.*

Assumption 4 *There exist sufficiently large positive constants C_0, C_1, C_2 and C_3 and $s_x, s_u > 0$ such that the covariates in the active set $\mathcal{S}_{nt}$ satisfy*

$$\sup_{i,t} \Pr(|x_{it}| > \alpha) \leq C_0 \exp\left(-C_1 \alpha^{s_x}\right), \textit{ for all } \alpha > 0, \tag{9}$$

and the errors, u_t, in DGP (6) satisfy

$$\sup_t \Pr(|u_t| > \alpha) \leq C_2 \exp\left(-C_3 \alpha^{s_u}\right), \textit{ for all } \alpha > 0. \tag{10}$$

Assumption 5 *Consider x_t and the $l_T \times 1$ vector of covariates $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$. $\boldsymbol{q}_{\cdot t}$ can contain a constant term, and x_t is a generic element of $\mathcal{S}_{nt}$ that does not belong to $\boldsymbol{q}_{\cdot t}$. It is assumed that $E(\boldsymbol{q}_{\cdot t} x_t)$ and $\boldsymbol{\Sigma}_{qq} = E(\boldsymbol{q}_{\cdot t}\boldsymbol{q}'_{\cdot t})$ exist and $\boldsymbol{\Sigma}_{qq}$ is invertible. Define $\boldsymbol{\gamma}_{qx,T} = \boldsymbol{\Sigma}_{qq}^{-1}[T^{-1}\sum_{t=1}^{T} E(\boldsymbol{q}_{\cdot t} x_t)]$ and*

$$u_{x,t,T} =: u_{x,t} = x_t - \boldsymbol{\gamma}'_{qx,T}\boldsymbol{q}_{\cdot t}. \tag{11}$$

All elements of the vector of projection coefficients, $\boldsymbol{\gamma}_{qx,T}$, are uniformly bounded and only a finite number of the elements of $\boldsymbol{\gamma}_{qx,T}$ are different from zero.

Assumption 6 *The number of signals, k, in (6) is finite, and their slope coefficients could change with T, such that for $i = 1, 2, ..., k$, $\beta_{i,T} = \ominus\left(T^{-\vartheta}\right)$, for some $0 \leq \vartheta < 1/2$.*

Before formally outlining OCMT procedure and presenting our theoretical results, we provide some remarks on the pros and cons of our assumptions as compared to the ones typically assumed in the penalised regression and boosting literature.

Assumption 1 ensures that regression coefficients in the model containing all signals and pseudo-signals and none of the noise variables are identified. Assumption 2 is slightly more general than the usual assumption in the regression analysis. Assumption 3 allows x_{it} to be a martingale difference sequence which is somewhat weaker than the IID assumption typically made in the literature on penalised regression. Relaxation of this assumption to allow for serially correlated covariates is discussed in Section 4.2.

The exponential bounds in Assumption 4 are sufficient for the existence of all moments of the covariates, x_{it}, and the error term, u_t. It is very common in the literature to assume some form of exponentially declining bound for probability tails of u_t and x_{it}. See, for example, Zheng et al. (2014).

Assumption 5 is a technical condition that is required for some results derived in the Appendix and in the online theory supplement, which consider a more general multiple regression context where subsets of regressors in $\boldsymbol{x}_{nt}$ are included in the regression equation. In the simple case where $\boldsymbol{q}_{\cdot t} = 1$, then Assumption 5 is trivially satisfied and follows from the rest of the assumptions, and we have $\boldsymbol{\gamma}_{qx,T} = \mu_{x,T} = \frac{1}{T}\sum_{t=1}^{T} E(x_t)$, and $u_{x,t,T} = x_t - \mu_{x,T}$.

Assumption 6 allows for the possibility of weak signal variables whose coefficients, $\beta_{i,T}$, for $i = 1, 2, ..., k$, decline with the sample size, T, at a sufficiently slow rate. To simplify notation, subscript T is dropped subsequently, and it is understood that the slope and net effect coefficients can change with the sample size according to this assumption. Using θ_i, we can refine our concept of pseudo-signals as variables with $\theta_i = \ominus\left(T^{-\vartheta}\right)$ for $i = k+1, k+2, ..., k+k^*$, for some $0 \leq \vartheta < 1/2$. Remark 1 discusses further how this condition enters the theoretical results.

Regarding our assumptions on the correlation between variables in the active set we note the following. The signal and noise variables are allowed to be correlated amongst themselves, so no restrictions are imposed on σ_{ij} for $i, j = 1, 2, ..., k$, and on σ_{ij} for $i, j = k+k^*+1, k+k^*+2, ..., n$.

Also, signals and pseudo-signals are allowed to be correlated; namely, σ_{ij} could be non-zero for $i, j = 1, 2, ..., k+k^*$. Therefore, signals and pseudo-signals as well as noise variables can contain common factors, but, under our definition of noise variables, the factors cannot be shared between the signals/pseudo-signals and noise variables, since the latter are uncorrelated with the former. If there are common factors affecting signal variables as well as a large number of the remaining variables in the active set, one can and should condition on such factors, as we do in our empirical illustration.[4] Without such conditioning, the size of the approximating model would be too large to be of practical use, when common factors affect both signal and a large number of the remaining variables in the active set.

In contrast, a number of crucial issues arise in the context of Lasso, or more generally when L_q penalty functions with $0 \leq q \leq 1$ are used. Firstly, it is customary to assume a framework of fixed-design regressor matrices, where in many cases a generalisation to stochastic regressors is not straightforward, requiring conditions such as the spark condition of Donoho and Elad (2003) and Zheng et al. (2014). Secondly, a frequent condition for Lasso to be a valid variable selection method is the irrepresentable condition which bounds the maximum of all regression coefficients, in regression of any noise or pseudo-signal variable on the signals, to be less than one in the case of normalised regressor variables. See, for example, Section 7.5 of Buhlmann and van de Geer (2011).

Further, most results for penalised regression essentially take as given the knowledge of the tuning parameter associated with the penalty function. In practice, cross-validation is used to determine this parameter but theoretical results on the properties of such cross-validation schemes are rare. Available theoretical results on boosting, as presented in Buhlmann (2006), are also limited to the case of bounded and IID regressors, while few restrictions are placed on their correlation structure.

We proceed next with formally describing the OCMT procedure. It is a multi-stage procedure. In the first stage, we consider the n bivariate regressions of y_t on a constant ($\boldsymbol{z}_t$ in the general case) and x_{it}, for $i = 1, 2, ..., n$,

$$y_t = c_i + \phi_i x_{it} + u_{it},\ t = 1, 2, ..., T, \tag{12}$$

where $\phi_i = \theta_i / \sigma_{ii}$, θ_i is defined in (8) and σ_{ii} is defined below (8). Denoting the t-ratio of ϕ_i in this regression by $t_{\hat{\phi}_{i,(1)}}$, we have

$$t_{\hat{\phi}_{i,(1)}} = \frac{\hat{\phi}_i}{s.e.\left(\hat{\phi}_i\right)} = \frac{\boldsymbol{x}_i' \boldsymbol{M}_{\boldsymbol{\tau}} \boldsymbol{y}}{\hat{\sigma}_i \sqrt{\boldsymbol{x}_i' \boldsymbol{M}_{\boldsymbol{\tau}} \boldsymbol{x}_i}}, \tag{13}$$

[4] Note that our theory allows for conditioning on observed common factors by incorporating them in $\boldsymbol{z}_t$. But when factors are unobserved they need to be replaced by their estimates using, for example, principal components. A formal argument that the associated estimation error is asymptotically negligible involves additional technical complications, and requires deriving exponential inequalities for the quantities analysed in Theorem 1 of Bai and Ng (2002) and Lemma A1 of Bai and Ng (2006), and then assuming that $\sqrt{T}/n \to 0$ as $n, T \to \infty$. While such a derivation is clearly feasible under appropriate regularity conditions, a formal analysis is beyond the scope of the present paper.

where $\hat{\phi}_i = (\boldsymbol{x}_i'\boldsymbol{M}_{\boldsymbol{\tau}}\boldsymbol{x}_i)^{-1}\boldsymbol{x}_i'\boldsymbol{M}_{\boldsymbol{\tau}}\boldsymbol{y}$ denotes the LS estimator of ϕ_i, $\hat{\sigma}_i^2 = \boldsymbol{e}_i'\boldsymbol{e}_i/T$, and $\boldsymbol{e}_i$ denotes the $T \times 1$ vector of residual of the regression of $\boldsymbol{y}$ on $\boldsymbol{\tau}_T$ and $\boldsymbol{x}_i$. The first stage OCMT selection indicator is given by

$$\widehat{\mathcal{J}}_{i,(1)} = I[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)], \text{ for } i = 1, 2, ..., n, \tag{14}$$

where $c_p(n,\delta)$ is a *critical value function* defined by

$$c_p(n,\delta) = \Phi^{-1}\left(1 - \frac{p}{2f(n,\delta)}\right), \tag{15}$$

$\Phi^{-1}(.)$ is the inverse of standard normal distribution function, $f(n,\delta) = cn^{\delta}$ for some positive constants δ and c, and p $(0 < p < 1)$ is the nominal size of the individual tests to be set by the investigator. We will refer to δ as the *critical value exponent*. One value of δ is used in the first stage, while another one (denoted by δ^*) is used in subsequent stages of OCMT. As we shall see, it will be required that $\delta^* > \delta$. Variables with $\widehat{\mathcal{J}}_{i,(1)} = 1$ are selected as signals and pseudo-signals in the first stage. Denote the number of covariates selected in the first stage by $\hat{k}_{(1)}^o$, the index set of the selected variables by $\mathcal{S}_{(1)}^o$, and the $T \times \hat{k}_{(1)}^o$ observation matrix of the $\hat{k}_{(1)}^o$ selected variables by $\boldsymbol{X}_{(1)}^o$. Further, let $\boldsymbol{X}_{(1)} = (\boldsymbol{\tau}_T, \boldsymbol{X}_{(1)}^o) = (\boldsymbol{x}_{(1),1}, ..., \boldsymbol{x}_{(1),T})'$, $\hat{k}_{(1)} = \hat{k}_{(1)}^o$, $\mathcal{S}_{(1)} = \mathcal{S}_{(1)}^o$, and $\mathfrak{A}_{(2)} = \{1, 2, ..., n\} \setminus \mathcal{S}_{(1)}$. For future reference, we also set $\boldsymbol{X}_{(0)} = \boldsymbol{\tau}_T$ and $\mathfrak{A}_{(1)} = \{1, 2, ..., n\}$. In stages $j = 2, 3, ...,$ we consider the $n - \hat{k}_{(j-1)}$ regressions of y_t on the variables in $\boldsymbol{X}_{(j-1)}$ and, one at the time, x_{it} for i belonging in the active set, $\mathfrak{A}_{(j)}$. We then compute the following t-ratios

$$t_{\hat{\phi}_{i,(j)}} = \frac{\hat{\phi}_{i,(j)}}{s.e.\left(\hat{\phi}_{i,(j)}\right)} = \frac{\boldsymbol{x}_i'\boldsymbol{M}_{(j-1)}\boldsymbol{y}}{\hat{\sigma}_{i,(j)}\sqrt{\boldsymbol{x}_i'\boldsymbol{M}_{(j-1)}\boldsymbol{x}_i}}, \text{ for } i \in \mathfrak{A}_{(j)},\ j = 2, 3, ..., \tag{16}$$

where $\hat{\phi}_{i,(j)} = \left(\boldsymbol{x}_i'\boldsymbol{M}_{(j-1)}\boldsymbol{x}_i\right)^{-1}\boldsymbol{x}_i'\boldsymbol{M}_{(j-1)}\boldsymbol{y}$ is the LS estimator of the conditional net effect of x_{it} on y_t in stage j, $\hat{\sigma}_{i,(j)}^2 = T^{-1}\boldsymbol{e}_{i,(j)}'\boldsymbol{e}_{i,(j)}$, $\boldsymbol{M}_{(j-1)} = \boldsymbol{I}_T - \boldsymbol{X}_{(j-1)}(\mathbf{X}_{(j-1)}'\boldsymbol{X}_{(j-1)})^{-1}\boldsymbol{X}_{(j-1)}'$, and $\boldsymbol{e}_{i,(j)}$ denotes the residual vector of the regression of $\boldsymbol{y}$ on $\boldsymbol{X}_{i,(j-1)} = \left(\boldsymbol{x}_i, \boldsymbol{X}_{(j-1)}\right)$. Regressors for which $\widehat{\mathcal{J}}_{i,(j)} = 1$, are then added to the set of already selected covariates from the previous stages, where $\widehat{\mathcal{J}}_{i,(j)} = I[|t_{\hat{\phi}_{i,(j)}}| > c_p(n,\delta^*)]$. Denote the number of variables selected in stage j by $\hat{k}_{(j)}^o$, their index set by $\mathcal{S}_{(j)}^o$, and the $T \times \hat{k}_{(j)}^o$ matrix of the $\hat{k}_{(j)}^o$ selected variables in stage j by $\mathbf{X}_{(j)}^o$. Also let $\mathbf{X}_{(j)} = (\mathbf{X}_{(j-1)}, \mathbf{X}_{(j)}^o) = (\boldsymbol{x}_{(j),1}, \boldsymbol{x}_{(j),2}, ..., \boldsymbol{x}_{(j),T})'$, $\hat{k}_{(j)} = \hat{k}_{(j-1)} + \hat{k}_{(j)}^o$, $\mathcal{S}_{(j)} = \mathcal{S}_{(j-1)} \cup \mathcal{S}_{(j)}^o$, define the $(j+1)$ stage active set by $\mathfrak{A}_{(j+1)} = \{1, 2, ..., n\} \setminus \mathcal{S}_{(j)}$, and then proceed to the next stage by increasing j by one. Note that $\hat{k}_{(j)}$ is the *total* number of variables selected up to and including stage j, $\hat{\phi}_{i,(j)} \to_p \theta_{i,(j)}/\sigma_{ii,(j)}$, where $\theta_{i,(j)}$ and $\sigma_{ii,(j)}$ are used in the remainder of this paper to denote $\theta_i\left(\boldsymbol{x}_{(j-1)}\right)$ and $\sigma_{ii}\left(\boldsymbol{x}_{(j-1)}\right)$ introduced in (3). Also to simplify the notation, $\theta_{i,(1)}$ is shown as θ_i. The procedure stops when no regressors are selected at a given stage, say $\hat{\jmath}$, in which case the final number of selected variables will be given, as before, by $\hat{k} = \hat{k}_{(\hat{j}-1)}$. The multi-stage OCMT selection indicator is thus given by $\widehat{\mathcal{J}}_i = \sum_{j=1}^{\hat{P}} \widehat{\mathcal{J}}_{i,(j)}$, where $\hat{P}$ denotes the number of stages at completion of OCMT, formally defined as

$$\hat{P} = \min_j \{j : \textstyle\sum_{i=1}^{n} \widehat{\mathcal{J}}_{i,(j)} = 0\} - 1. \tag{17}$$

It is important to note that the number of stages needed for OCMT is bounded in n. To show this we note that not all signals can be hidden, and once we condition on the set of signals that are not hidden, then there must exist i such that $\theta_i(\boldsymbol{z}) \neq 0$, while $\theta_i = 0$ and $\beta_i \neq 0$, where here $\boldsymbol{z}$ denotes the signal variables that are not hidden.[5] Using this result one can successively uncover all hidden signals. We denote by P the number of stages that need to be considered to uncover all hidden signals. Its true population value is denoted by P_0. This is defined as the index of the last stage where OCMT finds further signals (or pseudo-signals), assuming that $\Pr[|t_{\hat{\phi}_{i,(j)}}| > c_p(n,\delta) \, |\theta_{i,(j)} \neq 0] = 1$ and $\Pr[|t_{\hat{\phi}_{i,(j)}}| > c_p(n,\delta) \, |\theta_{i,(j)} = 0] = 0$, for all variables indexed by i, and OCMT stages indexed by j. Of course, these probabilities do not take the values 1 and 0 respectively, in small samples, but we will handle this complication later on. The following proposition provides an upper bound to P_0.

Proposition 1 *Suppose that y_t, $t = 1, 2, ..., T$, are generated according to (6), with $\beta_i \neq 0$ for $i = 1, 2, ..., k$, and that Assumption 1 holds. Then, there exists j, $1 \leq j \leq k$, for which $\theta_{i,(j)} \neq 0$, and the population value of the number of stages required to select all the signals, denoted as P_0, satisfies $1 \leq P_0 \leq k$.*

A proof is provided in Subsection A.2.1 of the Appendix.

In practice, $\hat{P}$ is likely to be small since hidden signals arise only in rare cases where $\theta_i = 0$ whilst the associated β_i is non-zero. Also, as we show all signals with nonzero θ will be picked up with probability tending to one in the first stage. Stopping after the first stage tends to improve the small sample performance of the OCMT approach, investigated in Section 5, only marginally when no hidden signals are present. Thus, allowing $P > 1$, using the stopping rule defined above, does not significantly deteriorate the small sample performance of OCMT when hidden signals are not present, while it picks-up all hidden signals with probability tending to one. Finally, using (7), note that the conditional net effect coefficient of variable i at stage j of OCMT, $\theta_{i,(j)}$, can be written as

$$\theta_{i,(j)} = E\left(T^{-1}\boldsymbol{x}_i'\boldsymbol{M}_{(j-1)}\boldsymbol{y}\right) = E\left(T^{-1}\boldsymbol{x}_i'\boldsymbol{M}_{(j-1)}\boldsymbol{X}_k\boldsymbol{\beta}_k\right) = \textstyle\sum_{\ell=1}^{k}\beta_\ell\sigma_{i\ell}\left(\boldsymbol{x}_{(j-1)}\right), \tag{18}$$

and to allow for the possibility of weak signals as defined by Assumption 6, pseudo-signal variables can be more generally defined as covariates $i = k+1, k+2, ..., k+k^*$ with $\theta_{i,(j)} = \ominus\left(T^{-\vartheta}\right)$, for some $0 \leq \vartheta < 1/2$ and some $1 \leq j \leq P_0$.

[5]For a proof see Lemma A1 in the online supplement. Note also that $\boldsymbol{z}_t$ may contain lagged values of y_t, principal components or other estimates of common effects as well as covariates that the investigator believes must be included.

Once the OCMT procedure is completed, the OCMT estimator of β_i, denoted by $\tilde{\beta}_i$, is set as

$$\tilde{\beta}_i = \begin{cases} \hat{\beta}_i^{(\hat{k})}, & \text{if } \widehat{\mathcal{J}}_i = 1 \\ 0, & \text{otherwise} \end{cases}, \text{ for } i = 1, 2, ..., n, \tag{19}$$

where $\hat{\beta}_i^{(\hat{k})}$ is the LS estimator of the coefficient of the i^{th} variable in a regression of y_t on all the selected covariates, namely *all* the covariates for which $\widehat{\mathcal{J}}_i = 1$, plus a constant term ($\boldsymbol{z}_t$ in the general case).

The choice of the critical value function, $c_p(n, \delta)$, given by (15), is important since it allows the investigator to relate the size and power of the selection procedure to the inferential problem in classical statistics, with the modification that p (type I error) is now scaled by a function of the number of covariates under consideration. As we shall see, the OCMT procedure applies irrespective of whether n is small or large relative to T, so long as $T = \ominus(n^{\kappa_1})$, for any finite $\kappa_1 > 0$. This follows from result (i) of Lemma A2 in the online supplement, which establishes that $c_p^2(n, \delta) = O[\delta \ln(n)]$. It is also helpful to bear in mind that, using result (ii) of Lemma A2 in the online supplement, $\exp\left[-\varkappa c_p^2(n, \delta)/2\right] = \ominus\left(n^{-\delta\varkappa}\right)$, and $c_p(n, \delta) = o\left(T^{C_0}\right)$, for all $C_0 > 0$, assuming there exists $\kappa_1 > 0$, such that $T = \ominus(n^{\kappa_1})$.

Note that setting $\delta = 1$ in the first stage, is equivalent to using a Bonferroni correction for the multiple testing problem. Of course, other c_p values can be used, such as those proposed by Holm (1979), Benjamini and Hochberg (1995), or Gavrilov et al. (2009) which are designed to control the family-wise error rate associated with a set of tests. However, since most impose some restriction on the dependence structure between the multiple tests (with the exception of the original Bonferroni procedure and the one proposed by Holm (1979)), we choose to use (15) which, furthermore, has a bespoke design, in terms of the conditions placed on δ, and is appropriate for the multi-stage OCMT method, where the number of tests carried out is not predetermined in advance.

We now consider the relationship of OCMT to sequential model selection procedures advanced in the literature. A notable example is L_2-Boosting by Buhlmann (2006) which starts with the same set of bivariate regressions, (12), but in the first step selects *only* the covariate with the maximum fit, as measured by the sum of squared residuals (SSR). Additional covariates are added sequentially by regressing a quasi-residual from the first step on the remaining covariates. The process is continued till convergence decided based on some information criterion.[6] Other sequential model selection approaches, such as those by Fithian et al. (2014), Tibshirani et al. (2014) and Fithian et al. (2015) build regression models by selecting variables from active sets, based on a sequence of tests. Variables are selected, and added to the model, one by one and selection stops once a test does not reject the latest null hypothesis in the sequence. It is important to note that these methods select one covariate (or at most a block

[6] The quasi-residuals are computed as $y_t - v\,\hat{y}_t$, where $\hat{y}_t$ is the fitted value in terms of the selected covariate, and v is a constant tuning parameter referred to as the step size. Buhlmann (2006) recommends choosing $v < 1$.

of covariates) in each of the steps. In contrast, OCMT operates as a 'hub and spoke' approach. It selects, in a single step, all variables whose t-ratios, in (12), exceed a threshold (given by $c_p(n, \delta)$), in absolute value. As a result, it is clear that in its main implementation OCMT is not a sequential approach. Only in the presence of hidden signals, does OCMT require subsequent stages. Even then, under our setting, where k is finite, the number of stages cannot exceed k with a high probability, and as a result in the vast majority of cases the number of additional stages required will be rather small.

We investigate the asymptotic properties of the OCMT procedure and the associated OCMT estimators, $\tilde{\beta}_i$, for $i = 1, 2, ..., n$, in terms of the probability of selecting the approximating model, and in terms of support recovery type statistics used in the Lasso literature, namely the true and false positive rates (TRP and FPR, respectively) defined by

$$TPR_{n,T} = \frac{\sum_{i=1}^{n} I(\widehat{\mathcal{J}}_i = 1 \text{ and } \beta_i \neq 0)}{\sum_{i=1}^{n} I(\beta_i \neq 0)}, \text{ and } FPR_{n,T} = \frac{\sum_{i=1}^{n} I(\widehat{\mathcal{J}}_i = 1, \text{ and } \beta_i = 0)}{\sum_{i=1}^{n} I(\beta_i = 0)}. \quad (20)$$

We also examine the following false discovery rate

$$FDR_{n,T} = \frac{\sum_{i=1}^{n} I(\widehat{\mathcal{J}}_i = 1, \text{ and } \beta_i = \theta_i = 0)}{\sum_{i=1}^{n} \widehat{\mathcal{J}}_i + 1}, \quad (21)$$

which applies to selection of signals and pseudo-signals. Further, we consider the error and the coefficient norms of the selected model, defined by

$$F_{\tilde{u}} = T^{-1}||\tilde{\boldsymbol{u}}||^2 = T^{-1}\textstyle\sum_{t=1}^{T} \tilde{u}_t^2, \text{ and } F_{\tilde{\boldsymbol{\beta}}} = ||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n|| = [\textstyle\sum_{i=1}^{n} (\tilde{\beta}_i - \beta_i)^2]^{1/2}, \quad (22)$$

respectively, where $\tilde{\boldsymbol{u}} = (\tilde{u}_1, \tilde{u}_2, ..., \tilde{u}_T)'$, $\tilde{u}_t = y_t - \hat{a} - \tilde{\boldsymbol{\beta}}'_n \boldsymbol{x}_{nt}$, $\boldsymbol{\beta}_n = (\beta_1, \beta_2, ..., \beta_n)'$, $\tilde{\boldsymbol{\beta}}_n = (\tilde{\beta}_1, \tilde{\beta}_2, ..., \tilde{\beta}_n)'$, $\tilde{\beta}_i$, for $i = 1, 2, ..., n$ are defined by (19), and $\hat{a}$ is the estimator of the constant term in the final regression.

We now present the main theoretical results using lemmas established in the online supplement. The key is Lemma A10 in the online supplement, which provides sharp bounds on the probability of $|t_{\hat{\phi}_{i,(j)}}| > c_p(n, \delta)$ conditional on whether the net effect coefficient $\theta_{i,(j)}$ is zero or not. Here we provide a simpler version of this lemma which focuses on the first-stage regressions and should provide a better understanding of the main mathematical results that lie behind the proofs in the more complicated multi-stage version of the OCMT.

Proposition 2 *Suppose y_t is given by (6) and Assumptions 2-4 hold. Let x_t be a generic element of the active set $\mathcal{S}_{nt}$, and suppose Assumption 5 holds for x_t and $\boldsymbol{q}_{\cdot t} = 1$. Consider the t-ratio of x_t in the regression of y_t on an intercept and x_t:*

$$t_x = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_\tau \boldsymbol{y}}{\sqrt{(T^{-1}\boldsymbol{e}'\boldsymbol{e})(T^{-1}\boldsymbol{x}'\mathbf{M}_\tau \boldsymbol{x})}},$$

where $\boldsymbol{e}$ is the $T \times 1$ vector of regression residuals. Let $\theta = E(T^{-1}\boldsymbol{x}'\mathbf{M}_\tau \boldsymbol{y})$ be the net impact effect of x_t, and suppose there exists $\kappa_1 > 0$ such that $T = \ominus(n^{\kappa_1})$. Then, for some finite *positive constants C_0 and C_1, we have*

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] \leq \exp\left[-\chi c_p^2(n,\delta)/2\right] + \exp\left(-C_0 T^{C_1}\right), \tag{23}$$

where $c_p(n,\delta)$ is the critical value function given by (15), and $\chi = [(1-\pi)/(1+d_T)]^2$, for any π in the range $0 < \pi < 1$, any $d_T > 0$ and bounded in T. Suppose further that in the case where $\theta \neq 0$, we have $\theta = \ominus\left(T^{-\vartheta}\right)$, for some $0 \leq \vartheta < 1/2$, where $c_p(n,\delta) = O\left(T^{1/2-\vartheta-C_4}\right)$, for some positive constant C_4. Then,

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta \neq 0\right] > 1 - \exp\left(-C_2 T^{C_3}\right). \tag{24}$$

Result (23) establishes a sharp probability bound for the absolute value of the t-ratio of x with zero net impact effect. The first term on the right side of (23) asymptotically dominates, and using result (ii) of Lemma A2 in the online supplement we have $\exp\left[-\chi c_p^2(n,\delta)/2\right] = \ominus\left(n^{-\delta\chi}\right)$. Result (24), on the other hand, establishes a lower bound on the probability of the event $|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)$ conditional on θ being sufficiently away from zero.

Since we wish to allow for the possibility of hidden signals for which $\theta = 0$ even if the associated $\beta \neq 0$, the results in Lemma A10 in the online supplement are obtained for t-ratios in multiple regression contexts where subsets of regressors in the active set are also included in the regression equation for y_t. Nevertheless, it is instructive to initially consider the OCMT in the absence of such hidden signals. Theorems 1 and 2 below provide the results for the general case where hidden signals are allowed.

We first examine $TPR_{n,T}$ defined by (20), under the assumption that $\theta_i \neq 0$ if $\beta_i \neq 0$. Note that by definition $TPR_{n,T} = k^{-1}\sum_{i=1}^{k} I(\widehat{\mathcal{J}}_{i,(1)} = 1 \text{ and } \beta_i \neq 0)$. Since the elements of this summation are 0 or 1, then taking expectations we have (note that in the present simple case $\theta_i \neq 0$ implies $\beta_i \neq 0$)

$$TPR_{n,T} = k^{-1}\textstyle\sum_{i=1}^{k} E[I(\widehat{\mathcal{J}}_{i,(1)} = 1 \text{ and } \beta_i \neq 0)] = k^{-1}\textstyle\sum_{i=1}^{k}\Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\,|\theta_i \neq 0].$$

Now using result (24) of Proposition 2, and recalling that $T = \ominus\left(n^{\kappa_1}\right)$, we have

$$TPR_{n,T} \geq 1 - \exp\left(-C_2 T^{C_3}\right) = 1 + O\left[\exp\left(-C_2 n^{C_3\kappa_1}\right)\right], \tag{25}$$

for some $C_2, C_3 > 0$. Hence, $TPR_{n,T} \to_p 1$ for any $\kappa_1 > 0$.

Consider now $FPR_{n,T}$ defined by (20). Again, note that the elements of $FPR_{n,T}$ are either 0 or 1 and hence $|FPR_{n,T}| = FPR_{n,T}$. Taking expectations of the right part of (20), and assuming $\theta_i = \ominus\left(T^{-\vartheta}\right)$, for $i = k+1, k+2, ..., k+k$, and some $0 \leq \vartheta < 1/2$, we have $(n-k)^{-1}\sum_{i=k+1}^{n}\Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\,|\beta_i = 0] = (n-k)^{-1}\sum_{i=k+1}^{k+k^*}\Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\,|\theta_i \neq 0] + (n-k)^{-1}\sum_{i=k+k^*+1}^{n}\Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\,|\theta_i] = 0$. Using (24) of Proposition 2 and assuming there exists $\kappa_1 > 0$ such that $T = \ominus\left(n^{\kappa_1}\right)$, we have $k^* - \sum_{i=k+1}^{k+k^*}\Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\,|\theta_i \neq 0] = O\left[\exp\left(-C_2 T^{C_3}\right)\right]$, for some finite positive constants C_2 and C_3. Moreover, (23) of Proposition

2, which holds uniformly over i, given the uniformity of (9) and (10) of Assumption 4, implies that for any $0 < \varkappa < 1$ there exist finite positive constants C_0 and C_1 such that

$$\sum_{i=k+k^*+1}^{n} \Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\,|\theta_i = 0] \leq \sum_{i=k+k^*+1}^{n} \left\{\exp\left[-\varkappa c_p^2(n,\delta)/2\right] + \exp\left(-C_0 T^{C_1}\right)\right\}. \tag{26}$$

Using these results we obtain

$$(n-k)^{-1}\sum_{i=k+1}^{n} \Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta)\,|\beta_i = 0] = k^*/(n-k) + O\left\{\exp\left[-\varkappa c_p^2(n,\delta)/2\right]\right\} + O\left[\exp(-C_0 T^{C_1})\right] + O\left[\exp\left(-C_2 T^{C_3}\right)/(n-k)\right]. \tag{27}$$

Next, we consider the probability of choosing the approximating model. A selected regression model is referred to as an approximating model if it contains the signal variables x_{it}, $i = 1, 2, ..., k$, and none of the noise variables, x_{it}, $i = k+k^*+1, k+k^*+2, ..., n$. The models in the set may contain one or more of the pseudo-signals, x_{it}, $i = k+1, k+2, ..., k+k^*$. We refer to all such regressions as the set of approximating models. So, the event of choosing the approximating model is given by

$$\mathcal{A}_0 = \{\textstyle\sum_{i=1}^{k} \widehat{\mathcal{J}}_i = k\} \cap \{\textstyle\sum_{i=k+k^*+1}^{n} \widehat{\mathcal{J}}_i = 0\}. \tag{28}$$

Theorem 1 below states the conditions under which $\Pr(\mathcal{A}_0) \to 1$. The results for the general multi-stage case that allows for the possibility of hidden signals are given in the following theorem. Since it is assumed that the expansion rates of T and n are related, the results that follow are reported in terms of n for presentational ease and consistency. They could, of course, be reported equally in terms of T, if required.

Theorem 1 *Consider the DGP (6) with k signals, k^* pseudo-signals, and $n-k-k^*$ noise variables, and suppose that Assumptions 1-4 and 6 hold, Assumption 5 holds for x_{it} and $\boldsymbol{q}_{\cdot t} = \boldsymbol{x}_{(j-1),t}$, $i \in \mathfrak{A}_{(j)}$, $j = 1, 2, ...k$, where $\mathfrak{A}_{(j)}$ is the active set at stage j of the OCMT procedure. $c_p(n,\delta)$ is given by (15) with $0 < p < 1$ and let $f(n,\delta) = cn^{\delta}$, for the first stage of OCMT, and $f(n,\delta^*) = cn^{\delta^*}$ for subsequent stages, for some $c > 0$, $\delta^* > \delta > 0$. $n, T \to \infty$, such that $T = \ominus(n^{\kappa_1})$, for some $\kappa_1 > 0$, and $k^* = \ominus(n^{\epsilon})$ for some positive $\epsilon < \min\{1, \kappa_1/3\}$. Then, for any $0 < \varkappa < 1$, and for some constant $C_0 > 0$,*

(a) the probability that the number of stages in the OCMT procedure, $\hat{P}$, defined by (17), exceeds k is given by

$$\Pr\left(\hat{P} > k\right) = O\left(n^{1-\varkappa\delta^*}\right) + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left[\exp\left(-n^{C_0\kappa_1}\right)\right], \tag{29}$$

(b) the probability of selecting the approximating model, $\mathcal{A}_0$, defined by (28), is given by

$$\Pr(\mathcal{A}_0) = 1 + O\left(n^{1-\delta\varkappa}\right) + O\left(n^{2-\delta^*\varkappa}\right) + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left[\exp\left(-n^{C_0\kappa_1}\right)\right], \tag{30}$$

(c) for the True Positive Rate, $TPR_{n,T}$, defined by (20), we have

$$E\left|TPR_{n,T}\right| = 1 + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left[\exp\left(-n^{C_0\kappa_1}\right)\right], \tag{31}$$

and if $\delta > 1 - \kappa_1/3$, then $TPR_{n,T} \to_p 1$; for the False Positive Rate, $FPR_{n,T}$, defined by (20), we have

$$E\left|FPR_{n,T}\right| = \frac{k^*}{n-k} + O\left(n^{-\varkappa\delta}\right) + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left(n^{\epsilon-1}\right) + O\left[\exp\left(-n^{C_0\kappa_1}\right)\right], \tag{32}$$

and if $\delta > \min\{0, 1-\kappa_1/3\}$, and $\delta^* > 1$, then $FPR_{n,T} \to_p 0$. For the False Discovery Rate, $FDR_{n,T}$, defined in (21), we have $FDR_{n,T} \to_p 0$, if $\delta > \max\{1, 2-\kappa_1/3\}$.

Since our proof requires that $0 < \varkappa < 1$, it is sufficient to set $\varkappa$ to be arbitrarily close to, but less than, unity. Also, κ_1 can be arbitrarily small which allows n to rise much faster than T. The condition $0 \leq \epsilon < \min\{1, \kappa_1/3\}$ ensures that $k^*/n \to 0$ and $k^* = o(T^{1/3})$.

Remark 1 *Assumption 6 allows for weak signals. In particular, we allow slope coefficients of order $\ominus\left(T^{-\vartheta}\right)$, for some $0 \leq \vartheta < 1/2$. Then, by (B.57) and (B.58) of Lemma A10 of the online supplement, it is seen that such weak signals can be picked up at no cost, in terms of rates, with respect to the exponential inequalities that underlie all the theoretical results. In particular, the power of the OCMT procedure in selecting the signal variable x_{it} rises with the ratio $\sqrt{T}\left|\theta_{i,(j)}\right|/\sigma_{e_i,(T)}\sigma_{x_i,(T)}$, so long as $\frac{c_p(n,\delta)}{\sqrt{T}\left|\theta_{i,(j)}\right|} \to 0$, as n and $T \to \infty$, where $\theta_{i,(j)}$ is given by (18), $\sigma_{e_i,(T)}$ and $\sigma_{x_i,(T)}$ are defined by (B.49), replacing $\mathbf{e}$, $\boldsymbol{x}$, and $\boldsymbol{M}_q$ by $\mathbf{e}_i$, $\boldsymbol{x}_i$, and $\boldsymbol{M}_{(j-1)}$, respectively. When this ratio is low, a large T will be required for the OCMT approach to select the i^{th} signal variable. This condition is similar to the so-called 'beta-min' condition assumed in the penalised regression literature. (See, for example, Section 7.4 of Buhlmann and van de Geer (2011) for a discussion.)*

Remark 2 *When the focus of the analysis is the true model, and not the approximating model that encompasses it, then the false discovery rate of the true model is given by*

$$FDR^*_{n,T} = \frac{\sum_{i=1}^{n} I(\widehat{\mathcal{J}}_i = 1, \text{ and } \beta_i = 0)}{\sum_{i=1}^{n} \widehat{\mathcal{J}}_i + 1}. \tag{33}$$

It is now easily seen that $FDR^*_{n,T}$ can tend to a nonzero value when pseudo-signals are present (i.e. if $k^* > 0$). In such cases, where the selection of the true model is the main objective of the analysis, a post-OCMT selection, using, for example, the Schwarz information criterion, could be considered to separate the signals from the pseudo-signals. However, when the norm of slope coefficients or the in-sample fit of the model is of main concern, then, under appropriate conditions on the rate at which k^* expands with n, the inclusion of pseudo-signals is asymptotically innocuous, as shown in Theorem 2 below.

Consider now the error and coefficient norms of the selected model, $F_{\boldsymbol{u}}$ and $F_{\tilde{\boldsymbol{\beta}}}$, defined in (22). We need the following additional regularity condition.

Assumption 7 *Let* $\mathbf{S}$ *denote the* $T \times l_T$ *observation matrix on the* l_T *regressors selected by the OCMT procedure. Then, let* $\mathbf{\Sigma}_{ss} = E\left(\mathbf{S}'\mathbf{S}/T\right)$ *with eigenvalues denoted by* $\mu_1 \leq \mu_2 \leq ... \leq \mu_{l_T}$. *Let* $\mu_i = O\left(l_T\right)$, $i = l_T - M + 1, l_T - M + 2, ..., l_T$, *for some finite* M, *and* $\sup_{1\leq i \leq l_T - M} \mu_i < C_0 < \infty$, *for some* $C_0 > 0$. *In addition,* $\inf_{1\leq i<l_T} \mu_i > C_1 > 0$, *for some* $C_1 > 0$.

Theorem 2 *Consider the DGP defined by (6), and the error and coefficient norms of the selected model,* $F_{\tilde{u}}$ *and* $F_{\tilde{\boldsymbol{\beta}}}$, *defined in (22). Suppose that Assumptions 1-4 and 6-7 hold, Assumption 5 holds for* x_{it} *and* $\boldsymbol{q}_{\cdot t} = \boldsymbol{x}_{(j-1),t}$, $i \in \mathfrak{A}_{(j)}$, $j = 1, 2, ...k$, *where* $\mathfrak{A}_{(j)}$ *is the active set at stage* j *of the OCMT procedure, and* k^* *(the number of pseudo-signals) is of order* $\ominus\left(n^{\epsilon}\right)$ *for some positive* ϵ. $c_p\left(n, \delta\right)$ *is given by (15) with* $0 < p < 1$ *and let* $f\left(n, \delta\right) = cn^{\delta}$, *for the first stage of OCMT, and* $f\left(n, \delta^*\right) = cn^{\delta^*}$ *for subsequent stages, for some* $c > 0$, $\delta^* > \delta > 0$. $n, T \to \infty$, *such that* $T = \ominus\left(n^{\kappa_1}\right)$, *for some* $\kappa_1 > 0$, *and* $k^* = \ominus(n^{\epsilon})$ *for some positive* $\epsilon < \min\left\{1, \kappa_1/3\right\}$. *Let* $\tilde{\boldsymbol{\beta}}_n$ *be the estimator of* $\boldsymbol{\beta}_n = (\beta_1, \beta_2, ..., \beta_n)'$ *in the final regression. Then, for any* $0 < \varkappa < 1$, *and some constant* $C_0 > 0$, *we have*

$$F_{\tilde{u}} = T^{-1}||\tilde{\boldsymbol{u}}||^2 = \sigma^2 + O_p(T^{-1/2}) + O(n^{3\epsilon}T^{-3/2}) = \sigma^2 + O_p(n^{-\kappa_1/2}) + O(n^{3\epsilon - 3\kappa_1/2}), \quad (34)$$

and

$$F_{\tilde{\boldsymbol{\beta}}} = ||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n|| = O_p(n^{5\epsilon/2 - \kappa_1}) + O_p(n^{\epsilon - \kappa_1/2}). \quad (35)$$

As can be seen from the above theorem, (34) and (35) require slightly stronger conditions than those needed for the proof of the earlier results in Theorem 1. In particular, a condition that relates to the eigenvalues of the population covariance of the selected regressors, denoted by $\mathbf{\Sigma}_{ss}$, is needed. It aims to control the rate at which $\|\mathbf{\Sigma}_{ss}^{-1}\|_F$ grows. It is mild in the sense that it allows for the presence of considerable collinearity between the regressors. Under this condition and $\epsilon < \min\left\{1, \kappa_1/3\right\}$, we in fact obtain an oracle rate of $T^{-1/2}$ for the error norm.

It is important to provide intuition on why we can get a consistency result for the coefficient norm of the selected model even though the selection process includes pseudo-signals. There are two reasons for this. First, since OCMT procedure selects all signals with probability approaching one as $n, T \to \infty$, then the coefficients of the additionally selected regressors (whether pseudo-signal or noise) will tend to zero with T. Second, restricting the rate at which k^* rises with n, as set out in Theorem 2, implies that the inclusion of pseudo-signals can be accommodated since their estimated coefficients will tend to zero and the variance of these estimated coefficients will be controlled.

In the case where hidden signals are not present, we have $P_0 = 1$, and as noted earlier further stages of the OCMT will not be required. Consequently, the results of Theorem 1 can be simplified and obtained under a less restrictive set of conditions. When $P_0 = 1$, and assuming that the conditions of Theorem 1 hold, with the exception of the condition on ϵ which could lie in $[0, 1)$, we obtain the following results, established in Section A.2.5 of the Appendix. The probability of selecting the approximating model is given by

$$\Pr\left(\mathcal{A}_0\right) = 1 + O\left(n^{1-\delta\varkappa}\right) + O\left[n \exp\left(-n^{C_0}\right)\right], \quad (36)$$

and $\Pr(\mathcal{A}_0) \to_p 1$, if $\delta > 1$. For the support recovery statistics, we have

$$E\left|TPR_{n,T}\right| = 1 + O\left[\exp\left(-n^{C_0}\right)\right], \text{ and} \tag{37}$$

$$E\left|FPR_{n,T}\right| = k^*/\left(n-k\right) + O\left(n^{-\delta\varkappa}\right) + O\left(n^{\epsilon-1}\right) + O\left[\exp(-n^{C_0})\right]. \tag{38}$$

Hence, if $\delta > 0$, then $TPR_{n,T} \to_p 1$, and $FPR_{n,T} \to_p 0$, and $FDR_{n,T} \to_p 0$, if $\delta > 1$.

4 Extensions

4.1 Alternative specifications for θ_i

Theorems 1 and 2, and the results discussed above relate to the first maintained assumption about the pseudo-signal variables where at most k^* of them have non-zero $\theta_{i,(j)}$ for some j. This result can be extended to the case where potentially all variables have non-zero θ_i, as long as θ_i's are absolutely summable. Two leading cases considered in the literature are to assume that there exists a (possibly unknown) ordering given by (4) or (5). The assumption that there is only a finite number of variables for which $\beta_i \neq 0$, is retained. The rationale for hidden signals is less clear for these cases, since rather than a discrete separation between variables with zero and non-zero θ_i, we consider a continuum that unites these two classes of variables. Essentially, we have no separation in terms of signals (or pseudo-signals) and noise variables, since under this setting there are no noise variables. Below, we provide some results for the settings implied by (4) and (5), proven in the online supplement.

Theorem 3 *Consider the DGP defined by (6), suppose that Assumptions 1-4 and 6 hold, Assumption 5 holds for x_{it} and $\boldsymbol{q}_{\cdot t} = 1$, $i = 1, 2, ..., n$, and condition (4) holds. Moreover, let $c_p(n, \delta)$ be given by (15) with $0 < p < 1$ and $f(n, \delta) = cn^{\delta}$, for some $c, \delta > 0$, and suppose there exists $\kappa_1 > 0$ such that $T = \ominus(n^{\kappa_1})$. Consider the variables selected by the OCMT procedure. Then, for all $\zeta > 0$, we have $E\left|FPR_{n,T}\right| = o(n^{\zeta-1}) + O\left[\exp(-n^{C_0})\right]$, for some finite positive constant C_0, where $FPR_{n,T}$ is defined by (20). If condition (5) holds instead of condition (4), then, assuming $\gamma > \frac{1}{2}\kappa_1^2$, we have $FPR_{n,T} \to_p 0$.*

4.2 Dynamic Extensions

An important assumption made so far is that noise variables are martingale difference processes which is restrictive in the case of time series applications. This assumption can be relaxed. In particular, under the less restrictive assumption that noise variables are exponentially mixing, it can be shown that all the theoretical results derived above hold. Details are provided in Section C of the online theory supplement. A further extension involves relaxing the martingale difference assumption for the signals and pseudo-signals. If we are willing to assume that either u_t is normally distributed or the covariates are deterministic, then a number of results become available. The relevant lemmas for the deterministic case are presented in Section

E of the online supplement. Alternatively, signals and pseudo-signals can be assumed to be exponentially mixing. In this general case, similar results to those in Theorems 1 and 2 can still be obtained. These are described in Section C of the online supplement. In the light of these theoretical extensions, one can also allow the DGP, (6), to include lagged dependent variables, $\mathbf{y}_{t,h} = (y_{t-1}, y_{t-2}, ..., y_{t-h})'$, where h is unknown. The OCMT procedure can now be applied to $\mathbf{x}_t$ augmented with $\mathbf{y}_{t,h_{\max}}$, where $h_{\max}$ is a maximum lag order selected by the investigator.

5 A Monte Carlo Study

We employ five different Monte Carlo (MC) designs, with or without lagged values of y_t. We allow the covariates to be serially correlated and consider different degrees of correlations across them. In addition, we experiment with Gaussian and non-Gaussian errors.

5.1 Data-generating processes (DGPs)

5.1.1 Design I (no hidden signals and no pseudo-signals)

y_t is generated as:

$$y_t = \varphi y_{t-1} + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \varsigma u_t, \tag{39}$$

where $u_t \sim IIDN(0,1)$ in the Gaussian case, and $u_t = [\chi_t^2(2) - 2]/2$ in the non-Gaussian case, in which $\chi_t^2(2)$ are independent draws from a χ^2-distribution with 2 degrees of freedom, for $t = 1, 2, ..., T$. We consider the 'static' specification with $\varphi = 0$, and two 'dynamic' specifications with $\varphi = 0.4$ and 0.8.[7] We set $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$ and consider the following alternative ways of generating $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$:

DGP-I(a) Temporally uncorrelated and weakly collinear covariates: Signal variables are generated as $x_{it} = (\varepsilon_{it} + \nu g_t)/\sqrt{1+\nu^2}$, for $i = 1, 2, 3, 4$, and noise variables are generated as $x_{5t} = \varepsilon_{5t}$, $x_{it} = (\varepsilon_{i-1,t} + \varepsilon_{it})/\sqrt{2}$, for $i > 5$, where g_t and ε_{it} are independent draws either from $N(0,1)$ or from $[\chi_t^2(2) - 2]/2$, for $t = 1, 2, ..., T$, and $i = 1, 2, ..., n$. We set $\nu = 1$, which implies 50% pair-wise correlation among the signal variables.

DGP-I(b) Temporally correlated and weakly collinear covariates: Covariates are generated as in DGP-I(a), but with $\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + \sqrt{1-\rho_i^2} e_{it}$, in which $e_{it} \sim IIDN(0,1)$ or $IID[\chi_t^2(2) - 2]/2$. We set $\rho_i = 0.5$ for all i.

DGP-I(c) Strongly collinear noise variables due to a persistent unobserved common factor: Signal variables are generated as $x_{it} = (\varepsilon_{it} + g_t)/\sqrt{2}$, for $i = 1, 2, 3, 4$, and noise variables are generated as $x_{5t} = (\varepsilon_{5t} + b_i f_t)/\sqrt{3}$ and $x_{it} = \left[(\varepsilon_{i-1,t} + \varepsilon_{it})/\sqrt{2} + b_i f_t\right]/\sqrt{3}$, for $i > 5$, where $b_i \sim IIDN(1,1)$, $f_t = 0.95 f_{t-1} + \sqrt{1-0.95^2} v_t$, and v_t, g_t and ε_{it} are independent draws from $N(0,1)$ or $[\chi_t^2(2) - 2]/2$.

[7] Dynamic processes are initialized from zero starting values and the first 100 observations are discarded.

DGP-I(d) Low or high pair-wise correlation of signal variables: Covariates are generated as in DGP-I(a), but we set $\nu = \sqrt{\omega / (1-\omega)}$, for $\omega = 0.2$ (low pair-wise correlation) and 0.8 (high pair-wise correlation). This ensures that average correlation among the signals is ω.

5.1.2 Design II (featuring pseudo-signals)

The DGP is given by (39) and $\boldsymbol{x}_{nt}$ is generated as:

DGP-II(a) Two pseudo-signals: Signal variables are generated as $x_{it} = (\varepsilon_{it} + g_t)/\sqrt{2}$, for $i = 1, 2, 3, 4$, pseudo-signal variables are generated as $x_{5t} = \varepsilon_{5t} + \kappa x_{1t}$, and $x_{6t} = \varepsilon_{6t} + \kappa x_{2t}$, and noise variables are generated as $x_{it} = (\varepsilon_{i-1,t} + \varepsilon_{it})/\sqrt{2}$, for $i > 6$, where, as before, g_t, and ε_{it} are independent draws from $N(0, 1)$ or $[\chi^2_t(2) - 2]/2$. We set $\kappa = 1.33$ (to achieve 80% correlation between the signal and the pseudo-signal variables).

DGP-II(b) All variables collinear with signals: $\boldsymbol{x}_{nt} \sim IID(\boldsymbol{0}, \boldsymbol{\Sigma}_x)$ with the elements of $\boldsymbol{\Sigma}_x$ given by $0.5^{|i-j|}$, $1 \leq i, j \leq n$. We generate $\boldsymbol{x}_{nt}$ with Gaussian and non-Gaussian innovations. In particular, $\boldsymbol{x}_{nt} = \boldsymbol{\Sigma}_x^{1/2} \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{nt})'$, and ε_{it} are generated as independent draws from $N(0, 1)$ or $[\chi^2_t(2) - 2]/2$.

5.1.3 Design III (featuring hidden signals)

y_t is generated by (39), $\boldsymbol{x}_{nt}$ is generated as in DGP-I(a), and the slope coefficients for the signals in (39) are selected so that, conditional on y_{t-1}, $\theta_4 = 0$:

DGP-III The fourth variable is hidden signal: We set $\beta_1 = \beta_2 = \beta_3 = 1$ and $\beta_4 = -1.5$. This implies $\theta_i \neq 0$ for $i = 1, 2, 3$ and $\theta_i = 0$ for $i \geq 4$, conditional on y_{t-1}.

5.1.4 Design IV (featuring both hidden signals and pseudo-signals)

In this case y_t is generated by (39), and:

DGP-IV(a) We generate $\boldsymbol{x}_{nt}$ in the same way as in DGP-II(a) which features two pseudo-signal variables. We generate slope coefficients β_i as in DGP-III to ensure $\theta_i \neq 0$ for $i = 1, 2, 3$, and $\theta_i = 0$ for $i = 4$, conditional on y_{t-1}.

DGP-IV(b) We generate $\boldsymbol{x}_{nt}$ in the same way as in DGP-II(b), where all covariates are collinear with signals. We set $\beta_1 = -0.875$ and $\beta_2 = \beta_3 = \beta_4 = 1$. This implies $\theta_i = 0$ for $i = 1$ and $\theta_i > 0$ for all $i > 1$, conditional on y_{t-1}.

5.1.5 Design V (Many signals)

For this design the DGP (**DGP-V**) is given by

$$y_t = \varphi y_{t-1} + \textstyle\sum_{i=1}^{n} i^{-2} x_{it} + \varsigma u_t, \tag{40}$$

where $\boldsymbol{x}_{nt}$ are generated as in design DGP-II(b), and u_t is generated in the same way as before. This design is inspired by the literature on approximately sparse models (Belloni et al. (2014b)).

Autoregressive processes are generated with zero starting values and 100 burn-in periods. ς is set so that $R^2 = 30\%$, 50% or 70% (on average) in static specifications ($\varphi = 0$). We do not change any parameters of the designs with an increase in φ, and we refer to the three R^2 measures corresponding to the three choices of ς as a low, medium and high fit. The sample combinations, $n = (100, 200, 300)$ and $T = (100, 300, 500)$ are considered, and all experiments are carried out using $R_{MC} = 2,000$ replications.

5.2 Variable selection methods

We consider six variable selection procedures, namely OCMT, Lasso, Adaptive Lasso (A-Lasso), Hard thresholding, SICA, and Boosting. In static specifications, the OCMT method is implemented as outlined in Section 3, where $c_p(n, \delta)$ is defined by (15) with $f(n, \delta) = n^{\delta}$ in the first stage and $f(n, \delta^*) = n^{\delta^*}$ in the subsequent stages. We use $p = 0.01$, and in line with the theoretical derivations we set $\delta = 1$ and $\delta^* = 2$. An online MC supplement provides results for other choices of $p \in \{0.01, 0.05, 0.1\}$ and $(\delta, \delta^*) \in \{(1, 1.5), (1, 2)\}$. It turns out that the choice of p is of second order importance. In the dynamic case, we augment the set of n covariates with $h_{\max} = 4$ lags of the dependent variable. Penalised regressions are implemented using the same set of possible values for the penalisation parameter λ as in Zheng et al. (2014), and following the literature λ is selected using 10-fold cross-validation. All methods are described in detail in the online MC supplement.

5.3 Monte Carlo results

We begin by reporting on the number of stages, denoted by $\hat{P}$, taken by OCMT before completion. This is important since our theory suggests that it should be close to P_0, which is 1 for DGPs I, II, and V without hidden signals, and 2 in the case of DGPs III and IV that do contain hidden signals. Realizations of $\hat{P}$ are very close to P_0 for both groups of experiments. The average number of stages in the two groups of experiments is $\overline{\hat{P}} = 1.03$ and 1.78, respectively. In addition, the frequency of MC replications with $\hat{P} > P_0$ and $\hat{P} > P_0 + 1$ turn out to be very small and amounted to 1.6%, and 0.003%, respectively.

Next, we focus on the average performance of Lasso, adaptive Lasso and OCMT methods, whilst the full set of results for all experiments and all six variable selection procedures is given in the online supplement. In our comparisons we focus on Lasso and adaptive Lasso since these are the main penalised regression methods used in the literature and also because they tend to perform better than Boosting. In our evaluation we use the following criteria: the true positive rate (TPR) defined by (20), the false positive rate (FPR) defined by (20), the false discovery rate of the true model (FDR*) defined by (33), the false discovery of the approximating model (FDR) defined by (21), the out-of-sample root mean square forecast error (RMSFE), and the root mean square error of $\tilde{\boldsymbol{\beta}}$ (RMSE$_{\tilde{\boldsymbol{\beta}}}$).[8] We find that no method uniformly outperforms in the set of experiments we consider. This is true for the full set of methods (OCMT, Lasso, adaptive Lasso, Hard thresholding, SICA and Boosting) reported in the online supplement. The performance of individual methods can be quite different for individual experiments, and a relative assessment of these methods is provided in Table 1, which reports the fraction of experiments (in percent) where OCMT is outperformed by Lasso and Adaptive Lasso. These results clearly show that no method universally dominates. But it is interesting that the fraction of such experiments where OCMT is beaten by its competitors is relatively small, at most 22% for RMSFE and RMSE$_{\tilde{\boldsymbol{\beta}}}$ entries, in all experiments with the exception of dynamic specifications with $\varphi = 0.8$.

Summary statistics across the three choices of R^2 (low medium and high) and all the sample sizes ($n = 100, 200, 300$ and $T = 100, 300, 500$), for each of the five DGPs and with or without the lagged dependent variable, are reported Table A.1 in the Appendix. Lasso's TPR is in the majority of experiments larger than OCMT's, but so is the FPR and FDR as Lasso tends to overestimate the number of signals, which is well known in the literature. Adaptive Lasso in turn achieves better FPR and FDR outcomes compared with Lasso, but the performance of adaptive Lasso can be worse for TPR, RMSFE and RMSE$_{\tilde{\boldsymbol{\beta}}}$ in these experiments. The reported RMSFE and RMSE$_{\tilde{\boldsymbol{\beta}}}$ averages of Lasso and Adaptive Lasso are outperformed by OCMT in static specifications and dynamic specifications with low value of $\varphi = 0.4$ in Table A.1, by about 1.6% to 3.4%, and 9.1% to 40%, respectively. OCMT is very successful at eliminating the noise variables. On the other hand, the power of OCMT procedure to pick up the signals rises with $\sqrt{T}\left|\theta_{i,(j)}\right| / \sigma_{e_i,(T)}\sigma_{x_i,(T)}$, see Remark 1.[9] Hence the magnitude of $\theta_{i,(j)}$, T and R^2 are all important for the power of the OCMT. For instance, detailed findings reported in the online supplement show that an increase in the collinearity among signal variables, which results in a larger $\theta_{i,(j)}$, improves the performance of OCMT, but it worsens the performance of Lasso, since a higher collinearity of signal variables diminishes the marginal contribution of signals to the fit of the model. The performance of OCMT method also deteriorates with an increase in φ, and we see that in dynamic specifications with $\varphi = 0.8$ reported in the bottom panel of Table A.1, OCMT is beaten by Lasso and/or Adaptive Lasso in some instances. Findings for the non-Gaussian experiments are presented in Table A.2 in Appendix, which shows that the effects of allowing for non-Gaussian innovations seem to be rather marginal.

Overall, the small sample evidence suggests that the OCMT method is a valuable alternative to penalised regressions, since, in many cases, it can outperform the penalised regressions, that have become the *de facto* benchmark in the literature.

[8] RMSE$_{\tilde{\boldsymbol{\beta}}}$ is the square root of the trace of the MSE matrix of $\tilde{\boldsymbol{\beta}}$. Additional summary statistics, including the frequency of selecting the true model, and the statistics summarizing the distribution of the number of selected covariates are reported in the online supplement.

[9] $\sigma_{e_i,(T)}$ and $\sigma_{x_i,(T)}$ are defined by (B.49) in the online theory supplement, replacing $\mathbf{e}$, $\boldsymbol{x}$, and $\boldsymbol{M}_q$ by $\mathbf{e}_i$, $\boldsymbol{x}_i$, and $\boldsymbol{M}_{(j-1)}$, respectively.

6 Empirical Illustration

In this section we present an empirical application that highlights the utility of OCMT. In particular, we present a macroeconomic forecasting exercise for US GDP growth and CPI inflation using a large set of macroeconomic variables. The data set is quarterly and comes from Stock and Watson (2012). We use the smaller data set considered in Stock and Watson (2012), which contains 109 series. The series are transformed by taking logarithms and/or differencing following Stock and Watson (2012).[10] The transformed series span 1960Q3 to 2008Q4 and are collected in the vector $\boldsymbol{\xi}_t$ together with the target variable y_t (either US GDP growth or differenced log CPI inflation). Our estimation period is from 1960Q3 to 1990Q2 (120 periods) while the forecast evaluation period is 1990Q3 to 2008Q4. We produce one step ahead forecasts using five different procedures:[11] (a) AR benchmark with the number of lags selected by Schwarz Bayesian criterion (SBC) with maximum lag set equal to $h_{\max}$; (AR), (b) AR augmented with one lag of principal components, and the number of lags of the dependent variable is selected by SBC with maximum lag $h_{\max}$; (factor-augmented AR), (c-d) Lasso and adaptive Lasso regressions of the target variable y_t on lagged principal components, $\boldsymbol{\xi}_{t-1}$, and $h_{\max}$ lags of y_t. For Lasso and adaptive Lasso regressions, both the target variable and regressors are demeaned, and the regressors are normalised to have unit variances. (e) OCMT procedure is applied to regressions of y_t conditional on lagged principal components (included as pre-selected regressors), with $\boldsymbol{\xi}_{t-1}$ and $h_{\max}$ lags of y_t considered for variable selection. We set $\delta = 1$ in the first stage of OCMT, and $\delta^* = 2$ in the subsequent stages. We consider $p = 0.05$ below and findings for $p = 0.01$ and 0.1 are reported in the online empirical supplement. In all three data-rich procedures (b) to (e), the principal components are selected in a rolling scheme by the PC_{p_1} Bai and Ng (2002) criterion (with the maximum number of PCs set to 5). The maximum number of lags for the dependent variable, $h_{\max}$, is set to 4. We generate rolling forecasts using a rolling window of 120 observations.

We evaluate the forecasting performance of the methods using relative RMSFE where the AR forecast is the benchmark. Relative RMSFE statistics for the whole evaluation sample as well as for the pre-crisis sub-period (1990Q3-2007Q2) are reported in Table 2. In the case of GDP growth forecasts, we note that factor-augmented AR, Lasso and OCMT methods perform better than the AR benchmark. OCMT performs the best while Adaptive Lasso is the worst performer. However, the performance of the best methods is very close.[12] The differences in RMSFE in the case of inflation, reported in the bottom half of Table 2, are also relatively small with the factor-augmented AR(1) performing the best followed by OCMT and Lasso.

Variable inclusion frequencies are reported in Table 3, using the full evaluation sample.

[10] For further details, see the online supplement of Stock and Watson (2012), in particular columns E and T of their Table B.1.

[11] Further detail is provided in the online empirical supplement.

[12] Diebold-Mariano test statistics for all pairwise method comparisons can be found in the online supplement. The RMSFE differences among the best performing methods are not generally statistically significant.

Interestingly, for forecasting growth, the first lag of the dependent variable is among the most selected variables using OCMT (with the inclusion frequency of 45.9%), while no lags of the dependent variable are selected in the case of Lasso in any of the rolling windows. Results are different when inflation is considered. In this case, the inclusion frequency of the first lag of the dependent variable is 100% for both OCMT and Lasso methods. OCMT selects considerably fewer number of variables as compared to Lasso, an outcome that mirrors the Monte Carlo findings. In summary, we see that there is no method that uniformly outperforms all competitor methods and that OCMT is not far behind the best performing method.

7 Conclusion

Model selection is a recurring and fundamental topic in econometric analysis. This problem has become considerably more difficult for large-dimensional data sets where the set of possible specifications rise exponentially with the number of available covariates. In the context of linear regression models, penalised regression has become the *de facto* benchmark method of choice. However, issues such as the choice of penalty function and tuning parameters remains contentious.

In this paper, we provide an alternative approach based on multiple testing that is computationally simple, fast, and effective for sparse regression functions. Extensive theoretical and Monte Carlo results highlight these properties. In particular, we find that although no single method dominates across the broad set of experiments we considered, our proposed method can in many instances outperform existing penalised regression methods, whilst at the same time being computationally much faster by some orders of magnitude.

There are a number of avenues for future research. We have already considered the possibility of allowing for dynamics, but further extensions to more general settings with weakly exogenous regressors is clearly desirable. For empirical economic applications it is also important to allow for the possibility of weak and strong common factors affecting both the signal and pseudo-signal variables. A further possibility is to extend the idea of considering regressors individually to other testing frameworks, such as tests of forecasting ability. It is hoped that the results presented in this paper provide a basis for such further developments and empirical applications.

Table 1: Fraction of experiments (in percent) where OCMT is beaten by Lasso (L) and Adpaptive Lasso (A-L)

	Experiments with Gaussian innovations										Experiments with non-Gaussian innovations									
DGP type:	**DGP-I**		**DGP-II**		**DGP-III**		**DGP-IV**		**DGP-V**		**DGP-I**		**DGP-II**		**DGP-III**		**DGP-IV**		**DGP-V**	
No. of experiments:	135		54		27		54		27		135		54		27		54		27	
OCMT beaten by(*):	L	A-L	L	A-L	L	A-L	L	A-L	L	A-L	L	A-L	L	A-L	L	A-L	L	A-L	L	A-L
Static Specifications																				
TPR	15.6	6.7	20.4	3.7	44.4	29.6	59.3	38.9	100.0	3.7	17.8	6.7	22.2	3.7	48.1	22.2	66.7	38.9	96.3	3.7
FPR	0.0	0.0	0.0	18.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	22.2	0.0	0.0	0.0	0.0	0.0	0.0
FDR* (true model)	0.0	0.0	0.0	46.3	0.0	0.0	0.0	13.0	0.0	0.0	0.0	0.0	0.0	48.1	0.0	0.0	0.0	16.7	0.0	0.0
FDR (approximating model)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RMSFE	2.2	0.7	0.0	0.0	11.1	3.7	5.6	1.9	0.0	0.0	2.2	1.5	0.0	0.0	11.1	3.7	5.6	1.9	0.0	0.0
$\text{RMSE}_{\tilde{\beta}}$	8.9	0.7	14.8	0.0	11.1	3.7	5.6	1.9	0.0	0.0	14.8	0.7	14.8	1.9	11.1	3.7	9.3	1.9	0.0	0.0
Dynamic Specifications																				
Experiments with $\varphi = 0.4$																				
TPR	30.4	13.3	38.9	16.7	55.6	40.7	64.8	51.9	100.0	44.4	33.3	16.3	38.9	18.5	55.6	44.4	70.4	57.4	100.0	33.3
FPR	0.0	0.0	0.0	9.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	0.0	0.0	0.0	0.0	0.0	0.0
FDR* (true model)	0.0	1.5	0.0	33.3	0.0	0.0	0.0	9.3	0.0	0.0	0.0	1.5	0.0	35.2	0.0	0.0	0.0	11.1	0.0	0.0
FDR (approximating model)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RMSFE	8.9	7.4	11.1	7.4	22.2	18.5	16.7	13.0	3.7	3.7	9.6	9.6	11.1	9.3	22.2	22.2	16.7	18.5	11.1	11.1
$\text{RMSE}_{\tilde{\beta}}$	14.8	2.2	11.1	0.0	11.1	11.1	11.1	5.6	0.0	0.0	14.8	2.2	13.0	0.0	11.1	14.8	14.8	7.4	0.0	0.0
Experiments with $\varphi = 0.8$																				
TPR	64.4	43.0	75.9	61.1	66.7	66.7	83.3	83.3	100.0	100.0	71.9	42.2	81.5	61.1	70.4	66.7	85.2	83.3	100.0	100.0
FPR	20.0	65.9	0.0	53.7	0.0	14.8	0.0	27.8	0.0	70.4	20.0	69.6	0.0	59.3	0.0	14.8	0.0	27.8	0.0	74.1
FDR* (true model)	10.4	90.4	1.9	85.2	3.7	40.7	0.0	66.7	3.7	100.0	11.1	91.9	1.9	90.7	0.0	37.0	0.0	70.4	3.7	100.0
FDR (approximating model)	0.0	10.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RMSFE	24.4	42.2	3.7	24.1	14.8	29.6	14.8	40.7	0.0	7.4	25.9	41.5	7.4	14.8	14.8	22.2	14.8	38.9	0.0	0.0
$\text{RMSE}_{\tilde{\beta}}$	60.0	45.2	55.6	44.4	44.4	37.0	55.6	51.9	40.7	0.0	60.7	44.4	55.6	37.0	44.4	44.4	55.6	57.4	29.6	0.0

Notes: (*) L: Lasso, A-L: Adaptive Lasso. DGPs I-IV are given by (39) and DGP V is given by (40). In the static case, DGP does not include lag dependent variable and selection of lags of the dependent variable is not considered. In the dynamic case, DGP includes one lag of the dependent variable, and the selection of up to $h_{\max} = 4$ lags of the dependent variable is considered. TPR (FPR) is the true (false) positive rate. FDR* is the false discovery rate for the true model and FDR is the false discovery rate for the approximating model. RMSFE is the root mean square forecast error. $\text{RMSE}_{\tilde{\beta}}$ is the root mean square error of $\tilde{\boldsymbol{\beta}}$. In DGP V, TPR is computed assuming that covariates $i = 1, 2, ..., 11$ are the signal variables, and FPR and FDR are computed assuming covariates $i > 11$ are the noise variables. In the case of Oracle method the identity of rue variables is known. In DGP V, Oracle* method assumes the first 11 covariates are the signal variables. Lasso is implemented using the same set of possible values for the penalisation parameter λ as in Zheng et al. (2014), and λ is selected using 10-fold cross-validation. Adaptive Lasso method is implemented as described in Section 2.8.4 of Buhlmann and van de Geer (2011) based on the implementation of the Lasso method described above. OCMT results are based on $p = 0.01$, $\delta = 1$ in the first stage, and $\delta^* = 2$ in the subsequent stages of the OCMT procedure. See Section 5 for further details. The complete set of findings is reported in the online MC supplement.

Table 2: RMSFE performance of the AR, factor-augmented AR, Lasso and OCMT methods

Evaluation sample:	Full		Pre-crisis	
	1990Q3-2008Q4		1990Q3-2007Q2	
	RMSFE (×100)	Relative RMSFE	RMSFE (×100)	Relative RMSFE
	Real output growth			
AR benchmark	0.561	1.000	0.505	1.000
Factor-augmented AR	0.484	0.862	0.470	0.930
Lasso	0.510	0.910	0.465	0.922
Adaptive Lasso	0.561	1.000	0.503	0.996
OCMT	0.477	0.850	0.461	0.912
	Inflation			
AR benchmark	0.601	1.000	0.435	1.000
Factor-augmented AR	0.557	0.927	0.415	0.954
Lasso	0.599	0.997	0.462	1.063
Adaptive Lasso	0.715	1.190	0.524	1.205
OCMT	0.590	0.982	0.464	1.068

Notes: RMSFE is computed based on rolling forecasts with a rolling window of 120 observations. The source of the data is the smaller data set with 109 time series provided by Stock and Watson (2012). The series are transformed by taking logarithms and/or differencing following Stock and Watson (2012). The transformed series span 1960Q3 to 2008Q4 and are collected in the vector $\boldsymbol{\xi}_t$. Set of regressors in Lasso and adaptive-Lasso contains $h_{\max} = 4$ lags of y_t (lagged target variables), $\boldsymbol{\xi}_{t-1}$, and a lagged set of principal components obtained from the large data set given by $(y_t, \boldsymbol{\xi}_t')'$. OCMT procedure is applied to regressions of y_t conditional on lagged principal components (included as pre-selected regressors) with $\boldsymbol{\xi}_{t-1}$ and $h_{\max} = 4$ lags of y_t considered for variable selection. OCMT is reported for $p = 0.05$ and $\delta = 1$ in the first stage, and $p = 0.05$ and $\delta^* = 2$ in the subsequent stages of the OCMT procedure. The number of principal components in the factor-augmented AR, Lasso, adaptive-Lasso, and OCMT methods is determined in a rolling scheme by using criterion PC_{p_1} of Bai and Ng (2002) (with the maximum number of PCs set to 5). See Section 6 and the online empirical supplement for further details.

Table 3: Top 5 variables with highest inclusion frequencies based on the Lasso and OCMT selection methods

Output growth			
Lasso		**OCMT**	
1. Real gross private domestic investment - residential (*)	100.0%	1. Residential price index	47.3%
2. Real personal consumption expenditures - services (*)	100.0%	2. First lag of the dependent variable	45.9%
3. Employees, nonfarm - mining	89.2%	3. Industrial production index - fuels	43.2%
4. Index of help - wanted advertising in newspapers	75.7%	4. Labor productivity (output per hour)	37.8%
5. Employment: Ratio; Help-wanted ads: No. unemployed CLF	56.8%	5. Employees, nonfarm - mining	27.0%
Average number of selected variables	8.1	Average number of selected variables (excluding pre-selected factors)	2.2
Inflation			
Lasso		**OCMT**	
1. Interest rate: U.S. Treasury bills, sec. mkt, 3-mo (% per ann)	100.0%	1. First lag of the dependent variable	100.0%
2. Real personal consumption expenditures - services (*)	100.0%	2. Third lag of the dependent variable	78.4%
3. First lag of the dependent variable	100.0%	3. MZM money stock (FRB St. Lois)	71.6%
4. Employees, nonfarm - mining	98.6%	4. Money stock: M2	45.9%
5. Second lag of the dependent variable	98.6%	5. Recreation price index	33.8%
Average number of selected variables	21.7	Average number of selected variables (excluding pre-selected factors)	4.0

Notes: This table reports the top 5 highest inclusion frequencies of the variables selected using the Lasso and OCMT procedure on the full evaluation sample, 1990Q3-2008Q4. OCMT is reported $p = 0.05$ and for $\delta = 1$ in the first stage, and $\delta^* = 2$ in the subsequent stages of the OCMT procedure.
(*) quantity index.

A Appendix

A.1 Additional notations and definitions

Throughout this appendix we consider the following events:

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

$\mathcal{A}_0$, also defined by (28), is the event of selecting the approximating model, $\mathcal{H}$ is the event that all signals are selected, and $\mathcal{G}$ is the event that no noise variable is selected. We also denote the event that exactly j noise variables are selected by $\mathcal{G}_j = \{\sum_{i=k+k^*+1}^{n} \widehat{\mathcal{J}}_i = j\}$, for $j = 0, 1, ..., n-k-k^*$, with $\mathcal{G} \equiv \mathcal{G}_0$. For the analysis of different stages of OCMT, we also introduce the event $\mathcal{B}_{i,s}$, which is the event that variable i is selected at the s^{th} stage of the OCMT procedure. $\mathcal{L}_{i,s} = \cup_{h=1}^{s} \mathcal{B}_{i,h}$ is the event that variable i is selected up to and including stage s, namely in any of the stages $j = 1, 2, ..., s$ of the OCMT procedure, and $\mathcal{L}_s = \cap_{i=1}^{k} \mathcal{L}_{i,s}$ is the event that all signals are selected up to and including stage s of the OCMT procedure. $\mathcal{T}_s$ is the event that OCMT stops after s stages or less. $\mathcal{D}_{s,T}$ is the event that the number of variables selected in the first s stages of OCMT ($\hat{k}_{(j)}$, $j = 1, 2, ..., s$) is smaller than or equal to l_T, where $l_T = \ominus(n^{\nu})$ and ν satisfies $\epsilon < \nu < \kappa_1/3$. Note that when $T = \ominus(n^{\kappa_1})$ then $l_T = \ominus\left(T^{\nu/\kappa_1}\right) = o\left(T^{1/3}\right)$ for $\nu < \kappa_1/3$.

Notations: Let $\mathbf{a} = (a_1, a_2, ..., a_n)'$ and $\mathbf{A} = (a_{ij})$ be an $n \times 1$ vector and an $n \times m$ matrix, respectively. Then, $\|\mathbf{a}\| = (\Sigma_{i=1}^{n} a_i^2)^{1/2}$ and $\|\mathbf{a}\|_1 = \Sigma_{i=1}^{n} |a_i|$ are the Euclidean (L_2) and L_1 norms of $\mathbf{a}$, respectively. $\|\boldsymbol{A}\|_F = [Tr(\boldsymbol{A}\boldsymbol{A}')]^{1/2}$ is the Frobenius norm of $\mathbf{A}$.

A.2 Proofs of Propositions and Theorems

All proofs are based on the set of lemmas presented and established in the online theory supplement. In particular, Lemmas A1-A9 are auxiliary ones, mostly providing supporting results for the main lemma of the paper, namely Lemma A10, which provides the basic exponential inequalities that underlie most of our results. A simple version of this lemma is included in the paper as Proposition 2.

A.2.1 Proof of Proposition 1

We recall that P_0 is a population quantity. This formally means that, to determine P_0, OCMT is carried out assuming $\Pr[|t_{\hat{\phi}_{i,(j)}}| > c_p(n,\delta) \,|\theta_{i,(j)} \neq 0] = 1$, and $\Pr[|t_{\hat{\phi}_{i,(j)}}| > c_p(n,\delta) \,|\theta_{i,(j)} = 0] = 0$ for all i, j. So, if $\theta_{i,(1)} \neq 0$, for all i for which $\beta_i \neq 0$, it obviously follows that $P_0 = 1$. Next, assume that the subset of signal variables in $\mathbf{X}_k$, such that for each element of this subset, $\theta_{i,(1)} = 0$, is not empty. Then, these signals will not be selected in the first stage of OCMT. By Lemma A1 in the online supplement, it follows that the subset of signals for which $\theta_{i,(1)} = 0$ is smaller than the set of signals and therefore at least one signal will be picked up in the first stage of OCMT. It then follows, by Lemma A1, that in the second stage of OCMT, at least one hidden signal, for which $\theta_{i,(1)} = 0$ will have $\theta_{i,(2)} \neq 0$. Therefore, such hidden signal(s) will be picked up in the second stage. Proceeding recursively using Lemma A1, it then follows that all hidden signals for which $\theta_{i,(1)} = 0$, will satisfy $\theta_{i,(j)} \neq 0$ for some $j \leq k$, proving the proposition.[13]

A.2.2 Proof of Theorem 1

Noting that $\mathcal{T}_k$ is the event that the OCMT procedure stops after k stages or less, we have $\Pr\left(\hat{P} > k\right) = \Pr\left(\mathcal{T}_k^c\right) = 1 - \Pr\left(\mathcal{T}_k\right)$, where $\hat{P}$ is defined by (17). Substituting (B.83) of Lemma A20 in the online supplement for $\Pr\left(\mathcal{T}_k\right)$, we obtain, $\Pr\left(\hat{P} > k\right) = O\left(n^{1-\nu-\varkappa\delta}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left[n\exp\left(-C_0 n^{C_1\kappa_1}\right)\right]$, for some $C_0, C_1 > 0$, any $\varkappa$ in $0 < \varkappa < 1$, and any ν in $0 \leq \epsilon < \nu < \kappa_1/3$, where $\kappa_1 > 0$ defines the rate for $T = \ominus\left(n^{\kappa_1}\right)$, and ϵ in $0 \leq \epsilon < \min\{1, \kappa_1/3\}$ defines the rate for $k^* = \ominus\left(n^{\epsilon}\right)$. But note that $O\left(n^{1-\nu-\varkappa\delta}\right)$ can be written equivalently as $O\left(n^{1-\kappa_1/3-\varkappa\delta}\right)$. This follows since $1 - \kappa_1/3 - \varkappa\delta = 1 - \left(\kappa_1/3 - \varepsilon\delta\right) - \left(\varkappa + \varepsilon\right)\delta = 1 - \tilde{\nu} - \tilde{\varkappa}\delta$, where $\tilde{\nu} = \kappa_1/3 - \varepsilon\delta$ and $\tilde{\varkappa} = \varkappa + \varepsilon$, for $\varepsilon > 0$ sufficiently small. Specifically, setting $\varepsilon < \min\{1 - \varkappa, \left(\kappa_1/3 - \epsilon\right)/\delta\}$, it follows that $\tilde{\varkappa}$ and $\tilde{\nu}$ satisfy $0 < \tilde{\varkappa} < 1$ and $\epsilon < \tilde{\nu} < \kappa_1/3$, respectively, as required. Hence

$$\Pr(\hat{P} > k) = \Pr\left(\mathcal{T}_k^c\right) = O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left[n\exp\left(-C_0 n^{C_1\kappa_1}\right)\right], \tag{A.2}$$

for some $C_0, C_1 > 0$ and any $\varkappa$ in $0 < \varkappa < 1$. Noting that $O\left[n\exp\left(-C_0 n^{C_1\kappa_1}\right)\right] = O\left[\exp\left(-n^{C_2\kappa_1}\right)\right]$ for any $0 < C_2 < C_1$, we have $\Pr\left(\hat{P} > k\right) = O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left[\exp\left(-n^{C_2\kappa_1}\right)\right]$, for some $C_2 > 0$, which establishes (29). Similarly, by (B.86) and noting that $n \geq n^{1-\nu}$ for $\nu \geq 0$, we also have (which is required subsequently)

$$\Pr\left(\mathcal{D}_{k,T}^c\right) = O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left(n^{1-\kappa_1/3-\varkappa\delta^*}\right) + O\left[n\exp\left(-C_0 T^{C_1\kappa_1}\right)\right], \tag{A.3}$$

for some $C_0, C_1 > 0$ and any $\varkappa$ in $0 < \varkappa < 1$.

To establish result (30), we first note that

$$\Pr(\mathcal{A}_0^c) = \Pr(\mathcal{A}_0^c|\mathcal{D}_{k,T})\Pr(\mathcal{D}_{k,T}) + \Pr(\mathcal{A}_0^c|\mathcal{D}_{k,T}^c)\Pr(\mathcal{D}_{k,T}^c) \leq \Pr(\mathcal{A}_0^c|\mathcal{D}_{k,T}) + \Pr(\mathcal{D}_{k,T}^c), \tag{A.4}$$

where $\Pr(\mathcal{D}_{k,T}^c)$ is given by (A.3). Also using (A.1) we have $\mathcal{A}_0^c = \mathcal{H}^c \cup \mathcal{G}^c$, and hence

$$\Pr(\mathcal{A}_0^c|\mathcal{D}_{k,T}) \leq \Pr\left(\mathcal{H}^c|\,\mathcal{D}_{k,T}\right) + \Pr\left(\mathcal{G}^c|\,\mathcal{D}_{k,T}\right) = A_{n,T} + B_{n,T}, \tag{A.5}$$

where $\mathcal{H}$ and $\mathcal{G}$ are given by (A.1). Therefore $\mathcal{H}^c = \{\sum_{i=1}^{k}\widehat{\mathcal{J}}_i < k\}$, and $\mathcal{G}^c = \{\sum_{i=k+k^*+1}^{n}\widehat{\mathcal{J}}_i > 0\}$. Consider the terms $A_{n,T}$ and $B_{n,T}$, in turn:

$$A_{n,T} = \Pr\left(\mathcal{H}^c|\,\mathcal{D}_{k,T}\right) \leq \textstyle\sum_{i=1}^{k}\Pr(\widehat{\mathcal{J}}_i = 0\Big|\,\mathcal{D}_{k,T}). \tag{A.6}$$

[13] Note that this proposition allows the net effects to tend to zero with T (or n) at a sufficiently slow rate as set out in Assumption 6, as long as they are not exactly zero. See also Lemma A1 in the online supplement.

But, the event $\{\widehat{\mathcal{J}}_i = 0|\mathcal{D}_{k,T}\}$ can occur only if $\{\cap_{j=1}^k \mathcal{B}_{i,j}^c|\mathcal{D}_{k,T}\}$ occurs, while $\{\cap_{j=1}^k \mathcal{B}_{i,j}^c|\mathcal{D}_{k,T}\}$ can occur without $\{\widehat{\mathcal{J}}_i = 0|\mathcal{D}_{k,T}\}$ occurring. Therefore, $\Pr[\widehat{\mathcal{J}}_i = 0|\mathcal{D}_{k,T}] \leq \Pr(\cap_{j=1}^k \mathcal{B}_{i,j}^c|\mathcal{D}_{k,T})$. Then,

$$\Pr\left(\cap_{j=1}^k \mathcal{B}_{i,j}^c \middle| \mathcal{D}_{k,T}\right) = \Pr\left(\mathcal{B}_{i,1}^c \middle| \mathcal{D}_{k,T}\right) \times \Pr\left(\mathcal{B}_{i,2}^c \middle| \mathcal{B}_{i,1}^c, \mathcal{D}_{k,T}\right) \times \Pr\left(\mathcal{B}_{i,3}^c \middle| \mathcal{B}_{i,2}^c \cap \mathcal{B}_{i,1}^c, \mathcal{D}_{k,T}\right) \times \ldots \times \Pr\left(\mathcal{B}_{i,k}^c \middle| \mathcal{B}_{i,k-1}^c \cap \ldots \cap \mathcal{B}_{i,1}^c, \mathcal{D}_{k,T}\right). \quad \text{(A.7)}$$

But, by Proposition 1 we are guaranteed that for some j in $1 \leq j \leq k$, $\theta_{i,(j)} \neq 0$, $i = 1, 2, ..., k$. Therefore, for some j in $1 \leq j \leq k$,

$$\Pr(\mathcal{B}_{i,j}^c \big| \mathcal{B}_{i,j-1}^c \cap \ldots \cap \mathcal{B}_{i,1}^c, \mathcal{D}_{k,T}) = \Pr(\mathcal{B}_{i,j}^c \big| \mathcal{B}_{i,j-1}^c \cap \ldots \cap \mathcal{B}_{i,1}^c, \theta_{i,(j)} \neq 0, \mathcal{D}_{k,T}),$$

and by (B.52) of Lemma A10 in the online supplement, $\Pr(\mathcal{B}_{i,j}^c \big| \mathcal{B}_{i,j-1}^c \cap \ldots \cap \mathcal{B}_{i,1}^c, \theta_{i,(j)} \neq 0, \mathcal{D}_{k,T}) = O\left[\exp\left(-C_0 T^{C_1}\right)\right]$, for $i = 1, 2, ..., k$, and some $C_0, C_1 > 0$. Therefore,

$$\Pr(\widehat{\mathcal{J}}_i = 0 \mid \mathcal{D}_{k,T}) = O\left[\exp\left(-C_0 T^{C_1}\right)\right], \text{ for } i = 1, 2, ..., k. \quad \text{(A.8)}$$

Substituting this result in (A.6), we have

$$A_{n,T} = \Pr\left(\mathcal{H}^c \middle| \mathcal{D}_{k,T}\right) \leq k \exp\left(-C_0 T^{C_1}\right). \quad \text{(A.9)}$$

Similarly, for $B_{n,T}$ we first note that

$$B_{n,T} = \Pr[\cup_{i=k+k^*+1}^n (\widehat{\mathcal{J}}_i > 0)|\mathcal{D}_{k,T}] \leq \textstyle\sum_{i=k+k^*+1}^n E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}). \quad \text{(A.10)}$$

Also, $E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}) = E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}, \mathcal{T}_k) \Pr(\mathcal{T}_k|\mathcal{D}_{k,T}) + E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}, \mathcal{T}_k^c) \Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}) \leq E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}, \mathcal{T}_k) + \Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T})$, since $E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}, \mathcal{T}_k^c) \leq 1$. Hence $B_{n,T} \leq \sum_{i=k+k^*+1}^n E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}, \mathcal{T}_k) + (n-k-k^*) \Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T})$. Consider now the first term of the above and note that

$$\textstyle\sum_{i=k+k^*+1}^n E(\widehat{\mathcal{J}}_i\,|\mathcal{D}_{k,T}, \mathcal{T}_k) = \sum_{i=k+k^*+1}^n \Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta) \left|\theta_{i,(1)} = 0, \mathcal{D}_{k,T}, \mathcal{T}_k\right.] + \sum_{i=k+k^*+1}^n \sum_{j=2}^k \Pr[|t_{\hat{\phi}_{i,(j)}}| > c_p(n,\delta^*) \left|\theta_{i,(j)} = 0, \mathcal{D}_{k,T}, \mathcal{T}_k\right.],$$

where we have made use of the fact that the net effect coefficients, $\theta_{i,(j)}$, of noise variables are zero for $i = k+k^*+1, k+k^*+2, ..., n$ and all j. Also by (B.51) of Lemma A10 and result (ii) of Lemma A2, we have

$$\sum_{i=k+k^*+1}^n \Pr\left(|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta) \left|\theta_{i,(1)} = 0, \mathcal{D}_{k,T}, \mathcal{T}_k\right.\right) + \sum_{i=k+k^*+1}^n \sum_{s=2}^k \Pr\left(|t_{\hat{\phi}_{i,(s)}}| > c_p(n,\delta^*) \left|\theta_{i,(s)} = 0, \mathcal{D}_{k,T}, \mathcal{T}_k\right.\right)$$
$$\leq (n-k-k^*) \exp\left[-\varkappa c_p^2(n,\delta)/2\right] + (k-1)(n-k-k^*) \exp\left[-\varkappa c_p^2(n,\delta^*)/2\right] + O\left[n \exp\left(-C_0 T^{C_1}\right)\right]$$
$$= O\left(n^{1-\varkappa\delta}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left[n \exp\left(-C_0 T^{C_1}\right)\right].$$

Further, by (B.92), $n \Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}) = O\left(n^{2-\varkappa\delta^*}\right) + O\left[n^2 \exp\left(-C_0 T^{C_1}\right)\right]$, giving, overall,

$$B_{n,T} = O\left(n^{1-\delta\varkappa}\right) + O\left(n^{2-\delta^*\varkappa}\right) + O\left[n^2 \exp\left(-C_0 T^{C_1}\right)\right], \quad \text{(A.11)}$$

where we used that $O\left[n\exp\left(-C_0T^{C_1}\right)\right]$ is dominated by $O\left[n^2\exp\left(-C_0T^{C_1}\right)\right]$, and $O\left(n^{1-\varkappa\delta^*}\right)$ is dominated by $O\left(n^{1-\varkappa\delta}\right)$ for $\delta^*>\delta>0$. Substituting for $A_{n,T}$ and $B_{n,T}$ from (A.9) and (A.11) in (A.5) and using (A.4) we obtain $\Pr(\mathcal{A}_0^c)\leq O\left(n^{1-\delta\varkappa}\right)+O\left(n^{2-\delta^*\varkappa}\right)+O\left[n^2\exp\left(-C_0T^{C_1}\right)\right]+\Pr(\mathcal{D}_{k,T}^c)$, where $\Pr(\mathcal{D}_{k,T}^c)$ is already given by (A.3), and $k\exp\left(-C_0T^{C_1}\right)$ is dominated by $O\left[n^2\exp\left(-C_0T^{C_1}\right)\right]$. Hence, noting that $\Pr\left(\mathcal{A}_0\right)=1-\Pr(\mathcal{A}_0^c)$, then

$$\Pr\left(\mathcal{A}_0\right)=1+O\left(n^{1-\delta\varkappa}\right)+O\left(n^{2-\delta^*\varkappa}\right)+O\left(n^{1-\kappa_1/3-\varkappa\delta}\right)+O\left[n^2\exp\left(-C_0T^{C_1}\right)\right], \tag{A.12}$$

since $O[n\exp\left(-C_0T^{C_1}\right)]$ is dominated by $O[n^2\exp\left(-C_0T^{C_1}\right)]$, and $O(n^{1-\kappa_1/3-\varkappa\delta^*})$ is dominated by $O(n^{1-\kappa_1/3-\varkappa\delta})$, for $\delta^*>\delta>0$. Result (30) now follows noting that $T=\ominus\left(n^{\kappa_1}\right)$ and that $O\left[n^2\exp\left(-C_0n^{C_1\kappa_1}\right)\right]=O\left[\exp\left(-n^{C_2\kappa_1}\right)\right]$ for some C_2 in $0<C_2<C_1$. If, in addition, $\delta>1$, and $\delta^*>2$, then $\Pr\left(\mathcal{A}_0\right)\to 1$, as $n,T\to\infty$, for any $\kappa_1>0$.

We establish result (32) next, before establishing results (31) and the result on FDR. Consider $FPR_{n,T}$ defined by (20), and note that the probability of noise or pseudo-signal variable i being selected in any stages of the OCMT procedure is given by $\Pr\left(\mathcal{L}_{i,n}\right)$, for $i=k+1,k+2,...,n$. Then

$$E\left|FPR_{n,T}\right|=\frac{\sum_{i=k+1}^{n}\Pr\left(\mathcal{L}_{i,n}\right)}{n-k}=\frac{\sum_{i=k+1}^{k+k^*}\Pr\left(\mathcal{L}_{i,n}\right)}{n-k}+\frac{\sum_{i=k+k^*+1}^{n}\Pr\left(\mathcal{L}_{i,n}\right)}{n-k}. \tag{A.13}$$

Since $\sum_{i=k+1}^{k+k^*}\Pr\left(\mathcal{L}_{i,n}\right)\leq k^*$ then

$$E\left|FPR_{n,T}\right|\leq (n-k)^{-1}k^*+(n-k)^{-1}\sum_{i=k+k^*+1}^{n}\Pr\left(\mathcal{L}_{i,n}\right). \tag{A.14}$$

Note that

$$(n-k)^{-1}\sum_{i=k+k^*+1}^{n}\Pr\left(\mathcal{L}_{i,n}\right)\leq (n-k)^{-1}\sum_{i=k+k^*+1}^{n}\Pr\left(\mathcal{L}_{i,n}|\mathcal{D}_{k,T}\right)+\Pr\left(\mathcal{D}_{k,T}^c\right). \tag{A.15}$$

Furthermore

$$\Pr\left(\mathcal{L}_{i,n}|\mathcal{D}_{k,T}\right)\leq\Pr\left(\mathcal{L}_{i,n}|\mathcal{D}_{k,T},\mathcal{T}_k\right)+\Pr\left(\mathcal{T}_k^c\right). \tag{A.16}$$

An upper bound to $\Pr\left(\mathcal{T}_k^c\right)=\Pr(\hat{P}>k)$ is established in the first part of this proof, see (A.2). We focus on $\Pr\left(\mathcal{L}_{i,n}|\mathcal{D}_{k,T},\mathcal{T}_k\right)$ next. Due to the conditioning on the event $\mathcal{T}_k$, we have $\Pr\left(\mathcal{L}_{i,n}|\mathcal{D}_{k,T},\mathcal{T}_k\right)=\Pr\left(\mathcal{L}_{i,k}|,\mathcal{D}_{k,T},\mathcal{T}_k\right)$, and in view of $\mathcal{L}_{i,k}=\cup_{h=1}^{k}\mathcal{B}_{i,h}$ we obtain

$$\Pr\left(\mathcal{L}_{i,k}|\mathcal{D}_{k,T},\mathcal{T}_k\right)\leq\sum_{s=1}^{k}\Pr\left(\mathcal{B}_{i,s}|\theta_{i,(s)}=0,\mathcal{D}_{k,T},\mathcal{T}_k\right),\ \text{for } i>k+k^*, \tag{A.17}$$

where we note that $\Pr\left(\mathcal{B}_{i,s}|\mathcal{D}_{k,T},\mathcal{T}_k\right)=\Pr\left(\mathcal{B}_{i,s}|\theta_{i,(s)}=0,\mathcal{D}_{k,T},\mathcal{T}_k\right)$, for $i>k+k^*$ since the net effect coefficients of the noise variables at any stage of OCMT are zero. Further, using (B.51) of Lemma A10, for $i=k+k^*+1,k+k^*+2,...,n$, we have

$$\Pr\left(\mathcal{B}_{i,s}|\theta_{i,(s)}=0,\mathcal{D}_{k,T},\mathcal{T}_k\right)=\begin{cases} O\left\{\exp\left[-\varkappa c_p^2(n,\delta)/2\right]\right\}+O\left[\exp(-C_0T^{C_1})\right], & s=1 \\ O\left\{\exp\left[-\varkappa c_p^2(n,\delta^*)/2\right]\right\}+O\left[\exp(-C_0T^{C_1})\right], & s>1\end{cases}, \tag{A.18}$$

where $\varkappa = [(1-\pi)/(1+d_T)]^2$. Clearly $0 < \varkappa < 1$, since $0 < \pi < 1$, and d_T is a bounded positive sequence. Hence, given result (ii) of Lemma A2 in the online supplement, for $i = k+k^*+1, k+k^*+2, ..., n$, we have

$$\sum_{s=1}^{k} \Pr\left(\mathcal{B}_{i,s}|\theta_{i,(s)} = 0, \mathcal{D}_{k,T}, \mathcal{T}_k\right) = O\left(n^{-\delta\varkappa}\right) + O\left(n^{-\delta^*\varkappa}\right) + O\left[\exp(-C_0 T^{C_1})\right].$$

Using this result in (A.17) and averaging across $i = k+k^*+1, k+k^*+2, ..., n$, we obtain

$$(n-k)^{-1}\sum_{i=k+k^*+1}^{n} \Pr\left(\mathcal{L}_{i,k}|\mathcal{D}_{k,T}, \mathcal{T}_k\right) = O\left(n^{-\varkappa\delta}\right) + O\left(n^{-\varkappa\delta^*}\right) + O\left[\exp(-C_0 T^{C_1})\right]. \quad \text{(A.19)}$$

Overall, with $\delta^* > \delta$, $T = \ominus(n^{\kappa_1})$, $k^* = \ominus(n^{\epsilon})$, and using (A.2), (A.3), (A.14)-(A.16) and (A.19), we have $E|FPR_{n,T}| = k^*/(n-k) + O\left(n^{-\varkappa\delta}\right) + O\left(n^{-\varkappa\delta^*}\right) + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left(n^{1-\kappa_1/3-\varkappa\delta^*}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left[\exp(-C_0 n^{C_1\kappa_1})\right] + O\left(n^{\epsilon-1}\right) + O\left[n\exp\left(-C_0 n^{C_1\kappa_1}\right)\right]$. But $O\left[\exp(-C_0 n^{C_1\kappa_1})\right]$ and $O\left[n\exp\left(-C_0 n^{C_1\kappa_1}\right)\right]$ are dominated by $\left[\exp\left(-n^{C_2\kappa_1}\right)\right]$ for some $0 < C_2 < C_1$. In addition, since $\delta^* > \delta$ and $\varkappa$ is positive, the terms $O\left(n^{-\varkappa\delta^*}\right)$ and $O\left(n^{1-\kappa_1/3-\varkappa\delta^*}\right)$ are dominated by $O\left(n^{-\varkappa\delta}\right)$ and $O\left(n^{1-\kappa_1/3-\varkappa\delta}\right)$, respectively. Hence, $E|FPR_{n,T}| = k^*/(n-k) + O\left(n^{-\varkappa\delta}\right) + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left(n^{\epsilon-1}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left[\exp\left(-n^{C_2\kappa_1}\right)\right]$, for some $C_2 > 0$, which completes the proof of (32).

To establish (31) we note from (20) that

$$E|TPR_{n,T}| = k^{-1}\sum_{i=1}^{k} \Pr[\widehat{\mathcal{J}}_i = 1]. \quad \text{(A.20)}$$

But $\Pr[\widehat{\mathcal{J}}_i = 1] = 1 - \Pr[\widehat{\mathcal{J}}_i = 0]$, and $\Pr[\widehat{\mathcal{J}}_i = 0] \leq \Pr[\widehat{\mathcal{J}}_i = 0|\mathcal{D}_{k,T}] + \Pr\left(\mathcal{D}_{k,T}^c\right)$. Using (A.8) and (A.3) and dropping the terms $O\left[\exp\left(-C_0 T^{C_1}\right)\right]$ and $O\left(n^{1-\kappa_1/3-\varkappa\delta^*}\right)$ that are dominated by $O\left[n\exp\left(-C_0 T^{C_1}\right)\right]$ and $O\left(n^{1-\kappa_1/3-\varkappa\delta}\right)$, respectively (noting that $\delta^* > \delta > 0$) we obtain $\Pr[\widehat{\mathcal{J}}_i = 0] = O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left[n\exp\left(-C_0 T^{C_1}\right)\right]$, for $i = 1, 2, ..., k$. Hence, $\sum_{i=1}^{k}\Pr[\widehat{\mathcal{J}}_i = 1] = k + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left[n\exp\left(-C_0 T^{C_1}\right)\right]$, which, after substituting this expression in (A.20), and noting that $T = \ominus(n^{\kappa_1})$, and
$O\left[n\exp\left(-C_0 n^{C_1\kappa_1}\right)\right] = O\left[\exp\left(-n^{C_2\kappa_1}\right)\right]$, for some C_2 in $0 < C_2 < C_1$ yields

$$E|TPR_{n,T}| = 1 + O\left(n^{1-\kappa_1/3-\varkappa\delta}\right) + O\left[\exp\left(-n^{C_2\kappa_1}\right)\right], \quad \text{(A.21)}$$

for some $C_2 > 0$, as required.

To establish the result on FDR, we first note that

$$FDR_{n,T} = \frac{\sum_{i=1}^{n} I\left(\widehat{\mathcal{J}}_i = 1, \text{ and } \beta_i = \theta_i = 0\right)}{(n-k)\,FPR_{n,T} + kTPR_{n,T} + 1}.$$

Consider the numerator first. Taking expectations $E\sum_{i=1}^{n} I[\widehat{\mathcal{J}}_i = 1, \text{ and } \beta_i = \theta_i = 0] = \sum_{i=k+k^*+1}^{n}\Pr\left(\mathcal{L}_{i,n}\right)$. Using (A.2),(A.3),(A.15), and (A.16), and noting $T = \ominus(n^{\kappa_1})$, we have

$$\sum_{i=k+k^*+1}^{n}\Pr\left(\mathcal{L}_{i,n}\right) = O\left(n^{1-\varkappa\delta}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left(n^{2-\kappa_1/3-\varkappa\delta}\right) + O\left(n^{2-\kappa_1/3-\varkappa\delta^*}\right) + O\left(n^{2-\varkappa\delta^*}\right) + O\left[n\exp(-C_0 n^{C_1\kappa_1})\right] + O\left[n^2\exp\left(-C_0 n^{C_1\kappa_1}\right)\right], \quad \text{(A.22)}$$

for some $C_0, C_1 > 0$. Hence, if $\delta > \max\{1, 2-\kappa_1/3\}$, and $\delta^* > 2$, then $\sum_{i=k+k^*+1}^{n} \Pr(\mathcal{L}_{i,n}) \to 0$, and

$$\textstyle\sum_{i=1}^{n} I[\widehat{\mathcal{J}}_i = 1, \text{ and } \beta_i = \theta_i = 0] \to_p 0. \tag{A.23}$$

Consider the term $kTPR_{n,T}$ in the denominator next. Using (A.21), we have

$$kTPR_{n,T} \to_p k, \tag{A.24}$$

if $\delta > 1-\kappa_1/3$. Using (A.23), (A.24), and noting that $(n-k)\,FPR_{n,T} \geq 0$, we have $FDR_{n,T} \to_p 0$, if $\delta > \max\{1, 2-\kappa_1/3\}$, and $\delta^* > 2$, as required.

A.2.3 Proof of Theorem 2

We prove the error norm result first. Define a sequence $r_{\tilde{u},n}$ such that $r_{\tilde{u},n} = O(n^{3\epsilon - 3\kappa_1/2}) + O\left(n^{-\kappa_1/2}\right)$. By the definition of convergence in probability, we need to show that, for any $\varepsilon > 0$, there exists some $B_\varepsilon < \infty$, such that $\Pr\left(r_{\tilde{u},n}^{-1}\left|F_{\tilde{u}} - \sigma^2\right| > B_\varepsilon\right) < \varepsilon$. We have $\Pr\left(r_{\tilde{u},n}^{-1}\left|F_{\tilde{u}} - \sigma^2\right| > B_\varepsilon\right) \leq \Pr\left(r_{\tilde{u},n}^{-1}\left|F_{\tilde{u}} - \sigma^2\right| > B_\varepsilon | \mathcal{A}_0\right) + \Pr\left(\mathcal{A}_0^c\right)$. By (A.12), $\lim_{n\to\infty}\Pr\left(\mathcal{A}_0^c\right) = 0$. Then, it is sufficient to show that, for any $\varepsilon > 0$, there exists some $B_\varepsilon < \infty$, such that $\Pr\left(r_{\tilde{u},n}^{-1}\left|F_{\tilde{u}} - \sigma^2\right| > B_\varepsilon | \mathcal{A}_0\right) < \varepsilon$. But, by (B.95) of Lemma A21 in the online supplement, the desired result follows immediately.

To prove the result for the coefficient norm, we proceed similarly. Recall that $k^* = \ominus(n^\epsilon)$ and define a sequence $r_{\beta,n}$, such that $r_{\beta,n} = O(n^{5\epsilon/2 - \kappa_1}) + O(n^{\epsilon - \kappa_1/2})$. To establish $\left\|\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\right\| = O_p(r_{\beta,n})$, we need to show that, for any $\varepsilon > 0$, there exists some $B_\varepsilon < \infty$, such that $\Pr(r_{\beta,n}^{-1}\left\|\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\right\| > B_\varepsilon) < \varepsilon$. We have $\Pr(r_{\beta,n}^{-1}\left\|\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\right\| > B_\varepsilon) \leq \Pr(r_{\beta,n}^{-1}\left\|\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\right\| > B_\varepsilon | \mathcal{A}_0) + \Pr\left(\mathcal{A}_0^c\right)$. Again, by (A.12), $\lim_{n\to\infty}\Pr\left(\mathcal{A}_0^c\right) = 0$. Then, it is sufficient to show that, for any $\varepsilon > 0$, there exists some $B_\varepsilon < \infty$, such that $\Pr(r_{\beta,n}^{-1}\left\|\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\right\| > B_\varepsilon | \mathcal{A}_0) < \varepsilon$. But this follows immediately from (B.96) of Lemma A21 in the online supplement, since, conditional on the event $\mathcal{A}_0$, the set of selected regressors includes all signals.

A.2.4 Proof of Theorem 3

See Section B of the online supplement.

A.2.5 Proofs of results for the single stage OCMT in the absence of hidden signals

Result (37) follows from (25), and (38) follows from the analysis preceding Theorem 1, using (26) and (27). The result on $FDR_{n,T}$ continues to hold using the same arguments as in the proof of Theorem 1. To obtain $\Pr(\mathcal{A}_0)$ we follow the derivations in the proof of the multi-stage version of OCMT provided in Section A.2.2, but note that we only need to consider the terms from the first stage of OCMT. Similarly to (A.5) and without the need to condition on $\mathcal{D}_{k,T}$, we have $\Pr(\mathcal{A}_0^c) \leq \Pr(\sum_{i=1}^{k}\widehat{\mathcal{J}}_i < k) + \Pr(\sum_{i=k+k^*+1}^{n}\widehat{\mathcal{J}}_i > 0) = A_{n,T} + B_{n,T}$, noting that $\widehat{\mathcal{J}}_i = \widehat{\mathcal{J}}_{i,(1)}$. Also, as with (A.9) and (A.10), we have $A_{n,T} \leq k\exp\left(-C_1 T^{C_2}\right)$. Similarly, for $B_{n,T}$ we first note that

$$B_{n,T} \leq \sum_{i=k+k^*+1}^{n} E(\widehat{\mathcal{J}}_{i,(1)}\,|\beta_i = 0) = \sum_{i=k+k^*+1}^{n} \Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p\,(n,\delta)\,|\theta_i = 0],$$

which, by (B.51) of Lemma A10 in the online supplement, yields $B_{n,T} \leq (n-k-k^*) \exp\left[-\varkappa c_p^2(n,\delta)/2\right]+ O\left[n \exp\left(-C_0 T^{C_1}\right)\right]$, or upon using result (ii) of Lemma A2, $\Pr\left(\mathcal{A}_0^c\right) \leq A_{n,T}+B_{n,T} \leq O\left(n^{1-\delta\varkappa}\right)+ O\left[n \exp\left(-C_0 T^{C_1}\right)\right]$, and hence $\Pr\left(\mathcal{A}_0\right) = O\left(n^{1-\delta\varkappa}\right) + O\left[\exp\left(-n^{C_2}\right)\right]$, for some $C_2 > 0$. If, in addition, $\delta > 1$, then $\Pr\left(\mathcal{A}_0\right) \to 1$, as n,$T \to \infty$, such that $T = O\left(n^{\kappa_1}\right)$ for some $\kappa_1 > 0$, as required.

References

Antoniadis, A. and J. Fan (2001). Regularization of wavelets approximations. *Journal of the American Statistical Association 96*, 939–967.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*, 1133–1150.

Belloni, A., V. Chernozhukov, and C. Hansen (2014a). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*, 29–50.

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

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the American Statistical Association 57*, 289–300.

Bickel, J. P., Y. Ritov, and A. Tsybakov (2009). Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics 37*, 1705–1732.

Buhlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics 34*(2), 599–583.

Buhlmann, P. and S. van de Geer (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer.

Candes, E. and T. Tao (2007). The dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics 35*, 2313–2404.

Donoho, D. and M. Elad (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. *Proceedings of the National Academy of Sciences 100*, 2197–2202.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Annals of Statistics 32*, 407–499.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348–1360.

Fan, J. and J. Lv (2008). Sure independence screening for ultra-high dimensional feature space. *Journal of Royal Statistical Society B 70*, 849–911.

Fan, J. and J. Lv (2013). Asymptotic equivalence of regularization methods in thresholded parameter space. *Journal of the American Statistical Association 108*, 1044–1061.

Fan, J., R. Samworth, and Y. Wu (2009). Ultra high dimensional variable selection: Beyond the linear model. *Journal of Machine Learning Research 10*, 1829–1853.

Fan, J. and R. Song (2010). Sure independence screening in generalized linear models with np-dimensionality. *Annals of Statistics 38*, 3567–3604.

Fan, J. and C. Tang (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society Series B 75*, 531–552.

Fithian, W., D. Sun, and J. Taylor (2014). Optimal inference after model selection. *arXiv:1410.2597v4*.

Fithian, W., J. Taylor, R. J. Tibshirani, and R. Tibshirani (2015). Selective sequential model selection. *arXiv:1512.02565*.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics 29*, 1189–1232.

Friedman, J., T. Hastie, and R. Tibshirani (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics 28*, 337–374.

Gavrilov, Y., Y. Benjamini, and S. K. Sarkar (2009). An adaptive step-down procedure with proven fdr control under independence. *Annals of Statistics 37*, 619–629.

G'Sell, M. G., S. Wager, A. Chouldechova, and R. Tibshirani (2016). Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society: Series B 78*, 423–444.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6*, 65–70.

Huang, J., J. Horowitz, and S. Ma (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics 36*, 587–613.

Leeb, H. and B. M. Pötscher (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Annals of Statistics 142*, 2554–2591.

Leeb, H. and B. M. Pötscher (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory 24*, 338–376.

Li, A. and R. Barber (2015). Accumulation tests for fdr control in ordered hypothesis testing. *arXiv:1505.07352*.

Lv, J. and Y. Fan (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Annals of Statistics 37*, 3498–3528.

Pesaran, M. H. and R. P. Smith (2014). Signs of impact effects in time series regression models. *Economics Letters 122*, 150–153.

Stock, J. H. and M. W. Watson (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics 30*, 481–493.

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

Tibshirani, R. J., J. Taylor, R. Lockhart, and R. Tibshirani (2014). Exact post-selection inference for sequential regression procedures. *arXiv:1401.3889*.

Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics 38*, 894–942.

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

Zhou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B 67*, 301–320.

Table A.1: Summary of Monte Carlo results for experiments with Gaussian innovations

	DGP-I				DGP-II				DGP-III				DGP-IV				DGP-V			
	Oracle	Lasso	A-Lasso	OCMT	Oracle	Lasso	A-Lasso	OCMT	Oracle	Lasso	A-Lasso	OCMT	Oracle	Lasso	A-Lasso	OCMT	Oracle*	Lasso	A-Lasso	OCMT
Static Specifications																				
TPR	1.000	0.962	0.883	0.964	1.000	0.966	0.907	0.958	1.000	0.948	0.922	0.899	1.000	0.929	0.887	0.873	1.000	0.326	0.211	0.275
FPR	0.000	0.039	0.013	0.000	0.000	0.044	0.015	0.007	0.000	0.098	0.033	0.000	0.000	0.085	0.030	0.007	0.000	0.037	0.012	0.000
FDR* (true model)	0.000	0.473	0.187	0.003	0.000	0.509	0.213	0.174	0.000	0.723	0.370	0.003	0.000	0.683	0.348	0.177	0.000	0.459	0.185	0.003
FDR (approximating model)	0.000	0.473	0.187	0.003	0.000	0.473	0.198	0.002	0.000	0.723	0.370	0.003	0.000	0.651	0.331	0.003	0.000	0.459	0.185	0.003
RMSFE	3.376	3.457	3.484	3.393	3.243	3.331	3.358	3.268	2.080	2.219	2.212	2.139	2.210	2.336	2.340	2.273	1.329	1.332	1.342	1.307
$RMSE_{\tilde{\beta}}$	0.639	0.824	1.143	0.693	0.550	0.786	1.022	0.707	0.356	0.995	0.863	0.601	0.373	0.958	0.912	0.703	0.382	0.275	0.342	0.219
Dynamic Specifications																				
Experiments with $\varphi = 0.4$																				
TPR	1.000	0.967	0.907	0.940	1.000	0.972	0.927	0.932	1.000	0.960	0.936	0.873	1.000	0.945	0.912	0.856	1.000	0.400	0.298	0.312
FPR	0.000	0.053	0.017	0.002	0.000	0.059	0.020	0.008	0.000	0.108	0.038	0.001	0.000	0.097	0.036	0.008	0.000	0.053	0.019	0.001
FDR* (true model)	0.000	0.532	0.222	0.041	0.000	0.563	0.248	0.158	0.000	0.721	0.373	0.026	0.000	0.692	0.363	0.157	0.000	0.525	0.237	0.028
FDR (approximating model)	0.000	0.518	0.217	0.013	0.000	0.518	0.229	0.002	0.000	0.705	0.365	0.003	0.000	0.647	0.340	0.003	0.000	0.512	0.232	0.003
RMSFE	3.386	3.530	3.538	3.466	3.253	3.392	3.402	3.336	2.087	2.255	2.243	2.201	2.216	2.373	2.370	2.331	1.333	1.356	1.361	1.332
$RMSE_{\tilde{\beta}}$	0.646	0.878	1.206	0.768	0.552	0.821	1.073	0.746	0.361	0.997	0.896	0.681	0.378	0.962	0.940	0.761	0.386	0.307	0.376	0.248
Experiments with $\varphi = 0.8$																				
TPR	1.000	0.962	0.881	0.868	1.000	0.968	0.896	0.847	1.000	0.946	0.887	0.810	1.000	0.935	0.874	0.786	1.000	0.394	0.282	0.250
FPR	0.000	0.049	0.016	0.028	0.000	0.054	0.018	0.021	0.000	0.101	0.031	0.017	0.000	0.090	0.027	0.020	0.000	0.050	0.013	0.017
FDR* (true model)	0.000	0.499	0.232	0.376	0.000	0.526	0.255	0.380	0.000	0.674	0.366	0.352	0.000	0.649	0.342	0.392	0.000	0.496	0.195	0.397
FDR (approximating model)	0.000	0.473	0.223	0.064	0.000	0.471	0.231	0.002	0.000	0.653	0.357	0.002	0.000	0.598	0.318	0.002	0.000	0.470	0.188	0.002
RMSFE	3.390	3.574	3.585	3.578	3.255	3.430	3.445	3.387	2.091	2.289	2.263	2.254	2.219	2.406	2.389	2.374	1.334	1.371	1.367	1.337
$RMSE_{\tilde{\beta}}$	0.645	0.877	1.130	1.313	0.551	0.819	1.010	0.993	0.360	1.030	0.900	0.929	0.376	0.985	0.906	0.971	0.383	0.303	0.366	0.294

Notes: The reported statistics represent averages across R^2 (low, medium and high), the sample sizes ($n = 100, 200, 300$ and $T = 100, 300, 500$) and all DGPs in a given design. This gives 135, 54, 27, 54 and 27 experiments for DGP-I to V, respectively. DGPs I-IV are given by (39) and DGP V is given by (40). See also notes to Table 1.

Table A.2: Summary of Monte Carlo results for experiments with non-Gaussian innovations

	DGP-I				DGP-II				DGP-III				DGP-IV				DGP-V			
	Oracle	Lasso	A-Lasso	OCMT	Oracle	Lasso	A-Lasso	OCMT	Oracle	Lasso	A-Lasso	OCMT	Oracle	Lasso	A-Lasso	OCMT	Oracle*	Lasso	A-Lasso	OCMT
Static Specifications																				
TPR	1.000	0.961	0.877	0.959	1.000	0.965	0.902	0.955	1.000	0.945	0.917	0.896	1.000	0.925	0.880	0.870	1.000	0.324	0.211	0.275
FPR	0.000	0.038	0.011	0.000	0.000	0.042	0.013	0.008	0.000	0.094	0.027	0.000	0.000	0.082	0.025	0.007	0.000	0.036	0.011	0.000
FDR* (true model)	0.000	0.464	0.182	0.005	0.000	0.503	0.208	0.176	0.000	0.716	0.339	0.005	0.000	0.675	0.322	0.179	0.000	0.456	0.180	0.007
FDR (approximating model)	0.000	0.464	0.182	0.005	0.000	0.467	0.192	0.004	0.000	0.716	0.339	0.005	0.000	0.643	0.304	0.005	0.000	0.456	0.180	0.007
RMSFE	3.376	3.460	3.480	3.400	3.243	3.333	3.352	3.274	2.081	2.223	2.205	2.145	2.209	2.337	2.331	2.276	1.330	1.334	1.339	1.310
RMSE$_{\tilde{\beta}}$	0.648	0.833	1.138	0.716	0.558	0.798	1.017	0.733	0.362	1.007	0.853	0.624	0.379	0.970	0.904	0.723	0.388	0.281	0.332	0.232
Dynamic Specifications																				
Experiments with $\varphi = 0.4$																				
TPR	1.000	0.966	0.903	0.936	1.000	0.971	0.923	0.930	1.000	0.957	0.931	0.871	1.000	0.943	0.908	0.856	1.000	0.399	0.296	0.313
FPR	0.000	0.052	0.016	0.002	0.000	0.057	0.018	0.008	0.000	0.105	0.032	0.001	0.000	0.094	0.031	0.008	0.000	0.052	0.017	0.001
FDR* (true model)	0.000	0.528	0.215	0.042	0.000	0.557	0.239	0.159	0.000	0.714	0.341	0.026	0.000	0.686	0.338	0.159	0.000	0.521	0.228	0.030
FDR (approximating model)	0.000	0.513	0.210	0.014	0.000	0.512	0.219	0.003	0.000	0.699	0.334	0.004	0.000	0.642	0.315	0.004	0.000	0.508	0.223	0.005
RMSFE	3.386	3.532	3.532	3.473	3.255	3.395	3.397	3.342	2.089	2.260	2.234	2.207	2.217	2.376	2.362	2.334	1.334	1.356	1.356	1.334
RMSE$_{\tilde{\beta}}$	0.653	0.887	1.199	0.789	0.563	0.833	1.066	0.770	0.367	1.009	0.882	0.702	0.383	0.970	0.927	0.775	0.393	0.311	0.366	0.258
Experiments with $\varphi = 0.8$																				
TPR	1.000	0.962	0.872	0.867	1.000	0.967	0.889	0.846	1.000	0.945	0.884	0.812	1.000	0.933	0.869	0.786	1.000	0.390	0.280	0.251
FPR	0.000	0.048	0.015	0.028	0.000	0.053	0.017	0.021	0.000	0.098	0.029	0.017	0.000	0.088	0.026	0.020	0.000	0.048	0.012	0.017
FDR* (true model)	0.000	0.492	0.221	0.376	0.000	0.521	0.247	0.382	0.000	0.669	0.356	0.351	0.000	0.645	0.333	0.393	0.000	0.491	0.187	0.397
FDR (approximating model)	0.000	0.466	0.213	0.064	0.000	0.465	0.223	0.002	0.000	0.647	0.347	0.002	0.000	0.594	0.308	0.002	0.000	0.465	0.181	0.003
RMSFE	3.391	3.578	3.594	3.584	3.256	3.431	3.451	3.393	2.091	2.290	2.265	2.255	2.218	2.406	2.389	2.373	1.335	1.371	1.366	1.339
RMSE$_{\tilde{\beta}}$	0.654	0.885	1.146	1.328	0.560	0.832	1.031	1.007	0.366	1.040	0.916	0.933	0.383	0.996	0.922	0.981	0.391	0.308	0.366	0.302

Notes: See notes to Tables 1 and A.1.

Online Theory Supplement to

"A One-Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models"

A. Chudik
Federal Reserve Bank of Dallas

G. Kapetanios
King's College, London

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

This online theory supplement is organised as follows: Section A provides lemmas for the Appendix of the main paper. Section B provides a proof of Theorem 3. Section C provides a discussion of various results related to the case where both signal and noise variables are mixing processes. Section D presents lemmas for regressions with covariates that are mixing processors. Section E provides lemmas for the case where the regressors are deterministic, while Section F provides some further supplementary lemmas needed for Sections B and C of this supplement.

A. Lemmas

Before presenting the lemmas and their proofs we give an outline of their use. Lemmas A1 and A2 are technical auxiliary lemmas. Lemmas A3-A5 provide extensions to existing results in the literature that form the building blocks for our exponential probability inequalities. Lemmas A6 and A7 provide exponential probability inequalities for squares and cross-products of sums of random variables. Lemmas A8 and A9 provide results that help handle the denominator of a t-statistic in the context of exponential inequalities. Lemma A10 is a key lemma that provides exponential inequalities for t-statistics. Lemmas A11-A21 are further auxiliary lemmas.

Lemma A1 *Let y_t, for $t = 1, 2, ..., T$, be given by DGP (6) and define $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$, for $i = 1, 2, ..., k$, and $\mathbf{X}_k = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$, and suppose that Assumption 1 holds. Moreover, let $\mathbf{q}_{i\cdot} = (q_{i1}, q_{i2}, ..., q_{iT})'$, for $i = 1, 2, ..., l_T$, $\mathbf{Q} = (\mathbf{q}_{1\cdot}, \mathbf{q}_{2\cdot}, ..., \mathbf{q}_{l_T\cdot})'$, and assume $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$ exists. Further, assume that $\boldsymbol{\tau}_T = (1, 1, ..., 1)'$ is included in $\mathbf{Q}$, a $(0 \leq a < k)$ column vectors of $\mathbf{X}_k$ belong to $\mathbf{Q}$, and the remaining $b = k - 1 > 0$ columns of $\mathbf{X}_k$ that do not belong in $\mathbf{Q}$ are collected in the $T \times b$ matrix $\mathbf{X}_b$. The slope coefficients that correspond to regressors in $\mathbf{X}_b$ are collected in the $b \times 1$ vector $\boldsymbol{\beta}_{b,T}$. Define $\boldsymbol{\theta}_{b,T} = \boldsymbol{\Omega}_{b,T}\boldsymbol{\beta}_{b,T}$, where $\boldsymbol{\Omega}_{b,T} = E(T^{-1}\mathbf{X}_b'\mathbf{M}_q\mathbf{X}_b)$. If $\boldsymbol{\Omega}_{b,T}$ is nonsingular, and $\boldsymbol{\beta}_{k,T} = (\beta_{1,T}, \beta_{2,T}, ..., \beta_{k,T})' \neq \mathbf{0}$, then at least one element of the $b \times 1$ vector $\boldsymbol{\theta}_{b,T}$ is nonzero.*

Proof. Since $\boldsymbol{\Omega}_{b,T}$ is nonsingular and $\boldsymbol{\beta}_{b,T} \neq \mathbf{0}$, it follows that $\boldsymbol{\theta}_{b,T} \neq \mathbf{0}$; otherwise $\boldsymbol{\beta}_{b,T} = \boldsymbol{\Omega}_{b,T}^{-1}\boldsymbol{\theta}_{b,T} = \mathbf{0}$, which contradicts the assumption that $\boldsymbol{\beta}_{b,T} \neq \mathbf{0}$. ■

Lemma A2 *Consider the critical value function* $c_p(n,\delta)$ *defined by (15), with* $0 < p < 1$ *and* $f(n,\delta) = cn^{\delta}$*, for some* $c, \delta > 0$*. Moreover, let* $a > 0$ *and* $0 < b \leq 1$*. Then: (i)* $c_p(n,\delta) = O\left([\delta \ln(n)]^{1/2}\right)$*, (ii)* $n^a \exp\left[-bc_p^2(n,\delta)\right] = \ominus\left(n^{a-2b\delta}\right)$*.*

Proof. Results follow from Lemma 3 of the Supplementary Appendix A of Bailey et al. (2018). ■

Lemma A3 *Let* z_t *be a martingale difference sequence with respect to the filtration* $\mathcal{F}_{t-1}^z = \sigma\left(\{z_s\}_{s=1}^{t-1}\right)$*, and suppose that there exist finite positive constants* C_0 *and* C_1*, and* $s > 0$ *such that* $\sup_t \Pr(|z_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$*, for all* $\alpha > 0$*. Let* $\sigma_{zt}^2 = E(z_t^2 \left| \mathcal{F}_{t-1}^z \right.)$ *and* $\sigma_z^2 = \frac{1}{T}\sum_{t=1}^T \sigma_{zt}^2$*. 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*

$$\Pr(|\textstyle\sum_{t=1}^T z_t| > \zeta_T) \leq \exp[-(1-\pi)^2 \zeta_T^2 T^{-1} \sigma_z^{-2}/2]. \tag{B.1}$$

If $\lambda > (s+1)/(s+2)$*, then for some finite positive constant* C_3*,*

$$\Pr(|\textstyle\sum_{t=1}^T z_t| > \zeta_T) \leq \exp[-C_3 \zeta_T^{s/(s+1)}]. \tag{B.2}$$

Proof. We proceed to prove (B.1) first and then prove (B.2). Decompose z_t as $z_t = w_t + v_t$, where $w_t = z_t I(|z_t| \leq D_T)$ and $v_t = z_t I(|z_t| > D_T)$, and note that

$$\begin{aligned}\Pr\{|\textstyle\sum_{t=1}^T [z_t - E(z_t)]| > \zeta_T\} \leq & \Pr\{|\textstyle\sum_{t=1}^T [w_t - E(w_t)]| > (1-\pi)\zeta_T\} \\ & + \Pr\{|\textstyle\sum_{t=1}^T [v_t - E(v_t)]| > \pi\zeta_T\},\end{aligned} \tag{B.3}$$

for any $0 < \pi < 1$.[1] Further, it is easily verified that $w_t - E(w_t)$ is a martingale difference process, and since $|w_t| \leq D_T$ then by setting $b = T\sigma_z^2$ and $a = (1-\pi)\zeta_T$ in Proposition 2.1 of Freedman (1975), for the first term on the RHS of (B.3) we obtain

$$\Pr\{|\textstyle\sum_{t=1}^T [w_t - E(w_t)]| > (1-\pi)\zeta_T\} \leq \exp\{-\zeta_T^2 \left[T\sigma_z^2 + (1-\pi) D_T \zeta_T\right]^{-1} (1-\pi)^2/2\}.$$

Consider now the second term on the RHS of (B.3) and first note that

$$\Pr\{|\textstyle\sum_{t=1}^T [v_t - E(v_t)]| > \pi\zeta_T\} \leq \Pr[\textstyle\sum_{t=1}^T |v_t - E(v_t)| > \pi\zeta_T], \tag{B.4}$$

and by Markov's inequality,

$$\Pr\{\textstyle\sum_{t=1}^T |[v_t - E(v_t)]| > \pi\zeta_T\} \leq \pi^{-1}\zeta_T^{-1}\textstyle\sum_{t=1}^T E|v_t - E(v_t)| \leq 2\pi^{-1}\zeta_T^{-1}\textstyle\sum_{t=1}^T E|v_t|. \tag{B.5}$$

[1] Let $A_T = \sum_{t=1}^T [z_t - E(z_t)] = B_{1,T} + B_{2,T}$, where $B_{1,T} = \sum_{t=1}^T [w_t - E(w_t)]$ and $B_{2,T} = \sum_{t=1}^T [v_t - E(v_t)]$. We have $|A_T| \leq |B_{1,T}| + |B_{2,T}|$ and, therefore, $\Pr(|A_T| > \zeta_T) \leq \Pr(|B_{1,T}| + |B_{2,T}| > \zeta_T)$. Equation (B.3) now readily follows using the same steps as in the proof of (B.59).

But by Holder's inequality, for any finite $p, q \geq 1$ such that $p^{-1}+q^{-1}=1$ we have $E\left|v_t\right| = E\left(\left|z_t I\left[\left|z_t\right|>D_T\right]\right|\right) \leq \left(E\left|z_t\right|^p\right)^{1/p}\left\{E\left[\left|I\left(\left|z_t\right|>D_T\right)\right|^q\right]\right\}^{1/q} = \left(E\left|z_t\right|^p\right)^{1/p}\left\{E\left[I\left(\left|z_t\right|>D_T\right)\right]\right\}^{1/q}$, and therefore

$$E\left|v_t\right| \leq \left(E\left|z_t\right|^p\right)^{1/p}\left[\Pr\left(\left|z_t\right|>D_T\right)\right]^{1/q}. \tag{B.6}$$

Also, for any finite $p \geq 1$ there exists a finite positive constant C_2 such that $E\left|z_t\right|^p \leq C_2 < \infty$, by Lemma A15. Further, by assumption $\sup_t \Pr\left(\left|z_t\right|>D_T\right) \leq C_0 \exp\left(-C_1 D_T^s\right)$. Using this upper bound in (B.6) together with the upper bound on $E\left|z_t\right|^p$, we have $\sup_t E\left|v_t\right| \leq C_2^{1/p} C_0^{1/q}\left[\exp\left(-C_1 D_T^s\right)\right]^{1/q}$. Therefore, using (B.4)-(B.5), $\Pr\{\left|\sum_{t=1}^T\left[v_t-E(v_t)\right]\right| > \pi\zeta_T\} \leq (2/\pi) C_2^{1/p} C_0^{1/q} \zeta_T^{-1} T\left[\exp\left(-C_1 D_T^s\right)\right]^{1/q}$. We need to determine D_T such that

$$(2/\pi) C_2^{1/p} C_0^{1/q} \zeta_T^{-1} T\left[\exp\left(-C_1 D_T^s\right)\right]^{1/q} \leq \exp\{-\zeta_T^2\left[T\sigma_z^2+(1-\pi) D_T\zeta_T\right]^{-1}(1-\pi)^2/2\}. \tag{B.7}$$

Taking logs, we have $\ln[(2/\pi) C_2^{1/p} C_0^{1/q}]+\ln\left(\zeta_T^{-1}T\right)-\left(C_1/q\right) D_T^s \leq -(1-\pi)^2\zeta_T^2/\{2[T\sigma_z^2+(1-\pi) D_T\zeta_T]\}$, or $C_1 q^{-1} D_T^s \geq \ln[(2/\pi) C_2^{1/p} C_0^{1/q}]+\ln\left(\zeta_T^{-1}T\right)+(1-\pi)^2\zeta_T^2/\left\{2\left[T\sigma_z^2+(1-\pi) D_T\zeta_T\right]\right\}$. Post-multiplying by $2\left[T\sigma_z^2+(1-\pi) D_T\zeta_T\right]>0$ we have

$$\begin{aligned}&\left(2\sigma_z^2 C_1 q^{-1}\right) T D_T^s+\left(2C_1 q^{-1}\right)(1-\pi) D_T^{s+1}\zeta_T-2(1-\pi) D_T\zeta_T\{\ln\left(\zeta_T^{-1}T\right)+\ln[(2/\pi) C_2^{1/p} C_0^{1/q}]\} \\ &\geq 2\sigma_z^2 T\ln[(2/\pi) C_2^{1/p} C_0^{1/q}]+2\sigma_z^2 T\ln\left(\zeta_T^{-1}T\right)+(1-\pi)^2\zeta_T^2.\end{aligned} \tag{B.8}$$

The above expression can now be simplified for values of $T\to\infty$, by dropping the constants and terms that are asymptotically dominated by other terms on the same side of the inequality.[2] Since $\zeta_T=\ominus\left(T^\lambda\right)$, for some $0<\lambda\leq(s+1)/(s+2)$, and considering values of D_T such that $D_T=\ominus\left(T^\psi\right)$, for some $\psi>0$, implies that the third and fourth term on the LHS of (B.8), which have the orders $\ominus\left[\ln(T)T^{\lambda+\psi}\right]$ and $\ominus\left(T^{\lambda+\psi}\right)$, respectively, are dominated by the second term on the LHS of (B.8) which is of order $\ominus\left(T^{\lambda+\psi+s\psi}\right)$. Further the first term on the RHS of (B.8) is dominated by the second term. Note that for $\zeta_T=\ominus\left(T^\lambda\right)$, we have $T\ln\left(\zeta_T^{-1}T\right)=\ominus[T\ln(T)]$, whilst the order of the first term on the RHS of (B.8) is $\ominus(T)$. Result (B.7) follows if we show that there exists D_T such that

$$\left(C_1 q^{-1}\right)\left[2\sigma_z^2 T D_T^s+2(1-\pi) D_T^{s+1}\zeta_T\right] \geq 2\sigma_z^2 T\ln\left(\zeta_T^{-1}T\right)+(1-\pi)^2\zeta_T^2. \tag{B.9}$$

Set $\left(C_1 q^{-1}\right) D_T^{s+1}=(1-\pi)\zeta_T/2$, or $D_T=\left(C_1^{-1}q(1-\pi)\zeta_T/2\right)^{1/(s+1)}$, and note that (B.9) can be written as $2\sigma_z^2\left(C_1 q^{-1}\right) T\left(C_1^{-1}q(1-\pi)\zeta_T/2\right)^{s/(s+1)}+(1-\pi)^2\zeta_T^2 \geq 2\sigma_z^2 T\ln\left(\zeta_T^{-1}T\right)+(1-\pi)^2\zeta_T^2$. Hence, the required condition is met if $\lim_{T\to\infty}[\left(C_1 q^{-1}\right)\left(C_1^{-1}q(1-\pi)\zeta_T/2\right)^{s/(s+1)}-\ln\left(\zeta_T^{-1}T\right)] \geq 0$. This condition is clearly satisfied noting that for values of $\zeta_T=\ominus\left(T^\lambda\right)$, $q>0$, $C_1>0$ and $0<\pi<1$,

$$\left(C_1 q^{-1}\right)\left(C_1^{-1}q(1-\pi)\zeta_T/2\right)^{s/(s+1)}-\ln\left(\zeta_T^{-1}T\right)=\ominus(T^{\frac{\lambda s}{1+s}})-\ominus[\ln(T)],$$

[2] A term A is said to be asymptotically dominant compared to a term B if both tend to infinity and $A/B\to\infty$.

since $s > 0$ and $\lambda > 0$, the first term on the RHS, which is positive, dominates the second term. Finally, we require that $D_T\zeta_T = o(T)$, since then the denominator of the fraction inside the exponential on the RHS of (B.7) is dominated by T which takes us back to the Exponential inequality with bounded random variables and proves (B.1). Consider $T^{-1}D_T\zeta_T = [C_1^{-1}q(1-\pi)/2]^{1/(s+1)}T^{-1}\zeta_T^{(2+s)/(1+s)}$, and since $\zeta_T = \ominus(T^\lambda)$ then $D_T\zeta_T = o(T)$, as long as $\lambda < (s+1)/(s+2)$, as required.

If $\lambda > (s+1)/(s+2)$, it follows that $D_T\zeta_T$ dominates T in the denominator of the fraction inside the exponential on the RHS of (B.7). So the bound takes the form $\exp[-(1-\pi)\zeta_T^2/(C_4D_T\zeta_T)]$, for some finite positive constant C_4. Noting that $D_T = \ominus(\zeta_T^{1/(s+1)})$, gives a bound of the form $\exp[-C_3\zeta_T^{s/(s+1)}]$ proving (B.2). ■

Lemma A4 *Let x_t and u_t be sequences of random variables and suppose that there exist $C_0, C_1 > 0$, and $s > 0$ such that $\sup_t \Pr(|x_t| > \alpha) \leq C_0\exp(-C_1\alpha^s)$ and $\sup_t \Pr(|u_t| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, for all $\alpha > 0$. Let $\mathcal{F}_{t-1}^{(1)} = \sigma(\{u_s\}_{s=1}^{t-1}, \{x_s\}_{s=1}^{t-1})$ and $\mathcal{F}_t^{(2)} = \sigma(\{u_s\}_{s=1}^{t-1}, \{x_s\}_{s=1}^{t})$. Then, assume either that (i) $E(u_t|\mathcal{F}_t^{(2)}) = 0$ or (ii) $E(x_tu_t - \mu_t|\mathcal{F}_{t-1}^{(1)}) = 0$, where $\mu_t = E(x_tu_t)$. Let $\zeta_T = \ominus(T^\lambda)$, for some λ such that $0 < \lambda \leq (s/2+1)/(s/2+2)$. Then, for any π in the range $0 < \pi < 1$ we have*

$$\Pr(|\sum_{t=1}^T (x_tu_t - \mu_t)| > \zeta_T) \leq \exp[-(1-\pi)^2\zeta_T^2/(2T\sigma_{(T)}^2)], \tag{B.10}$$

where $\sigma_{(T)}^2 = T^{-1}\sum_{t=1}^T \sigma_t^2$ and $\sigma_t^2 = E[(x_tu_t - \mu_t)^2|\mathcal{F}_{t-1}^{(1)}]$. If $\lambda > (s/2+1)/(s/2+2)$, then for some finite positive constant C_2,

$$\Pr(|\sum_{t=1}^T (x_tu_t - \mu_t)| > \zeta_T) \leq \exp[-C_2\zeta_T^{s/(s+2)}]. \tag{B.11}$$

Proof. Let $\tilde{\mathcal{F}}_{t-1} = \sigma(\{x_su_s\}_{s=1}^{t-1})$ and note that under (i), $E(x_tu_t|\tilde{\mathcal{F}}_{t-1}) = E[E(u_t|\mathcal{F}_t^{(2)})x_t|\tilde{\mathcal{F}}_{t-1}] = 0$. Therefore, x_tu_t is a martingale difference process. Under (ii) we simply note that $x_tu_t - \mu_t$ is a martingale difference process by assumption. Next, for any $\alpha > 0$ we have (using (B.60) with C_0 set equal to α and C_1 set equal to $\sqrt{\alpha}$)

$$\Pr[|x_tu_t| > \alpha] \leq \Pr[|x_t| > \alpha^{1/2}] + \Pr[|u_t| > \alpha^{1/2}]. \tag{B.12}$$

But, under the assumptions of the lemma, $\sup_t \Pr[|x_t| > \alpha^{1/2}] \leq C_0e^{-C_1\alpha^{s/2}}$, and $\sup_t \Pr[|u_t| > \alpha^{1/2}] \leq C_0e^{-C_1\alpha^{s/2}}$. Hence $\sup_t \Pr[|x_tu_t| > \alpha] \leq 2C_0e^{-C_1\alpha^{s/2}}$. Therefore, the process x_tu_t satisfies the conditions of Lemma A3 and the results of the lemma apply. ■

Lemma A5 *Let $\boldsymbol{x} = (x_1, x_2, ..., x_T)'$ and $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ be sequences of random variables and suppose that there exist finite positive constants C_0 and C_1, and $s > 0$ such that $\sup_t \Pr(|x_t| > \alpha) \leq C_0\exp(-C_1\alpha^s)$ and $\sup_{i,t} \Pr(|q_{i,t}| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, for all $a > 0$. Consider the linear projection $x_t = \sum_{j=1}^{l_T}\beta_jq_{jt} + u_{x,t}$, and assume that only a finite number of slope coefficients β's are nonzero and bounded, and the remaining β's are zero. Then, there exist finite positive constants C_2 and C_3, such that $\sup_t \Pr(|u_{x,t}| > \alpha) \leq C_2\exp(-C_3\alpha^s)$.*

Proof. We assume without any loss of generality that the $|\beta_i| < C_0$ for $i = 1, 2, ..., M$, M is a finite positive integer and $\beta_i = 0$ for $i = M+1, M+2, ..., l_T$. Note that for some $0 < \pi < 1$, $\sup_t \Pr(|u_{x,t}| > \alpha) \leq \sup_t \Pr(|x_t - \sum_{j=1}^{M} \beta_j q_{jt}| > \alpha) \leq \sup_t \Pr(|x_t| > (1-\pi)\alpha) + \sup_t \Pr(|\sum_{j=1}^{M} \beta_j q_{jt}| > \pi\alpha) \leq \sup_t \Pr(|x_t| > (1-\pi)\alpha) + \sup_t \sum_{j=1}^{M} \Pr(|\beta_j q_{jt}| > \pi\alpha/M)$, and since $|\beta_j| > 0$, then $\sup_t \Pr(|u_{x,t}| > \alpha) \leq \sup_t \Pr(|x_t| > (1-\pi)\alpha) + M \sup_{j,t} \Pr[|q_{jt}| > \pi\alpha/(M|\beta_j|)]$. But $\sup_{j,t} \Pr[|q_{jt}| > \pi\alpha/(M|\beta_j|)] \leq \sup_{j,t} \Pr[|q_{jt}| > \pi\alpha/(M\beta_{\max})] \leq C_0 \exp\{-C_1[\pi\alpha/(M\beta_{\max})]^s\}$, and, for fixed M, the probability bound condition is clearly met. ■

Lemma A6 *Let x_{it}, $i = 1, 2, ..., n$, $t = 1, 2, ..., T$, and η_t be processes that satisfy exponential tail probability bounds of the form (9) and (10), with tail exponents s_x and s_η, where $s = \min(s_x, s_\eta) > 0$. Further, let $x_{it}\eta_t$, $i = 1, 2, ..., n$, be martingale difference processes. Let $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ contain a constant and a subset of $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$. Let $\boldsymbol{\Sigma}_{qq} = T^{-1}\sum_{t=1}^{T} E(\boldsymbol{q}_{\cdot t}\boldsymbol{q}'_{\cdot t})$ and $\hat{\boldsymbol{\Sigma}}_{qq} = \mathbf{Q}'\mathbf{Q}/T$ be both invertible, where $\mathbf{Q} = (\boldsymbol{q}_{1\cdot}, \boldsymbol{q}_{2\cdot}, ..., \boldsymbol{q}_{l_T\cdot})$ and $\boldsymbol{q}_{i\cdot} = (q_{i1}, q_{i2}, ..., q_{iT})'$, for $i = 1, 2, ..., l_T$. Suppose that Assumption 5 holds for x_{it} and $\boldsymbol{q}_{\cdot t}$, $i = 1, 2, ..., n$, and for η_t and $\boldsymbol{q}_{\cdot t}$, and denote the corresponding projection residuals defined by (11) as $u_{x_i,t} = x_{it} - \boldsymbol{\gamma}'_{qx_i,T}\boldsymbol{q}_{\cdot t}$ and $u_{\eta,t} = \eta_t - \boldsymbol{\gamma}'_{q\eta,T}\boldsymbol{q}_{\cdot t}$, respectively. Let $\hat{\boldsymbol{u}}_{x_i} = (\hat{u}_{x_i,1}, \hat{u}_{x_i,2}, ..., \hat{u}_{x_i,T})' = \mathbf{M}_q\boldsymbol{x}_i$, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$, $\hat{\boldsymbol{u}}_\eta = (\hat{u}_{\eta,1}, \hat{u}_{\eta,2}, ..., \hat{u}_{\eta,T})' = \mathbf{M}_q\boldsymbol{\eta}$, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_T)'$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}$, $\mathcal{F}_t = \mathcal{F}_t^\eta \cup \mathcal{F}_t^x$, $\mu_{x_i\eta,t} = E(u_{x_i,t}u_{\eta,t}|\mathcal{F}_{t-1})$, $\omega^2_{x_i\eta,1,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[(x_{it}\eta_t - E(x_{it}\eta_t|\mathcal{F}_{t-1}))^2\right]$, and $\omega^2_{x_i\eta,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[(u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t})^2\right]$. Let $\zeta_T = \ominus(T^\lambda)$. Then, for any π in the range $0 < \pi < 1$, we have,*

$$\Pr[|\textstyle\sum_{t=1}^{T} x_{it}\eta_t - E(x_{it}\eta_t|\mathcal{F}_{t-1})| > \zeta_T] \leq \exp[-(1-\pi)^2\zeta_T^2/(2T\omega^2_{x_i\eta,1,T})], \tag{B.13}$$

if $0 < \lambda \leq (s/2+1)/(s/2+2)$. Further, if $\lambda > (s/2+1)/(s/2+2)$, we have,

$$\Pr[|\textstyle\sum_{t=1}^{T} x_{it}\eta_t - E(x_{it}\eta_t|\mathcal{F}_{t-1})| > \zeta_T] \leq \exp[-C_0\zeta_T^{s/(s+2)}], \tag{B.14}$$

for some finite positive constant C_0. If it is further assumed that $l_T = \ominus(T^d)$, such that $0 \leq d < 1/3$, then, if $3d/2 < \lambda \leq (s/2+1)/(s/2+2)$,

$$\Pr[|\textstyle\sum_{t=1}^{T}(\hat{u}_{x_i,t}\hat{u}_{\eta,t} - \mu_{x_i\eta,t})| > \zeta_T] \leq C_0\exp[-(1-\pi)^2\zeta_T^2/(2T\omega^2_{x_i\eta,T})] + \exp\left[-C_1T^{C_2}\right]. \tag{B.15}$$

for some finite positive constants C_0, C_1 and C_2, and, if $\lambda > (s/2+1)/(s/2+2)$ we have

$$\Pr[|\textstyle\sum_{t=1}^{T}(\hat{u}_{x_i,t}\hat{u}_{\eta,t} - \mu_{x_i\eta,t})| > \zeta_T] \leq C_0\exp[-C_3\zeta_T^{s/(s+2)}] + \exp\left[-C_1T^{C_2}\right], \tag{B.16}$$

for some finite positive constants C_0, C_1, C_2 and C_3.

Proof. Note that all the results in the proofs below hold both for sequences and for triangular arrays of random variables. If $\boldsymbol{q}_{\cdot t}$ contains x_{it}, all results follow trivially, so, without loss of generality, we assume that, if this is the case, the relevant column of $\mathbf{Q}$ is removed. (B.13) and (B.14) follow immediately given our assumptions and Lemma A4. We proceed to prove the rest of the lemma. Let $\boldsymbol{u}_{x_i} = (u_{x_i,1}, u_{x_i,2}, ..., u_{x_i,T})'$ and $\boldsymbol{u}_\eta = (u_{\eta,1}, u_{\eta,2}, ..., u_{\eta,T})'$. We first note that $\sum_{t=1}^{T} (\hat{u}_{x_i,t}\hat{u}_{\eta,t} - \mu_{x_i\eta,t}) = \hat{\boldsymbol{u}}'_{x_i}\hat{\boldsymbol{u}}_\eta - \sum_{t=1}^{T} \mu_{x_i\eta,t} = \boldsymbol{u}'_{x_i}\mathbf{M}_q\boldsymbol{u}_\eta - \sum_{t=1}^{T} \mu_{x_i\eta,t}$, and

$$\sum_{t=1}^{T} (\hat{u}_{x_i,t}\hat{u}_{\eta,t} - \mu_{x_i\eta,t}) = \sum_{t=1}^{T} (u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t}) - \left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)\hat{\mathbf{\Sigma}}_{qq}^{-1}(\mathbf{Q}'\boldsymbol{u}_\eta), \tag{B.17}$$

where $\hat{\mathbf{\Sigma}}_{qq} = T^{-1}(\mathbf{Q}'\mathbf{Q})$. The second term of the above expression can now be decomposed as

$$\left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)\hat{\mathbf{\Sigma}}_{qq}^{-1}(\mathbf{Q}'\boldsymbol{u}_\eta) = \left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)(\hat{\mathbf{\Sigma}}_{qq}^{-1} - \mathbf{\Sigma}_{qq}^{-1})(\mathbf{Q}'\boldsymbol{u}_\eta) + \left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)\mathbf{\Sigma}_{qq}^{-1}(\mathbf{Q}'\boldsymbol{u}_\eta). \tag{B.18}$$

By (B.59) and for any $0 < \pi_1, \pi_2, \pi_3 < 1$ such that $\sum_{i=1}^{3}\pi_i = 1$, we have $\Pr[|\sum_{t=1}^{T} (\hat{u}_{x_i,t}\hat{u}_{\eta,t} - \mu_{x_i\eta,t})| > \zeta_T] \leq \Pr[|\sum_{t=1}^{T} (u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t})| > \pi_1\zeta_T] +$
$\Pr[|\left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)(\hat{\mathbf{\Sigma}}_{qq}^{-1} - \mathbf{\Sigma}_{qq}^{-1})(\mathbf{Q}'\boldsymbol{u}_\eta)| > \pi_2\zeta_T] + \Pr[|\left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)\mathbf{\Sigma}_{qq}^{-1}(\mathbf{Q}'\boldsymbol{u}_\eta)| > \pi_3\zeta_T]$. Also applying (B.60) to the last two terms of the above we obtain
$\Pr[|\left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)(\hat{\mathbf{\Sigma}}_{qq}^{-1} - \mathbf{\Sigma}_{qq}^{-1})(\mathbf{Q}'\boldsymbol{u}_\eta)| > \pi_2\zeta_T] \leq \Pr(\|\hat{\mathbf{\Sigma}}_{qq}^{-1} - \mathbf{\Sigma}_{qq}^{-1}\|_F \left\|T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right\|_F \|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > \pi_2\zeta_T) \leq$
$\Pr(\|\hat{\mathbf{\Sigma}}_{qq}^{-1} - \mathbf{\Sigma}_{qq}^{-1}\|_F > \zeta_T/\delta_T) + \Pr(T^{-1}\left\|\boldsymbol{u}'_{x_i}\mathbf{Q}\right\|_F \|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > \pi_2\delta_T) \leq \Pr(\|\hat{\mathbf{\Sigma}}_{qq}^{-1} - \mathbf{\Sigma}_{qq}^{-1}\|_F > \zeta_T/\delta_T) +$
$\Pr[\left\|\boldsymbol{u}'_{x_i}\mathbf{Q}\right\|_F > (\pi_2\delta_T T)^{1/2}] + \Pr[\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > (\pi_2\delta_T T)^{1/2}]$, where $\delta_T > 0$ is a deterministic sequence. In what follows, we set $\delta_T = \ominus(\zeta_T^\alpha)$, for some $\alpha > 0$. Similarly,
$\Pr[|\left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)\mathbf{\Sigma}_{qq}^{-1}(\mathbf{Q}'\boldsymbol{u}_\eta)| > \pi_3\zeta_T] \leq \Pr(\left\|\mathbf{\Sigma}_{qq}^{-1}\right\|_F \left\|T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right\|_F \|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > \pi_3\zeta_T) \leq$
$\Pr[\left\|\boldsymbol{u}'_{x_i}\mathbf{Q}\right\|_F \|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > \pi_3\zeta_T T/\left\|\mathbf{\Sigma}_{qq}^{-1}\right\|_F] \leq \Pr(\left\|\boldsymbol{u}'_{x_i}\mathbf{Q}\right\|_F > \pi_3^{1/2}\zeta_T^{1/2}T^{1/2}\left\|\mathbf{\Sigma}_{qq}^{-1}\right\|_F^{-1/2}) +$
$\Pr(\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > \pi_3^{1/2}\zeta_T^{1/2}T^{1/2}\left\|\mathbf{\Sigma}_{qq}^{-1}\right\|_F^{-1/2})$. Overall

$$\begin{aligned}
&\Pr(|\textstyle\sum_{t=1}^{T} (\hat{u}_{x,t}\hat{u}_{\eta,t} - \mu_{x\eta,t})| > \zeta_T) \leq \Pr(|\textstyle\sum_{t=1}^{T} (u_{x,t}u_{\eta,t} - \mu_{x\eta,t})| > \pi_1\zeta_T) \\
&+ \Pr\left(\left\|\hat{\mathbf{\Sigma}}_{qq}^{-1} - \mathbf{\Sigma}_{qq}^{-1}\right\|_F > \zeta_T/\delta_T\right) + \Pr\left(\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > (\pi_2\delta_T T)^{1/2}\right) + \Pr\left(\|\boldsymbol{u}'_x\mathbf{Q}\|_F > (\pi_2\delta_T T)^{1/2}\right) \\
&+ \Pr(\|\boldsymbol{u}'_x\mathbf{Q}\|_F > \pi_3^{1/2}\zeta_T^{1/2}T^{1/2}\left\|\mathbf{\Sigma}_{qq}^{-1}\right\|_F^{-1/2}) + \Pr(\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > \pi_3^{1/2}\zeta_T^{1/2}T^{1/2}\left\|\mathbf{\Sigma}_{qq}^{-1}\right\|_F^{-1/2}).
\end{aligned} \tag{B.19}$$

First, since $u_{x,t}u_{\eta,t} - \mu_{x\eta,t}$ is a martingale difference process with respect to $\sigma(\{\eta_s\}_{s=1}^{t-1}, \{x_s\}_{s=1}^{t-1}, \{q_s\}_{s=1}^{t-1})$, by Lemma A4, we have, for any π in the range $0 < \pi < 1$,

$$\Pr[|\textstyle\sum_{t=1}^{T} (u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t})| > \pi_1\zeta_T] \leq \exp[-(1-\pi)^2\zeta_T^2/(2T\omega_{x\eta,T}^2)], \tag{B.20}$$

if $0 < \lambda \leq (s/2+1)/(s/2+2)$, and

$$\Pr[|\textstyle\sum_{t=1}^{T} (u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t})| > \pi_1\zeta_T] \leq \exp[-C_0\zeta_T^{s/(s+1)}], \tag{B.21}$$

if $\lambda > (s/2+1)/(s/2+2)$, for some finite positive constant C_0. We now show that the last five terms on the RHS of (B.19) are of order $\exp\left[-C_1 T^{C_2}\right]$, for some finite positive constants C_1 and C_2. We will make use of Lemma A4 since by assumption $\{q_{it}u_{\eta,t}\}$ and $\{q_{it}u_{x_i,t}\}$ are martingale difference sequences. We note that some of the bounds of the last five terms exceed, in order, $T^{1/2}$. Since we do not know the value of s, we need to consider the possibility that either (B.10) or (B.11) of Lemma A4, apply. We start with (B.10). Then, for some finite positive constant C_0, we have[3]

$$\sup_i \Pr[\|\mathbf{q}'_i\boldsymbol{u}_\eta\| > (\pi_2\delta_T T)^{1/2}] \leq \exp(-C_0\delta_T). \tag{B.22}$$

[3]The required probability bound on u_{xt} follows from the probability bound assumptions on x_t and on q_{it}, for $i = 1, 2, ..., l_T$, even if $l_T \to \infty$. See also Lemma A5.

Also, using $\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F^2 = \sum_{j=1}^{l_T}(\sum_{t=1}^T q_{jt}u_t)^2$ and (B.59), $\Pr[\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > (\pi_2\delta_T T)^{1/2}] = \Pr(\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F^2 > \pi_2\delta_T T) \leq \sum_{j=1}^{l_T}\Pr[(\sum_{t=1}^T q_{jt}u_{\eta,t})^2 > \pi_2\delta_T T/l_T] = \sum_{j=1}^{l_T}\Pr[|\sum_{t=1}^T q_{jt}u_{\eta,t}| > (\pi_2\delta_T T/l_T)^{1/2}]$, which upon using (B.22) yields (for some finite positive constant C_0)

$$\Pr[\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > (\pi_2\delta_T T)^{1/2}] \leq l_T \exp\left(-C_0\delta_T/l_T\right), \;\; \Pr[\|\mathbf{Q}'\boldsymbol{u}_x\| > (\pi_2\delta_T T)^{1/2}] \leq l_T \exp\left(-C_0\delta_T/l_T\right). \tag{B.23}$$

Similarly,

$$\Pr(\|\mathbf{Q}'\boldsymbol{u}_\eta\|_F > \pi_3^{1/2}\zeta_T^{1/2}T^{1/2}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-1/2}) \leq l_T\exp[-C_0\zeta_T/(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T)], \tag{B.24}$$
$$\Pr(\|\mathbf{Q}'\boldsymbol{u}_x\| > \pi_3^{1/2}\zeta_T^{1/2}T^{1/2}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-1/2}) \leq l_T\exp[-C_0\zeta_T/(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T)].$$

Turning to the second term of (B.19), since for all i and j, $\{q_{it}q_{jt} - E(q_{it}q_{jt})\}$ is a martingale difference process and q_{it} satisfy the required probability bound then

$$\sup_{ij}\Pr\{|T^{-1}\textstyle\sum_{t=1}^T[q_{it}q_{jt} - E(q_{it}q_{jt})]| > \pi_2\zeta_T/\delta_T\} \leq \exp(-C_0T\zeta_T^2/\delta_T^2). \tag{B.25}$$

Therefore, by Lemma A16, for some finite positive constant C_0, we have

$$\begin{aligned}\Pr(\|\hat{\boldsymbol{\Sigma}}_{qq}^{-1} - \boldsymbol{\Sigma}_{qq}^{-1}\| > \zeta_T/\delta_T) &\leq l_T^2\exp[-C_0T\zeta_T^2\delta_T^{-2}l_T^{-2}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \delta_T^{-1}\zeta_T)^{-2}] \\ &\quad + l_T^2\exp(-C_0T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}l_T^{-2}).\end{aligned} \tag{B.26}$$

Further by Lemma A14, $\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F = \ominus\left(l_T^{1/2}\right)$, and $T\zeta_T^2\delta_T^{-2}l_T^{-2}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \delta_T^{-1}\zeta_T)^{-2} = Tl_T^{-2}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}(\delta_T\zeta_T^{-1}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + 1)^{-2}$. Consider now the different terms in the above expression and let $P_{11} = \delta_T/l_T$, $P_{12} = \zeta_T/(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T)$, $P_{13} = Tl_T^{-2}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}[\delta_T\zeta_T^{-1}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + 1]^{-2}$, and $P_{14} = T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}l_T^{-2}$. Under $\delta_T = \ominus(\zeta_T^\alpha)$, $l_T = \ominus(T^d)$, and $\zeta_T = \ominus(T^\lambda)$, we have $P_{11} = \delta_T/l_T = \ominus\left(T^{\alpha-d}\right)$,

$$P_{12} = \zeta_T/(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T) = \ominus\left(T^{\lambda-3d/2}\right), \tag{B.27}$$

$P_{13} = Tl_T^{-2}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}[\delta_T\zeta_T^{-1}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + 1]^{-2} = \ominus\left(T^{\max\{1+2\lambda-4d-2\alpha,1+\lambda-7d/2-\alpha,1-3d\}}\right)$, and $P_{14} = T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-2}l_T^{-2} = \ominus\left(T^{1-3d}\right)$. Suppose that $d < 1/3$, and by (B.27) note that $\lambda \geq 3d/2$. Then, setting $\alpha = 1/3$, ensures that all the above four terms tend to infinity polynomially with T. Therefore, it also follows that they can be represented as terms of order $\exp\left[-C_1T^{C_2}\right]$, for some finite positive constants C_1 and C_2, and (B.15) follows. The same analysis can be repeated under (B.11). In this case, (B.23), (B.24), (B.25) and (B.26) are replaced by

$$\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F > (\pi_2\delta_T T)^{1/2}\right) \leq l_T \exp\left(-\frac{C_0\delta_T^{s/2(s+2)}T^{s/2(s+2)}}{l_T^{s/2(s+2)}}\right) = l_T \exp\left[-C_0\left(\frac{\delta_T T}{l_T}\right)^{s/2(s+2)}\right],$$

$$\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{x}\right\| > (\pi_2\delta_T T)^{1/2}\right) \leq l_T \exp\left(-\frac{C_0\delta_T^{s/2(s+2)}T^{s/2(s+2)}}{l_T^{s/2(s+2)}}\right) = l_T \exp\left[-C_0\left(\frac{\delta_T T}{l_T}\right)^{s/2(s+2)}\right],$$

$$\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F > \frac{\pi_3^{1/2}\zeta_T^{1/2}T^{1/2}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{1/2}}\right) \leq l_T \exp\left(\frac{-C_0\zeta_T^{s/2(s+2)}T^{s/2(s+2)}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{s/2(s+2)} l_T^{s/2(s+2)}}\right) = l_T \exp\left[-C_0\left(\frac{\zeta_T T}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T}\right)^{s/2(s+2)}\right]$$

$$\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{x}\right\| > \frac{\pi_3^{1/2}\zeta_T^{1/2}T^{1/2}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{1/2}}\right) \leq l_T \exp\left(\frac{-C_0\zeta_T^{s/2(s+2)}T^{s/2(s+2)}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{s/2(s+2)} l_T^{s/2(s+2)}}\right) = l_T \exp\left[-C_0\left(\frac{\zeta_T T}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T}\right)^{s/2(s+2)}\right],$$

$\sup_{ij}\Pr\{|T^{-1}\sum_{t=1}^{T}[q_{it}q_{jt} - E(q_{it}q_{jt})]| > \pi_2\zeta_T/\delta_T\} \leq \exp[-C_0 T^{s/(s+2)}\zeta_T^{s/(s+2)}\delta_T^{-s/(s+2)}]$, and, using Lemma A17, $\Pr[\|(\hat{\boldsymbol{\Sigma}}_{qq}^{-1} - \boldsymbol{\Sigma}_{qq}^{-1}\| > \pi_2\zeta_T/\delta_T] \leq$ $l_T^2\exp[-C_0 T^{s/(s+2)}\zeta_T^{s/(s+2)}\delta_T^{-s/(s+2)}l_T^{-s/(s+2)}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-s/(s+2)}(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \delta_T^{-1}\zeta_T)^{-s/(s+2)}]+$ $l_T^2\exp[-C_0 T^{s/(s+2)}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-s/(s+2)}l_T^{-s/(s+2)}] = l_T^2\exp\left(-C_0\{T\zeta_T/[\delta_T l_T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \delta_T^{-1}\zeta_T)]\}^{s/(s+2)}\right)+$ $l_T^2\exp[-C_0(T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-1}l_T^{-1})^{s/(s+2)}]$, respectively. Once again, we need to derive conditions that imply that $P_{21} = \delta_T T/l_T$, $P_{22} = \zeta_T T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-1}l_T^{-1}$, $P_{23} = T\zeta_T[\delta_T l_T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \delta_T^{-1}\zeta_T)]^{-1}$ and $P_{24} = T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-1}l_T^{-1}$ are terms that tend to infinity polynomially with T. If that is the case then, as before, the relevant terms are of order $\exp\left[-C_1T^{C_2}\right]$, for some finite positive constants C_1 and C_2, and (B.16) follows, completing the proof of the lemma. P_{22} dominates P_{23} so we focus on P_{21}, P_{23} and P_{24}. We have $\delta_T T/l_T = \ominus\left(T^{1+\alpha-d/2}\right)$, $T\zeta_T[\delta_T l_T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \delta_T^{-1}\zeta_T)]^{-1} = \ominus\left[T^{\max(1+\lambda-\alpha-2d,1-3d/2)}\right]$, and $T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{-1}l_T^{-1} = \ominus\left(T^{1-3d/2}\right)$. It immediately follows that under the conditions set when using (B.10), which were that $\alpha = 1/3$, $d < 1/3$ and $\lambda > 3d/2$, and as long as $s > 0$, P_{21} to P_{24} tend to infinity polynomially with T, proving the lemma.[4] ■

Lemma A7 *Let x_{it}, $i = 1, 2, ..., n$, be processes that satisfy exponential tail probability bounds of the form (9), with positive tail exponent s. Let $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ contain a constant and a subset of $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$. Suppose that Assumption 5 holds for x_{it} and $\boldsymbol{q}_{\cdot t}$, $i = 1, 2, ..., n$, and denote the corresponding projection residuals defined by (11) as $u_{x_it} = x_{it} - \boldsymbol{\gamma}'_{qx_i,T}\boldsymbol{q}_{\cdot t}$. Let $\boldsymbol{\Sigma}_{qq} = T^{-1}\sum_{t=1}^{T}E\left(\boldsymbol{q}_{\cdot t}\boldsymbol{q}'_{\cdot t}\right)$ and $\hat{\boldsymbol{\Sigma}}_{qq} = \mathbf{Q}'\mathbf{Q}/T$ be both invertible, where $\mathbf{Q} = (\boldsymbol{q}_{1\cdot}, \boldsymbol{q}_{2\cdot}, ..., \boldsymbol{q}_{l_T\cdot})$ and $\boldsymbol{q}_{i\cdot} = (q_{i1}, q_{i2}, .., q_{iT})'$, for $i = 1, 2, ..., l_T$. Let $\hat{\boldsymbol{u}}_{x_i} = (\hat{u}_{x_i,1}, \hat{u}_{x_i,2}, ..., \hat{u}_{x_i,T})' = \mathbf{M}_q\boldsymbol{x}_i$, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$ and $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}$. Moreover, suppose that $E\left(u_{x_i,t}^2 - \sigma_{x_it}^2|\mathcal{F}_{t-1}\right) = 0$,*

[4] It is important to highlight one particular feature of the above proof. In (B.23), $q_{it}u_{x,t}$ needs to be a martingale difference process. Note that if q_{it} is a martingale difference process distributed independently of $u_{x,t}$, then $q_{it}u_{x,t}$ is also a martingale difference process irrespective of the nature of $u_{x,t}$. This implies that one may not need to impose a martingale difference assumption on $u_{x,t}$ if x_{it} is a noise variable. Unfortunately, a leading case for which this lemma is used is one where $q_{it} = 1$. It is then clear that one needs to impose a martingale difference assumption on $u_{x,t}$, to deal with covariates that cannot be represented as martingale difference processes. We relax this assumption in Section C of the online theory supplement where we allow noise variables to be mixing processes.

where $\mathcal{F}_t = \mathcal{F}_t^x$ *and* $\sigma_{x_it}^2 = E(u_{x_i,t}^2)$*. Let* $\zeta_T = \ominus(T^\lambda)$*. Then, if* $0 < \lambda \leq (s/2+1)/(s/2+2)$*, for any* π *in the range* $0 < \pi < 1$*, and some finite positive constant* C_0*, we have,*

$$\Pr\left[\left|\sum_{t=1}^T \left(x_{it}^2 - \sigma_{x_it}^2\right)\right| > \zeta_T\right] \leq C_0 \exp\left[-\left(1-\pi\right)^2 \zeta_T^2 T^{-1} \omega_{i,1,T}^{-2}/2\right]. \quad \text{(B.28)}$$

Otherwise, if $\lambda > (s/2+1)/(s/2+2)$*, for some finite positive constant* C_0*, we have*

$$\Pr\left[\left|\sum_{t=1}^T \left(x_{it}^2 - \sigma_{x_it}^2\right)\right| > \zeta_T\right] \leq \exp\left[-C_0 \zeta_T^{s/(s+2)}\right]. \quad \text{(B.29)}$$

If it is further assumed that $l_T = \ominus\left(T^d\right)$*, such that* $0 \leq d < 1/3$*, then, if* $3d/2 < \lambda \leq (s/2+1)/(s/2+2)$*,*

$$\Pr\left[\left|\sum_{t=1}^T \left(\hat{u}_{x_i,t}^2 - \sigma_{x_it}^2\right)\right| > \zeta_T\right] \leq C_0 \exp\left[-\left(1-\pi\right)^2 \zeta_T^2 T^{-1} \omega_{i,T}^{-2}/2\right] + \exp\left[-C_1 T^{C_2}\right], \quad \text{(B.30)}$$

for some finite positive constants C_0*,* C_1 *and* C_2*, and, if* $\lambda > (s/2+1)/(s/2+2)$*,*

$$\Pr\left[\left|\sum_{t=1}^T \left(\hat{u}_{x_i,t}^2 - \sigma_{x_it}^2\right)\right| > \zeta_T\right] \leq C_0 \exp\left[-C_3 \zeta_T^{s/(s+2)}\right] + \exp\left[-C_1 T^{C_2}\right], \quad \text{(B.31)}$$

for some finite positive constants C_0*,* C_1*,* C_2 *and* C_3*, where* $\omega_{i,1,T}^2 = T^{-1}\sum_{t=1}^T E\left[\left(x_{it}^2 - \sigma_{x_it}^2\right)^2\right]$ *and* $\omega_{i,T}^2 = T^{-1}\sum_{t=1}^T E\left[\left(u_{x_i,t}^2 - \sigma_{x_it}^2\right)^2\right]$*.*

Proof. If $\boldsymbol{q}_{\cdot t}$ contains x_{it}, all results follow trivially, so, without loss of generality, we assume that, if this is the case, the relevant column of $\mathbf{Q}$ is removed. (B.28) and (B.29) follow similarly to (B.13) and (B.14). For (B.30) and (B.31), we first note that $|\sum_{t=1}^T \left(\hat{u}_{x_i,t}^2 - \sigma_{x_it}^2\right)| \leq |\sum_{t=1}^T \left(u_{x_i,t}^2 - \sigma_{x_it}^2\right)| + |\left(T^{-1}\boldsymbol{u}_{x_i}'\mathbf{Q}\right)\left(T^{-1}\mathbf{Q}'\mathbf{Q}\right)^{-1}\left(\mathbf{Q}'\boldsymbol{u}_{x_i}\right)|$. Since $\left\{u_{x_i,t}^2 - \sigma_{x_it}^2\right\}$ is a martingale difference process and for $\alpha > 0$ and $s > 0$, $\sup_t \Pr\left(\left|u_{x_i,t}^2\right| > \alpha^2\right) = \sup_t \Pr\left(|u_{x_i,t}| > \alpha\right) \leq C_0 \exp\left(-C_1\alpha^s\right)$, by Lemma A5, then the conditions of Lemma A3 are met and we have $\Pr[|\sum_{t=1}^T \left(u_{x_i,t}^2 - \sigma_{x_it}^2\right)| > \zeta_T] \leq \exp[-\left(1-\pi\right)^2 \zeta_T^2 T^{-1}\omega_{i,T}^{-2}/2]$, if $0 < \lambda \leq (s/2+1)/(s/2+2)$, and $\Pr[|\sum_{t=1}^T \left(u_{x_i,t}^2 - \sigma_{x_it}^2\right)| > \zeta_T] \leq \exp[-C_0\zeta_T^{s/(s+2)}]$, if $\lambda > (s/2+1)/(s/2+2)$. Then, using the same line of reasoning as in the proof of Lemma A6 we establish the desired result. ■

Lemma A8 *Let* y_t*, for* $t = 1, 2, ..., T$*, be given by the data generating process (6) and suppose that* u_t *and* $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$ *satisfy Assumptions 2-4, with* $s = \min(s_x, s_u) > 0$*. Let* $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ *contain a constant and a subset of* $\boldsymbol{x}_{nt}$*. Assume that* $\mathbf{\Sigma}_{qq} = \frac{1}{T}\sum_{t=1}^T E\left(\boldsymbol{q}_{\cdot t}\boldsymbol{q}_{\cdot t}'\right)$ *and* $\hat{\mathbf{\Sigma}}_{qq} = \mathbf{Q}'\mathbf{Q}/T$ *are both invertible, where* $\mathbf{Q} = (\boldsymbol{q}_{1\cdot}, \boldsymbol{q}_{2\cdot}, ..., \boldsymbol{q}_{l_T\cdot})$ *and* $\boldsymbol{q}_{i\cdot} = (q_{i1}, q_{i2}, ..., q_{iT})'$*, for* $i = 1, 2, ..., l_T$*. Moreover, suppose that Assumption 5 holds for* x_t *and* $\boldsymbol{q}_{\cdot t}$*, where* x_t *is a generic element of* $\{x_{1t}, x_{2t}, ..., x_{nt}\}$ *that does not belong to* $\boldsymbol{q}_{\cdot t}$*. Denote the corresponding projection residuals defined by (11) as* $u_{x,t} = x_t - \boldsymbol{\gamma}_{qx,T}'\boldsymbol{q}_{\cdot t}$*, and the projection residuals of* y_t *on* $(\boldsymbol{q}_{\cdot t}', x_t)'$ *as* $e_t = y_t - \boldsymbol{\gamma}_{yqx,T}'(\boldsymbol{q}_{\cdot t}', x_t)'$*. Define* $\boldsymbol{x} = (x_1, x_2, ..., x_T)'$*, and* $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$*, and let* $a_T = \ominus\left(T^{\lambda-1}\right)$*. Then, for any* π *in the range* $0 < \pi < 1$*, and as long as* $l_T = \ominus\left(T^d\right)$*, such that* $0 \leq d < 1/3$*, we have, that, if* $3d/2 < \lambda \leq (s/2+1)/(s/2+2)$*,*

$$\Pr\left(\left|T^{-1}\sigma_{x,(T)}^{-2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x} - 1\right| > a_T\right) \leq \exp\left[-\sigma_{x,(T)}^4\left(1-\pi\right)^2 T a_T^2 \omega_{x,(T)}^{-2}/2\right] + \exp\left[-C_0 T^{C_1}\right], \text{ and} \quad \text{(B.32)}$$

$$\Pr[|(T^{-1}\sigma_{x,(T)}^{-2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x})^{-1/2}-1|>a_T]\leq \exp[-\sigma_{x,(T)}^4\left(1-\pi\right)^2 Ta_T^2\omega_{x,(T)}^{-2}/2]+\exp\left[-C_0T^{C_1}\right], \tag{B.33}$$

where

$$\sigma_{x,(T)}^2=T^{-1}\textstyle\sum_{t=1}^T E\left(u_{x,t}^2\right),\ \omega_{x,(T)}^2=T^{-1}\textstyle\sum_{t=1}^T E\left[\left(u_{x,t}^2-\sigma_{xt}^2\right)^2\right]. \tag{B.34}$$

If $\lambda>(s/2+1)/(s/2+2)$,

$$\Pr(|T^{-1}\sigma_{x,(T)}^{-2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}-1|>a_T)\leq \exp[-C_0\left(Ta_T\right)^{s/(s+2)}]+\exp\left[-C_1T^{C_2}\right], \tag{B.35}$$

and

$$\Pr[|(T^{-1}\sigma_{x,(T)}^{-2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x})^{-1/2}-1|>a_T]\leq \exp[-C_0\left(Ta_T\right)^{s/(s+2)}]+\exp\left[-C_1T^{C_2}\right]. \tag{B.36}$$

Also, if $3d/2<\lambda\leq(s/2+1)/(s/2+2)$,

$$\Pr(|T^{-1}\sigma_{u,(T)}^{-2}\boldsymbol{e}'\boldsymbol{e}-1|>a_T)\leq \exp[-\sigma_{u,(T)}^4\left(1-\pi\right)^2 Ta_T^2\omega_{u,(T)}^{-2}/2]+\exp\left[-C_0T^{C_1}\right], \tag{B.37}$$

and

$$\Pr[|(\sigma_{u,(T)}^{-2}\boldsymbol{e}'\boldsymbol{e}/T)^{-1/2}-1|>a_T]\leq \exp[-\sigma_{u,(T)}^4\left(1-\pi\right)^2 Ta_T^2\omega_{u,T}^{-2}/2]+\exp\left[-C_0T^{C_1}\right], \tag{B.38}$$

where $\boldsymbol{e}=(e_1,e_2,...,e_T)'$,

$$\sigma_{u,(T)}^2=T^{-1}\textstyle\sum_{t=1}^T\sigma_t^2,\ \textit{and}\ \omega_{u,T}^2=T^{-1}\textstyle\sum_{t=1}^T E[\left(u_t^2-\sigma_t^2\right)^2]. \tag{B.39}$$

If $\lambda>(s/2+1)/(s/2+2)$,

$$\Pr(|T^{-1}\sigma_{u,(T)}^{-2}\boldsymbol{e}'\boldsymbol{e}-1|>a_T)\leq \exp[-C_0\left(Ta_T\right)^{s/(s+2)}]+\exp\left[-C_1T^{C_2}\right],\ \textit{and} \tag{B.40}$$

$$\Pr[|(\sigma_{u,(T)}^{-2}\boldsymbol{e}'\boldsymbol{e}/T)^{-1/2}-1|>a_T]\leq \exp[-C_0\left(Ta_T\right)^{s/(s+2)}]+\exp\left[-C_1T^{C_2}\right], \tag{B.41}$$

Proof. First note that $T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}-\sigma_{x,(T)}^2=T^{-1}\sum_{t=1}^T\left(\hat{u}_{x,t}^2-\sigma_{xt}^2\right)$**,** where $\hat{u}_{x,t}$, for $t=1,2,...,T$,. is the t-th element of $\hat{\mathbf{u}}_x=\mathbf{M}_q\boldsymbol{x}$. Now applying Lemma A7 to $\sum_{t=1}^T\left(\hat{u}_{x,t}^2-\sigma_{xt}^2\right)$ with $\zeta_T=Ta_T$ we have $\Pr(|\sum_{t=1}^T\left(\hat{u}_{x,t}^2-\sigma_{xt}^2\right)|>\zeta_T)\leq\exp[-\left(1-\pi\right)^2\zeta_T^2\omega_{x,(T)}^{-2}/(2T)]+\exp\left[-C_0T^{C_1}\right]$, if $3d/2<\lambda\leq(s/2+1)/(s/2+2)$, and $\Pr(|\sum_{t=1}^T\left(\hat{u}_{x,t}^2-\sigma_{xt}^2\right)|>\zeta_T)\leq\exp[-C_0\zeta_T^{s/(s+2)}]+\exp\left[-C_1T^{C_2}\right]$, if $\lambda>(s/2+1)/(s/2+2)$, where $\omega_{x,(T)}^2$ is defined by (B.34). Also $\Pr[|T^{-1}\sigma_{x,(T)}^{-2}\sum_{t=1}^T\left(\hat{u}_{x,t}^2-\sigma_{xt}^2\right)|>T^{-1}\sigma_{x,(T)}^{-2}\zeta_T]\leq\exp[-\left(1-\pi\right)^2\zeta_T^2\omega_{x,(T)}^{-2}T^{-1}/2]+\exp\left[-C_0T^{C_1}\right]$, if $3d/2<\lambda\leq(s/2+1)/(s/2+2)$, and $\Pr[|T^{-1}\sigma_{x,(T)}^{-2}\sum_{t=1}^T\left(\hat{u}_{x,t}^2-\sigma_{xt}^2\right)|>\zeta_T T^{-1}\sigma_{x,(T)}^{-2}]\leq\exp[-C_0\zeta_T^{s/(s+2)}]+\exp\left[-C_1T^{C_2}\right]$, if $\lambda>(s/2+1)/(s/2+2)$. Therefore, setting $a_T=\zeta_T/T\sigma_{x,(T)}^2=\ominus\left(T^{\lambda-1}\right)$, we have

$$\Pr(|\sigma_{x,(T)}^{-2}T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}-1|>a_T)\leq \exp[-\sigma_{x,(T)}^4\left(1-\pi\right)^2 Ta_T^2\omega_{x,(T)}^{-2}/2]+\exp\left[-C_0T^{C_1}\right], \tag{B.42}$$

if $3d/2<\lambda\leq(s/2+1)/(s/2+2)$, and

$$\Pr(|\sigma_{x,(T)}^{-2}T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}-1|>a_T)\leq \exp[-C_0\zeta_T^{s/(s+2)}]+\exp\left[-C_1T^{C_2}\right],$$

if $\lambda > (s/2+1)/(s/2+2)$, as required. Now setting $\omega_T = \sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x}$, and using Lemma A13, we have $\Pr[|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{-1/2} - 1| > a_T] \leq \Pr(|\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x} - 1| > a_T)$, and hence

$$\Pr[|(\sigma_{u,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{-1/2} - 1| > a_T] \leq \exp[-\sigma_{x,(T)}^{4} (1-\pi)^2 T a_T^2 \omega_{x,(T)}^{-2}] + \exp\left[-C_0 T^{C_1}\right], \quad (B.43)$$

if $3d/2 < \lambda \leq (s/2+1)/(s/2+2)$, and

$$\Pr[|(\sigma_{u,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{-1/2} - 1| > a_T] \leq \exp[-C_0 \zeta_T^{s/(s+2)}] + \exp\left[-C_1 T^{C_2}\right],$$

if $\lambda > (s/2+1)/(s/2+2)$. Furthermore

$$\Pr[|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} - 1| > a_T] = \Pr\left[\frac{|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x}) - 1|}{(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} + 1} > a_T\right],$$

and using Lemma A11 for some finite positive constant C, we have $\Pr[|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} - 1| > a_T] \leq \Pr[|\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x} - 1| > a_T C^{-1}] + \Pr[(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} + 1 < C^{-1}]$. Let $C = 1$, and note that for this choice of C, $\Pr[(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} + 1 < C^{-1}] = \Pr[(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} < 0] = 0$. Hence $\Pr[|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} - 1| > a_T] \leq \Pr[|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x}) - 1| > a_T]$, and using (B.42),

$$\Pr[|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} - 1| > a_T] \leq \exp[-\sigma_{x,(T)}^{4} (1-\pi)^2 T a_T^2 \omega_{x,(T)}^{-2}/2] + \exp\left[-C_0 T^{C_1}\right], \quad (B.44)$$

if $3d/2 < \lambda \leq (s/2+1)/(s/2+2)$, and

$$\Pr[|(\sigma_{x,(T)}^{-2} T^{-1} \boldsymbol{x}' \mathbf{M}_q \boldsymbol{x})^{1/2} - 1| > a_T] \leq \exp[-C_0 \zeta_T^{s/(s+2)}] + \exp\left[-C_1 T^{C_2}\right],$$

if $\lambda > (s/2+1)/(s/2+2)$. Consider now $\boldsymbol{e}'\boldsymbol{e} = \sum_{t=1}^T e_t^2$ and note that $|\sum_{t=1}^T (e_t^2 - \sigma_t^2)| \leq |\sum_{t=1}^T (u_t^2 - \sigma_t^2)| + |(T^{-1}\boldsymbol{u}'\mathbf{W})(T^{-1}\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\boldsymbol{u})|$, where $\mathbf{W} = (\mathbf{Q}, \boldsymbol{x})$. As before, applying Lemma A7 to $\sum_{t=1}^T (e_t^2 - \sigma_t^2)$, and following similar lines of reasoning we have

$$\Pr[|\textstyle\sum_{t=1}^T \left(e_t^2 - \sigma_t^2\right)| > \zeta_T] \leq \exp[-(1-\pi)^2 \zeta_T^2 T^{-1} \omega_{u,(T)}^{-2}/2] + \exp\left[-C_0 T^{C_1}\right],$$

if $3d/2 < \lambda \leq (s/2+1)/(s/2+2)$, and

$$\Pr[|\textstyle\sum_{t=1}^T \left(e_t^2 - \sigma_t^2\right)| > \zeta_T] \leq \exp[-C_0 \zeta_T^{s/(s+2)}] + \exp\left[-C_1 T^{C_2}\right],$$

if $\lambda > (s/2+1)/(s/2+2)$, which yield (B.37) and (B.40). Result (B.38) also follows along similar lines as used above to prove (B.33). ■

Lemma A9 *Let y_t, for $t = 1, 2, ..., T$, be given by the data generating process (6) and suppose that u_t and $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$ satisfy Assumptions 2-4. Let $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ contain a constant and a subset of $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$, and $l_T = o(T^{1/3})$. Assume that $\mathbf{\Sigma}_{qq} = \frac{1}{T}\sum_{t=1}^T E(\boldsymbol{q}_{\cdot t}\boldsymbol{q}_{\cdot t}')$ and $\hat{\mathbf{\Sigma}}_{qq} = \mathbf{Q}'\mathbf{Q}/T$ are both invertible, where $\mathbf{Q} = (\boldsymbol{q}_{1\cdot}, \boldsymbol{q}_{2\cdot}, ..., \boldsymbol{q}_{l_T\cdot})$ and $\boldsymbol{q}_{i\cdot} =$*

$(q_{i1}, q_{i2}, .., q_{iT})'$, for $i = 1, 2, ..., l_T$. Suppose that Assumption 5 holds for x_t and $\boldsymbol{q}_{\cdot t}$, where x_t is a generic element of $\{x_{1t}, x_{2t}, ..., x_{nt}\}$ that does not belong to $\boldsymbol{q}_{\cdot t}$. Denote the projection residuals of y_t on $(\boldsymbol{q}'_{\cdot t}, x_t)'$ as $e_t = y_t - \boldsymbol{\gamma}'_{yqx,T}(\boldsymbol{q}'_{\cdot t}, x_t)'$. Define $\boldsymbol{x} = (x_1, x_2, ..., x_T)'$, $\boldsymbol{e} = (e_1, e_2, ..., e_T)'$, and $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$. Moreover, let $E(\boldsymbol{e}'\boldsymbol{e}/T) = \sigma^2_{e,(T)}$ and $E(\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}/T) = \sigma^2_{x,(T)}$. Then

$$\Pr\left[\left|\frac{a_T}{\sqrt{(T^{-1}\boldsymbol{e}'\boldsymbol{e})(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x})}}\right| > c_p(n,\delta)\right] \le \Pr\left(\left|\frac{a_T}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left[-C_0 T^{C_1}\right] \tag{B.45}$$

for any random variable a_T, some finite positive constants C_0 and C_1, and some bounded sequence $d_T > 0$, where $c_p(n,\delta)$ is defined in (15). Similarly,

$$\Pr\left[\left|\frac{a_T}{\sqrt{(T^{-1}\boldsymbol{e}'\boldsymbol{e})}}\right| > c_p(n,\delta)\right] \le \Pr\left(\left|\frac{a_T}{\sigma_{e,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left[-C_0 T^{C_1}\right] \tag{B.46}$$

Proof. We prove (B.45). (B.46) follows similarly. Define

$$g_T = [\sigma^2_{e,(T)}/(T^{-1}\boldsymbol{e}'\boldsymbol{e})]^{1/2} - 1,\ h_T = [\sigma^2_{x,(T)}/(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x})]^{1/2} - 1.$$

Using results in Lemma A11, note that for any $d_T > 0$ bounded in T,

$$\Pr\left[\left|\frac{a_T}{\sqrt{(T^{-1}\boldsymbol{e}'\boldsymbol{e})(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x})}}\right| > c_p(n,\delta)\,|\theta = 0\right] \le \Pr\left(\left|\frac{a_T}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \Pr\left(\left|(1+g_T)(1+h_T)\right| > 1+d_T\right).$$

Since $(1+g_T)(1+h_T) > 0$, then

$$\Pr\left(\left|(1+g_T)(1+h_T)\right| > 1+d_T\right) = \Pr\left[(1+g_T)(1+h_T) > 1+d_T\right] = \Pr\left(g_T h_T + g_T + h_T\right) > d_T).$$

Using (B.33), (B.36), (B.38) and (B.41),

$$\Pr\left[|h_T| > d_T\right] \le \exp\left[-C_0 T^{C_1}\right],\ \Pr\left[|h_T| > c\right] \le \exp\left[-C_0 T^{C_1}\right],$$
$$\Pr\left[|g_T| > d_T\right] \le \exp\left[-C_0 T^{C_1}\right],\ \Pr\left[|g_T| > d_T/c\right] \le \exp\left[-C_0 T^{C_1}\right],$$

for some finite positive constants C_0 and C_1. Using the above results, for some finite positive constants C_0 and C_1, we have,

$$\Pr\left[\left|\frac{a_T}{\sqrt{(T^{-1}\boldsymbol{e}'\boldsymbol{e})(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x})}}\right| > c_p(n,\delta)\,|\theta = 0\right] \le \Pr\left(\left|\frac{a_T}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left[-C_0 T^{C_1}\right],$$

which establishes the desired result. ■

Lemma A10 *Let y_t, for $t = 1, 2, ..., T$, be given by the data generating process (6) and suppose that u_t and $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$ satisfy Assumptions 2-4, with $s = \min(s_x, s_u) > 0$. Let $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ contain a constant and a subset of $\boldsymbol{x}_{nt}$, and let $\eta_t = \boldsymbol{x}'_{b,t}\boldsymbol{\beta}_b + u_t$, where $\boldsymbol{x}_{b,t}$ is $k_b \times 1$ dimensional vector of signal variables that do not belong to $\boldsymbol{q}_{\cdot t}$, with the associated coefficients, $\boldsymbol{\beta}_b$. Assume that $\boldsymbol{\Sigma}_{qq} = \frac{1}{T}\sum_{t=1}^{T} E\left(\boldsymbol{q}_{\cdot t}\boldsymbol{q}'_{\cdot t}\right)$ and $\hat{\boldsymbol{\Sigma}}_{qq} = \mathbf{Q}'\mathbf{Q}/T$ are both invertible, where $\mathbf{Q} = (\boldsymbol{q}_{1\cdot}, \boldsymbol{q}_{2\cdot}, ..., \boldsymbol{q}_{l_T\cdot})$ and $\boldsymbol{q}_{i\cdot} = (q_{i1}, q_{i2}, ..., q_{iT})'$, for $i = 1, 2, ..., l_T$. Moreover, let $l_T = o(T^{1/3})$ and suppose that Assumption 5 holds for x_{it} and $\boldsymbol{q}_{\cdot t}$, $i = 1, 2, ..., n$, where x_t is a generic element of $\{x_{1t}, x_{2t}, ..., x_{nt}\}$ that does not belong to $\boldsymbol{q}_{\cdot t}$. Denote the corresponding projection residuals defined by (11) as $u_{x,t} = x_t - \boldsymbol{\gamma}'_{qx,T}\boldsymbol{q}_{\cdot t}$, and the projection residuals of y_t on $(\boldsymbol{q}'_{\cdot t}, x_t)'$ as $e_t = y_t - \boldsymbol{\gamma}'_{yqx,T}(\boldsymbol{q}'_{\cdot t}, x_t)'$. Define $\boldsymbol{x} = (x_1, x_2, ..., x_T)'$, $\boldsymbol{y} = (y_1, y_2, ..., y_T)'$, $\boldsymbol{e} = (e_1, e_2, ..., e_T)'$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, and $\theta_T = E\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\right)\boldsymbol{\beta}_b$, where $\mathbf{X}_b$ is $T \times k_b$ matrix of observations on $\boldsymbol{x}_{b,t}$. Finally, $c_p(n, \delta)$ is given by (15) with $0 < p < 1$ and $f(n, \delta) = cn^{\delta}$, for some $c, \delta > 0$, and there exists $\kappa_1 > 0$ such that $T = \ominus\left(n^{\kappa_1}\right)$. Then, for any π in the range $0 < \pi < 1$, any $d_T > 0$ and bounded in T, and for some finite positive constants C_0 and C_1,*

$$\Pr\left[|t_x| > c_p(n, \delta)\,|\theta_T = 0\right] \le \exp\left[\frac{-(1-\pi)^2\sigma^2_{e,(T)}\sigma^2_{x,(T)}c_p^2(n, \delta)}{2(1+d_T)^2\omega^2_{xe,T}}\right] + \exp\left[-C_0T^{C_1}\right], \quad \text{(B.47)}$$

where

$$t_x = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}}, \quad \text{(B.48)}$$

$$\sigma^2_{e,(T)} = E\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right),\ \sigma^2_{x,(T)} = E\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right), \quad \text{(B.49)}$$

and

$$\omega^2_{xe,T} = T^{-1}\textstyle\sum_{t=1}^{T}E\left[(u_{x,t}\eta_t)^2\right]. \quad \text{(B.50)}$$

Under $\sigma_t^2 = \sigma^2$ and/or $E\left(u^2_{x,t}\right) = \sigma^2_{xt} = \sigma^2_x$, for all $t = 1, 2, ..., T$,

$$\Pr\left[|t_x| > c_p(n, \delta)\,|\theta_T = 0\right] \le \exp\left[-(1-\pi)^2 c_p^2(n, \delta)(1+d_T)^{-2}/2\right] + \exp\left(-C_0T^{C_1}\right). \quad \text{(B.51)}$$

In the case where $\theta_T \neq 0$, let $\theta_T = \ominus\left(T^{-\vartheta}\right)$, for some $0 \le \vartheta < 1/2$, where $c_p(n, \delta) = O\left(T^{1/2-\vartheta-C_8}\right)$, for some positive C_8. Then, for some bounded positive sequence d_T, and for some $C_2, C_3 > 0$, we have

$$\Pr\left[|t_x| > c_p(n, \delta)\,|\theta_T \neq 0\right] > 1 - \exp\left(-C_2T^{C_3}\right). \quad \text{(B.52)}$$

Proof. The DGP, given by (7), can be written as $\boldsymbol{y} = a\boldsymbol{\tau}_T + \mathbf{X}_k\boldsymbol{\beta} + \boldsymbol{u} = a\boldsymbol{\tau}_T + \mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}$, where $\mathbf{X}_a$ is a subset of $\mathbf{Q}$. Let $\mathbf{Q}_x = (\mathbf{Q}, \boldsymbol{x})$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, $\mathbf{M}_{qx} = \mathbf{I}_T - \mathbf{Q}_x(\mathbf{Q}'_x\mathbf{Q}_x)^{-1}\mathbf{Q}'_x$. Then, $\mathbf{M}_q\mathbf{X}_a = \mathbf{0}$, and let $\mathbf{M}_q\mathbf{X}_b = (\boldsymbol{x}_{bq,1}, \boldsymbol{x}_{bq,2}, ..., \boldsymbol{x}_{bq,T})'$. Then,

$$t_x = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}/\right)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}} = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b}{\sqrt{\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}} + \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sqrt{\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}}. \quad \text{(B.53)}$$

Let $\theta_T = E\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\right)\boldsymbol{\beta}_b$, $\boldsymbol{\eta} = \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}$, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_T)'$, and write (B.53) as

$$t_x = \frac{\sqrt{T}\theta_T}{\sqrt{\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}/\right)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}} + \frac{\sqrt{T}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta_T\right)}{\sqrt{\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}/\right)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}}. \tag{B.54}$$

First, consider the case where $\theta_T = 0$ and note that in this case
$t_x = \left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)^{-1/2}\left(T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}\right)\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right)^{-1/2}$. Now by Lemma A9, we have

$$\begin{aligned}\Pr\left[|t_x| > c_p(n,\delta)\,|\theta_T = 0\right] &= \Pr\left[\left|\frac{\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)^{-1/2}\left(T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}\right)}{\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right)^{1/2}}\right| > c_p(n,\delta)\,|\theta_T = 0\right] \\ &\leq \Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left(-C_0T^{C_1}\right).\end{aligned}$$

where $\sigma^2_{e,(T)}$ and $\sigma^2_{x,(T)}$ are defined by (B.49). Hence, noting that $c_p(n,\delta) = o(T^{C_0})$, for all $C_0 > 0$, under Assumption 3, and by Lemma A6, we have

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta_T = 0\right] \leq \exp\left[\frac{-(1-\pi)^2\sigma^2_{e,(T)}\sigma^2_{x,(T)}c_p^2(n,\delta)}{2(1+d_T)^2\omega^2_{xe,T}}\right] + \exp\left(-C_0T^{C_1}\right),$$

where $\omega^2_{xe,T} = T^{-1}\sum_{t=1}^{T}E[(u_{x,t}\eta_t)^2] = T^{-1}\sum_{t=1}^{T}E[u^2_{x,t}\left(\boldsymbol{x}'_{b,t}\boldsymbol{\beta}_b + u_t\right)^2]$, and $u_{x,t}$, being the error in the regression of x_t on $\mathbf{Q}$, is defined by (11). Since by assumption u_t are distributed independently of $u_{x,t}$ and $\boldsymbol{x}_{b,t}$, then

$$\omega^2_{xe,T} = T^{-1}\sum_{t=1}^{T}E[u^2_{x,t}(\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b)^2] + T^{-1}\sum_{t=1}^{T}E\left(u^2_{xt}\right)E\left(u^2_t\right),$$

where $\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b$ is the t-th element of $\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b$. Furthermore, $E[u^2_{x,t}(\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b)^2] = E\left(u^2_{x,t}\right)E(\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b)^2 = E\left(u^2_{x,t}\right)\boldsymbol{\beta}'_bE(\boldsymbol{x}_{bq,t}\boldsymbol{x}'_{bq,t})\boldsymbol{\beta}_b$, noting that under $\theta = 0$, $u_{x,t}$ and $\boldsymbol{x}_{b,t}$ are independently distributed. Hence

$$\omega^2_{xe,T} = T^{-1}\sum_{t=1}^{T}E\left(u^2_{x,t}\right)\boldsymbol{\beta}'_bE(\boldsymbol{x}_{bq,t}\boldsymbol{x}'_{bq,t})\boldsymbol{\beta}_b + T^{-1}\sum_{t=1}^{T}E\left(u^2_{xt}\right)E\left(u^2_t\right). \tag{B.55}$$

Similarly

$$\begin{aligned}\sigma^2_{e,(T)} &= E\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right) = E\left(T^{-1}\boldsymbol{\eta}'\mathbf{M}_{qx}\boldsymbol{\eta}\right) = E[T^{-1}\left(\mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}\right)'\mathbf{M}_{qx}\left(\mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}\right)] \\ &= \boldsymbol{\beta}'_bE\left(T^{-1}\mathbf{X}'_b\mathbf{M}_{qx}\mathbf{X}_b\right)\boldsymbol{\beta}_b + T^{-1}\sum_{t=1}^{T}E\left(u^2_t\right),\end{aligned}$$

and since under $\theta = 0$, $\boldsymbol{x}$ being a noise variable will be distributed independently of $\mathbf{X}_b$, then $E\left(T^{-1}\mathbf{X}'_b\mathbf{M}_{qx}\mathbf{X}_b\right) = E\left(T^{-1}\mathbf{X}'_b\mathbf{M}_q\mathbf{X}_b\right)$, and we have

$$\sigma^2_{e,(T)} = \boldsymbol{\beta}'_bE\left(T^{-1}\mathbf{X}'_b\mathbf{M}_q\mathbf{X}_b\right)\boldsymbol{\beta}_b + T^{-1}\sum_{t=1}^{T}E\left(u^2_t\right) = T^{-1}\sum_{t=1}^{T}\boldsymbol{\beta}'_bE\left(\boldsymbol{x}_{bq,t}\boldsymbol{x}'_{bq,t}\right)\boldsymbol{\beta}_b + T^{-1}\sum_{t=1}^{T}E\left(u^2_t\right). \tag{B.56}$$

Using (B.55) and (B.56), it is now easily seen that if either $E\left(u^2_{x,t}\right) = \sigma^2_{ux}$ or $E\left(u^2_t\right) = \sigma^2$, for all t, then we have $\omega^2_{xe,T} = \sigma^2_{e,(T)}\sigma^2_{x,(T)}$, and hence

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta_T = 0\right] \leq \exp\left[-(1-\pi)^2c_p^2(n,\delta)(1+d_T)^{-2}/2\right] + \exp\left(-C_0T^{C_1}\right),$$

giving a rate that does not depend on error variances. Next, we consider $\theta_T \neq 0$. By (B.45) of Lemma A9, for $d_T > 0$ and bounded in T,

$$\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{(T^{-1}\boldsymbol{e}'\boldsymbol{e}/)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}}\right| > c_p(n,\delta)\right] \leq \Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left(-C_0T^{C_1}\right).$$

We then have

$$\begin{aligned}\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sigma_{e,(T)}\sigma_{x,(T)}} &= \frac{T^{1/2}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b - \theta_T\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta_T}{\sigma_{e,(T)}\sigma_{x,(T)}} \\ &= \frac{T^{1/2}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta_T\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta_T}{\sigma_{e,(T)}\sigma_{x,(T)}}.\end{aligned}$$

Then $\Pr[|T^{1/2}\sigma_{e,(T)}^{-1}\sigma_{x,(T)}^{-1}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta_T\right) + T^{1/2}\sigma_{e,(T)}^{-1}\sigma_{x,(T)}^{-1}\theta_T| > c_p(n,\delta)/(1+d_T)] = 1 - \Pr\left[\left|T^{1/2}\sigma_{e,(T)}^{-1}\sigma_{x,(T)}^{-1}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta_T\right) + T^{1/2}\sigma_{e,(T)}^{-1}\sigma_{x,(T)}^{-1}\theta_T\right| \leq c_p(n,\delta)/(1+d_T)\right]$. Note that since $c_p(n,\delta)$ is given by (15), then, $T^{1/2}|\theta_T|/(\sigma_{e,(T)}\sigma_{x,(T)}) - c_p(n,\delta)/(1+d_T) > 0$. Then by Lemma A12,

$$\begin{aligned}&\Pr\left[\left|\frac{T^{1/2}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta_T\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta_T}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| \leq \frac{c_p(n,\delta)}{1+d_T}\right] \\ &\leq \Pr\left[\left|\frac{T^{1/2}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta_T\right)}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{T^{1/2}|\theta_T|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p(n,\delta)}{1+d_T}\right].\end{aligned}$$

But, setting $\zeta_T = T^{1/2}\left[T^{1/2}|\theta_T|/[\sigma_{e,(T)}\sigma_{x,(T)}] - c_p(n,\delta)/(1+d_T)\right]$ and noting that $\theta_T = O(T^{-\vartheta})$, $0 \leq \vartheta < 1/2$, implies that this choice of ζ_T satisfies $\zeta_T = \ominus\left(T^{\lambda}\right)$ with $\lambda = 1-\vartheta$, (B.16) of Lemma A6 applies regardless of $s > 0$, which gives us

$$\begin{aligned}&\Pr\left[\left|\frac{T^{1/2}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta_T\right)}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{T^{1/2}|\theta_T|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p(n,\delta)}{1+d_T}\right] \\ &\leq C_4\exp\left\{-C_5\left[T^{1/2}\left(\frac{T^{1/2}|\theta_T|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p(n,\delta)}{1+d_T}\right)\right]^{s/(s+2)}\right\} + \exp\left(-C_6T^{C_7}\right),\end{aligned} \tag{B.57}$$

for some C_4, C_5, C_6 and $C_7 > 0$. Hence, as long as the assumption that $c_p(n,\delta) = O\left(T^{1/2-\vartheta-C_8}\right)$ holds, for some positive C_8, there must exist positive finite constants C_2 and C_3, such that

$$\Pr\left[\left|\frac{T^{1/2}\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta} - \theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{T^{1/2}|\theta_T|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p(n,\delta)}{1+d_T}\right] \leq \exp\left(-C_2T^{C_3}\right) \tag{B.58}$$

for any $s > 0$. So overall

$$\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{(T^{-1}\boldsymbol{e}'\boldsymbol{e}/)\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right)}}\right| > c_p(n,\delta)\right] > 1 - \exp\left(-C_2T^{C_3}\right).$$

■

Lemma A11 *Let X_{iT}, for $i = 1, 2, ..., l_T$, Y_T and Z_T be random variables. Then, for some finite positive constants C_0, C_1 and C_2, and any constants π_i, for $i = 1, 2, ..., l_T$, satisfying $0 < \pi_i < 1$ and $\sum_{i=1}^{l_T} \pi_i = 1$, we have*

$$\Pr\left(\sum_{i=1}^{l_T} |X_{iT}| > C_0\right) \leq \sum_{i=1}^{l_T} \Pr\left(|X_{iT}| > \pi_i C_0\right), \tag{B.59}$$

$$\Pr\left(|X_T| \times |Y_T| > C_0\right) \leq \Pr\left(|X_T| > C_0/C_1\right) + \Pr\left(|Y_T| > C_1\right), \tag{B.60}$$

and

$$\begin{aligned}\Pr\left(|X_T| \times |Y_T| \times |Z_T| > C_0\right) \leq \Pr\left(|X_T| > C_0/\left(C_1 C_2\right)\right) + \Pr\left(|Y_T| > C_1\right) + \\ \Pr\left(|Z_T| > C_2\right).\end{aligned} \tag{B.61}$$

Proof. Without loss of generality we consider the case $l_T = 2$. Consider the two random variables X_{1T} and X_{2T}. Then, for some finite positive constants C_0 and C_1, we have

$$\begin{aligned}\Pr\left(|X_{1T}| + |X_{2T}| > C_0\right) &\leq \Pr\left(\{|X_{1T}| > (1-\pi)C_0\} \cup \{|X_{2T}| > \pi C_0\}\right) \\ &\leq \Pr\left(|X_{1T}| > (1-\pi)C_0\right) + \Pr\left(|X_{2T}| > \pi C_0\right),\end{aligned}$$

proving the first result of the lemma.

Define events $\mathfrak{H} = \{|X_T| \times |Y_T| > C_0\}$, $\mathfrak{B} = \{|X_T| > C_0/C_1\}$ and $\mathfrak{C} = \{|Y_T| > C_1\}$. Then $\mathfrak{H} \subset (\mathfrak{B} \cup \mathfrak{C})$, namely $\mathfrak{H}$ must be contained in $\mathfrak{B} \cup \mathfrak{C}$. Hence $P(\mathfrak{H}) \leq P(\mathfrak{B} \cup \mathfrak{C})$. But $P(\mathfrak{B} \cup \mathfrak{C}) \leq P(\mathfrak{B}) + P(\mathfrak{C})$. Therefore, $P(\mathfrak{H}) \leq P(\mathfrak{B}) + P(\mathfrak{C})$, proving the second result of the lemma. The third result follows by a repeated application of the second result. ■

Lemma A12 *Consider the scalar random variable X, and the constants B and C. Then, if $|B| \geq C > 0$,*

$$\Pr\left(|X + B| \leq C\right) \leq \Pr\left(|X| > |B| - C\right). \tag{B.62}$$

Proof. We note that the event we are concerned with is of the form $\mathcal{A} = \{|X + B| \leq C\}$. Consider two cases: (i) $B > 0$. Then, $\mathcal{A}$ can occur only if $X < 0$ and $|X| > B - C = |B| - C$. (ii) $B < 0$. Then, $\mathcal{A}$ can occur only if $X > 0$ and $X = |X| > |B| - C$. It therefore follows that the event $\{|X| > |B| - C\}$ implies $\mathcal{A}$ proving (B.62). ■

Lemma A13 *Consider the scalar random variable, ω_T, and the deterministic sequence, $\alpha_T > 0$, such that $\alpha_T \to 0$ as $T \to \infty$. Then there exists $T_0 > 0$ such that for all $T > T_0$ we have*

$$\Pr\left(\left|\frac{1}{\sqrt{\omega_T}} - 1\right| > \alpha_T\right) \leq \Pr\left(|\omega_T - 1| > \alpha_T\right). \tag{B.63}$$

Proof. We first note that for $\alpha_T < 1/2$

$$\left|\frac{1}{\sqrt{\omega_T}} - 1\right| < |\omega_T - 1| \text{ for any } \omega_T \in [1-\alpha_T, 1+\alpha_T].$$

Also, since $a_T \to 0$ then there must exist a $T_0 > 0$ such that $a_T < 1/2$, for all $T > T_0$, and hence if event $A : |\omega_T - 1| \leq a_T$ is satisfied, then it must be the case that event $B : \left|\frac{1}{\sqrt{\omega_T}} - 1\right| \leq a_T$ is also satisfied for all $T > T_0$. Further, since $A \Rightarrow B$, then $B^c \Rightarrow A^c$, where A^c denotes the complement of A. Therefore, $\left|\frac{1}{\sqrt{\omega_T}} - 1\right| > a_T$ implies $|\omega_T - 1| > a_T$, for all $T > T_0$, and we have $\Pr\left(\left|\frac{1}{\sqrt{\omega_T}} - 1\right| > \alpha_T\right) \leq \Pr(|\omega_T - 1| > \alpha_T)$, as required. ■

Lemma A14 *Let $\mathbf{A}_T = (a_{ij,T})$ be a symmetric $l_T \times l_T$ matrix with eigenvalues $\mu_1 \leq \mu_2 \leq ... \leq \mu_{l_T}$. Let $\mu_i = \ominus(l_T)$, $i = l_T - M + 1, l_T - M + 2, ..., l_T$, for some finite M, and $\sup_{1\leq i\leq l_T-M} \mu_i < C_0 < \infty$, for some finite positive C_0. Then,*

$$\|\mathbf{A}_T\|_F = \ominus(l_T). \tag{B.64}$$

If, in addition, $\inf_{1\leq i<l_T} \mu_i > C_1 > 0$, for some finite positive C_1, then

$$\left\|\mathbf{A}_T^{-1}\right\|_F = \ominus\left(\sqrt{l_T}\right). \tag{B.65}$$

Proof. We have

$$\|\mathbf{A}_T\|_F^2 = Tr(\mathbf{A}_T\mathbf{A}_T') = Tr\left(\mathbf{A}_T^2\right) = \sum_{i=1}^{l_T} \mu_i^2,$$

where μ_i, for $i = 1, 2, ..., l_T$, are the eigenvalues of $\mathbf{A}_T$. But by assumption $\mu_i = \ominus(l_T)$, for $i = l_T - M + 1, l_T - M + 2, ..., l_T$, and $\sup_{1\leq i\leq l_T-M} \mu_i < C_0 < \infty$, then $\sum_{i=1}^{l_T} \mu_i^2 = M \ominus(l_T^2) + O(l_T - M) = \ominus(l_T^2)$, and since M is fixed then (B.64) follows. Note that $\mathbf{A}_T^{-1}$ is also symmetric, and using similar arguments as above, we have

$$\left\|\mathbf{A}_T^{-1}\right\|_F^2 = Tr\left(\mathbf{A}_T^{-2}\right) = \sum_{i=1}^{l_T} \mu_i^{-2},$$

but all eigenvalues of $\mathbf{A}_T$ are bounded away from zero under the assumptions of the lemma, which implies $\mu_i^{-2} = \ominus(1)$ and therefore $\left\|\mathbf{A}_T^{-1}\right\|_F = \ominus\left(\sqrt{l_T}\right)$, which establishes (B.65). ■

Lemma A15 *Let z be a random variable and suppose there exists finite positive constants C_0, C_1 and $s > 0$ such that*

$$\Pr(|z| > \alpha) \leq C_0 \exp(-C_1\alpha^s), \text{ for all } \alpha > 0. \tag{B.66}$$

Then for any finite $p > 0$ and p/s finite, there exists $C_2 > 0$ such that

$$E|z|^p \leq C_2. \tag{B.67}$$

Proof. We have that

$$E|z|^p = \int_0^\infty \alpha^p d\Pr(|z| \leq \alpha).$$

Using integration by parts, we get

$$\int_0^\infty \alpha^p d\Pr(|z| \leq \alpha) = p\int_0^\infty \alpha^{p-1}\Pr(|z| > \alpha)\, d\alpha.$$

But, using (B.66), and a change of variables, implies

$$E|z|^p \leq pC_0\int_0^\infty \alpha^{p-1}\exp(-C_1\alpha^s)\, d\alpha = \frac{pC_0}{s}\int_0^\infty u^{\frac{p-s}{s}}\exp(-C_1 u)\, du = C_0 C_1^{-p/s}\left(\frac{p}{s}\right)\Gamma\left(\frac{p}{s}\right),$$

where $\Gamma(\cdot)$ is a gamma function. But for a finite positive p/s, $\Gamma(p/s)$ is bounded and (B.67) follows. ■

Lemma A16 *Let $\boldsymbol{A}_T = (a_{ij,T})$ be an $l_T \times l_T$ matrix and $\hat{\boldsymbol{A}}_T = (\hat{a}_{ij,T})$ be an estimator of $\boldsymbol{A}_T$. Suppose that $\boldsymbol{A}_T$ is invertible and there exists a finite positive C_0, such that*

$$\sup_{i,j}\Pr(|\hat{a}_{ij,T} - a_{ij,T}| > b_T) \leq \exp\left(-C_0 T b_T^2\right), \tag{B.68}$$

for all $b_T > 0$. Then

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > b_T\right) \leq l_T^2\exp\left(-C_0\frac{Tb_T^2}{l_T^2}\right), \tag{B.69}$$

and

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\|_F > b_T\right) \leq l_T^2\exp\left(\frac{-C_0 T b_T^2}{l_T^2\left\|\boldsymbol{A}_T^{-1}\right\|_F^2\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)^2}\right) + l_T^2\exp\left(-C_0\frac{T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 l_T^2}\right). \tag{B.70}$$

Proof. First note that since $b_T > 0$, then

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > b_T\right) = \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F^2 > b_T^2\right) = \Pr\left(\left[\sum_{j=1}^{l_T}\sum_{i=1}^{l_T}(\hat{a}_{ij,T} - a_{ij,T})^2 > b_T^2\right]\right),$$

and using the probability bound result, (B.59), and setting $\pi_i = 1/l_T$, we have

$$\begin{aligned}\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > b_T\right) &\leq \sum_{j=1}^{l_T}\sum_{i=1}^{l_T}\Pr\left(|\hat{a}_{ij,T} - a_{ij,T}|^2 > l_T^{-2}b_T^2\right)\\ &= \sum_{j=1}^{l_T}\sum_{i=1}^{l_T}\Pr\left(|\hat{a}_{ij,T} - a_{ij,T}| > l_T^{-1}b_T\right)\\ &\leq l_T^2\sup_{ij=1,2,\ldots,l_T}\left[\Pr\left(|\hat{a}_{ij,T} - a_{ij,T}| > l_T^{-1}b_T\right)\right].\end{aligned}$$

Hence by (B.68) we obtain (B.69). To establish (B.70) define the events

$$\mathcal{A}_T = \left\{ \left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F < 1 \right\} \text{ and } \mathcal{B}_T = \left\{ \left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\|_F > b_T \right\}$$

and note that by (2.15) of Berk (1974) if $\mathcal{A}_T$ holds we have

$$\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\|_F \leq \frac{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}{1 - \left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}. \tag{B.71}$$

Hence

$$\begin{aligned}\Pr\left(\mathcal{B}_T \left|\mathcal{A}_T\right.\right) &\leq \Pr\left(\frac{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}{1 - \left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F} > b_T\right) \\ &= \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F \left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right).\end{aligned} \tag{B.72}$$

Note also that

$$\Pr\left(\mathcal{B}_T\right) = \Pr\left(\left\{\mathcal{B}_T \cap \mathcal{A}_T\right\} \cup \left\{\mathcal{B}_T \cap \mathcal{A}_T^C\right\}\right) = \Pr\left(\mathcal{B}_T|\mathcal{A}_T\right)\Pr\left(\mathcal{A}_T\right) + \Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right)\Pr\left(\mathcal{A}_T^C\right). \tag{B.73}$$

Furthermore

$$\begin{aligned}\Pr\left(\mathcal{A}_T^C\right) &= \Pr\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > 1\right) \\ &= \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \left\|\boldsymbol{A}_T^{-1}\right\|_F^{-1}\right),\end{aligned}$$

and by (B.69) we have

$$\Pr\left(\mathcal{A}_T^C\right) \leq l_T^2 \exp\left(-C_0 \frac{T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 l_T^2}\right).$$

Using the above result and (B.72) in (B.73), we now have

$$\begin{aligned}\Pr\left(\mathcal{B}_T\right) \leq \Pr&\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F \left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right)\Pr\left(\mathcal{A}_T\right) \\ &+ \Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right) l_T^2 \exp\left(-C_0 \frac{T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 l_T^2}\right).\end{aligned}$$

Furthermore, since $\Pr\left(\mathcal{A}_T\right) \leq 1$ and $\Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right) \leq 1$ then

$$\begin{aligned}\Pr\left(\mathcal{B}_T\right) = \Pr\left(\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\|_F > b_T\right) &\leq \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F \left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right) \\ &+ l_T^2 \exp\left(-C_0 \frac{T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 l_T^2}\right).\end{aligned}$$

Result (B.70) now follows if we apply (B.69) to the first term on the RHS of the above. ■

Lemma A17 *Let $\boldsymbol{A}_T = (a_{ij,T})$ be a $l_T \times l_T$ matrix and $\hat{\boldsymbol{A}}_T = (\hat{a}_{ij,T})$ be an estimator of $\boldsymbol{A}_T$. Let $\left\|\boldsymbol{A}_T^{-1}\right\|_F > 0$ and suppose that for some $s > 0$, any $b_T > 0$ and some finite positive constant C_0,*

$$\sup_{i,j} \Pr\left(|\hat{a}_{ij,T} - a_{ij,T}| > b_T\right) \leq \exp\left[-C_0 \left(Tb_T\right)^{s/(s+2)}\right].$$

Then

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\|_F > b_T\right) \leq l_T^2 \exp\left(\frac{-C_0\left(Tb_T\right)^{s/(s+2)}}{l_T^{s/(s+2)}\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)}\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)^{s/(s+2)}}\right) \tag{B.74}$$
$$+ l_T^2 \exp\left(-C_0 \frac{T^{s/(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right).$$

Proof. First note that since $b_T > 0$, then

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > b_T\right) = \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F^2 > b_T^2\right)$$
$$= \Pr\left[\sum_{j=1}^{l_T}\sum_{i=1}^{l_T}\left(\hat{a}_{ij,T} - a_{ij,T}\right)^2 > b_T^2\right],$$

and using the probability bound result, (B.59), and setting $\pi_i = 1/l_T^2$, we have

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > b_T\right) \leq \sum_{j=1}^{l_T}\sum_{i=1}^{l_T} \Pr\left(\left|\hat{a}_{ij,T} - a_{ij,T}\right|^2 > l_T^{-2} b_T^2\right) \tag{B.75}$$
$$= \sum_{j=1}^{l_T}\sum_{i=1}^{l_T} \Pr\left(\left|\hat{a}_{ij,T} - a_{ij,T}\right| > l_T^{-1} b_T\right)$$
$$\leq l_T^2 \sup_{ij}\left[\Pr\left(\left|\hat{a}_{ij,T} - a_{ij,T}\right| > l_T^{-1} b_T\right)\right] = l_T^2 \exp\left(-C_0 T^{s/(s+1)} \frac{b_T^{s/(s+2)}}{l_t^{s/(s+2)}}\right).$$

To establish (B.74) define the events

$$\mathcal{A}_T = \left\{\left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F < 1\right\} \text{ and } \mathcal{B}_T = \left\{\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\| > b_T\right\}$$

and note that by (2.15) of Berk (1974) if $\mathcal{A}_T$ holds we have

$$\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\| \leq \frac{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}{1 - \left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}.$$

Hence

$$\Pr\left(\mathcal{B}_T \left|\mathcal{A}_T\right.\right) \leq \Pr\left(\frac{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2 \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}{1 - \left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F} > b_T\right)$$
$$= \Pr\left[\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right].$$

Note also that

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

Furthermore

$$\begin{aligned}\Pr\left(\mathcal{A}_T^C\right) &= \Pr\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F \left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > 1\right) \\ &= \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \left\|\boldsymbol{A}_T^{-1}\right\|_F^{-1}\right),\end{aligned}$$

and by (B.75) we have

$$\Pr\left(\mathcal{A}_T^C\right) \leq l_T^2 \exp\left(-C_0 T^{s/(s+1)} \frac{b_T^{s/(s+2)}}{l_t^{s/(s+2)}}\right) = \exp\left(-C_0 \frac{T^{s/(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right).$$

Using the above result, we now have

$$\begin{aligned}\Pr\left(\mathcal{B}_T\right) \leq &\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F \left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right)\Pr\left(\mathcal{A}_T\right) \\ &+ \Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right)\exp\left(-C_0 \frac{T^{s/(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right).\end{aligned}$$

Furthermore, since $\Pr\left(\mathcal{A}_T\right) \leq 1$ and $\Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right) \leq 1$ then

$$\begin{aligned}\Pr\left(\mathcal{B}_T\right) = \Pr\left(\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\| > b_T\right) \leq &\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F \left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right) \\ &+ \exp\left(-C_0 \frac{T^{s/(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right).\end{aligned}$$

Result (B.74) now follows if we apply (B.75) to the first term on the RHS of the above. ■

Lemma A18 *Let $\mathbf{S}_a$ and $\mathbf{S}_b$, respectively, be $T \times l_{a,T}$ and $T \times l_{b,T}$ matrices of observations on $s_{a,it}$, and $s_{b,it}$, for $i = 1, 2, ..., l_T$, $t = 1, 2, ..., T$, and suppose that $\{s_{a,it}, s_{b,it}\}$ are either non-stochastic and bounded, or random with finite 8^{th} order moments. Consider the sample covariance matrix $\hat{\mathbf{\Sigma}}_{ab} = T^{-1}\mathbf{S}_a'\mathbf{S}_b$ and denote its expectations by $\mathbf{\Sigma}_{ab} = T^{-1}E\left(\mathbf{S}_a'\mathbf{S}_b\right)$. Let*

$$z_{ij,t} = s_{a,it}s_{b,jt} - E\left(s_{a,it}s_{b,jt}\right),$$

and suppose that

$$\sup_{i,j}\left[\sum_{t=1}^{T}\sum_{t'=1}^{T} E(z_{ij,t}z_{ij,t'})\right] = O\left(T\right). \tag{B.76}$$

Then,

$$E\left\|\hat{\mathbf{\Sigma}}_{ab} - \mathbf{\Sigma}_{ab}\right\|_F^2 = O\left(\frac{l_{a,T}l_{b,T}}{T}\right). \tag{B.77}$$

If, in addition,

$$\sup_{i,j,i',j'}\left[\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}E(z_{ij,t}z_{ij,t'}z_{i'j',s}z_{i'j',s'})\right]=O\left(T^{2}\right), \tag{B.78}$$

then

$$E\left\|\hat{\boldsymbol{\Sigma}}_{ab}-\boldsymbol{\Sigma}_{ab}\right\|_F^4=O\left(\frac{l_{a,T}^2 l_{b,T}^2}{T^2}\right). \tag{B.79}$$

Proof. We first note that $E(z_{ij,t}z_{ij,t'})$ and $E\left(z_{ij,t}z_{ij,t'}z_{i'j',s}z_{i'j',s'}\right)$ exist since by assumption $\{s_{a,it}, s_{b,it}\}$ have finite 8^{th} order moments. The (i,j) element of $\hat{\boldsymbol{\Sigma}}_{ab}-\boldsymbol{\Sigma}_{ab}$ is given by

$$a_{ij,T}=T^{-1}\sum_{t=1}^{T}z_{ij,t}, \tag{B.80}$$

and hence

$$E\left\|\hat{\boldsymbol{\Sigma}}_{ab}-\boldsymbol{\Sigma}_{ab}\right\|_F^2=\sum_{i=1}^{l_{a,T}}\sum_{j=1}^{l_{b,T}}E\left(a_{ij,T}^2\right)=T^{-2}\sum_{i=1}^{l_{a,T}}\sum_{j=1}^{l_{b,T}}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(z_{ij,t}z_{ij,t'}\right)$$
$$\leq\frac{l_{a,T}l_{b,T}}{T^2}\sup_{i,j}\left[\sum_{t=1}^{T}\sum_{t'=1}^{T}E(z_{ij,t}z_{ij,t'})\right],$$

and (B.77) follows from (B.76). Similarly,

$$\left\|\hat{\boldsymbol{\Sigma}}_{ab}-\boldsymbol{\Sigma}_{ab}\right\|_F^4=\left(\sum_{i=1}^{l_{a,T}}\sum_{j=1}^{l_{b,T}}a_{ij,T}^2\right)^2$$
$$=\sum_{i=1}^{l_{a,T}}\sum_{j=1}^{l_{b,T}}\sum_{i'=1}^{l_{a,T}}\sum_{j'=1}^{l_{b,T}}a_{ij,T}^2a_{i'j',T}^2.$$

But using (B.80) we have

$$a_{ij,T}^2a_{i'j',T}^2=T^{-4}\left(\sum_{t=1}^{T}\sum_{t'=1}^{T}z_{ij,t}z_{ij,t'}\right)\left(\sum_{s=1}^{T}\sum_{s'=1}^{T}z_{i'j',s}z_{i'j',s'}\right)$$
$$=T^{-4}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}z_{ij,t}z_{ij,t'}z_{i'j',s}z_{i'j',s'},$$

and

$$E\left\|\hat{\boldsymbol{\Sigma}}_{ab}-\boldsymbol{\Sigma}_{ab}\right\|_F^4=T^{-4}\sum_{i=1}^{l_{a,T}}\sum_{j=1}^{l_{b,T}}\sum_{i'=1}^{l_{a,T}}\sum_{j'=1}^{l_{b,T}}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}E\left(z_{ij,t}z_{ij,t'}z_{i'j',s}z_{i'j',s'}\right)$$
$$\leq\frac{l_{a,T}^2l_{b,T}^2}{T^4}\sup_{i,j,i',j'}\left[\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}E\left(z_{ij,t}z_{ij,t'}z_{i'j',s}z_{i'j',s'}\right)\right].$$

Result (B.79) now follows from (B.78). ■

Remark 1 *It is clear that conditions (B.76) and (B.78) are met under Assumption 3 that requires* z_{it} *to be a martingale difference process. But it is easily seen that condition (B.76) also follows if we assume that* $s_{a,it}$ *and* $s_{b,jt}$ *are stationary processes with finite* 8*-th moments, since the product of stationary processes is also a stationary process under a certain additional cross-moment conditions (Wecker (1978)). The results of the lemma also follow readily if we assume that* $s_{a,it}$ *and* $s_{b,jt'}$ *are independently distributed for all* $i \neq j$ *and all* t *and* t'.

Lemma A19 *Consider the data generating process (6) with* k *signal variables,* k^* *pseudo-signal variables, and* $n-k-k^*$ *noise variables. Let* $\hat{k}^o_{(s)}$ *be the number of variables selected at the stage* s *of the OCMT procedure and suppose that conditions of Lemma A10 hold. Let* $k^* = \ominus(n^{\epsilon})$ *for some* $0 \leq \epsilon < \min\{1, \kappa_1/3\}$*, where* κ_1 *is the positive constant that defines the rate for* $T = \ominus(n^{\kappa_1})$ *in Lemma A10. Let* $\mathcal{D}_{s,T}$*, be the event that the number of variables selected in the first* s *stages of OCMT is smaller than or equal to* l_T*, where* $l_T = \ominus(n^{\nu})$ *and* ν *satisfies* $\epsilon < \nu < \kappa_1/3$*. Then there exist constants* $C_0, C_1 > 0$ *such that for any* $0 < \varkappa < 1$*, any* $\delta_s > 0$*, and any* $j > 0$*, it follows that*

$$\Pr\left(\hat{k}^o_{(s)} - k - k^* > j|\mathcal{D}_{s-1,T}\right) \leq \frac{n-k-k^*}{j}\left\{\exp\left[-\frac{\varkappa c_p^2(n,\delta_s)}{2}\right] + \exp(-C_0 T^{C_1})\right\}, \quad \text{(B.81)}$$

for $s = 1, 2, ..., k$.

Proof. By convention, the number of variables selected at the stage zero of OCMT is zero. Conditioning on $\mathcal{D}_{s-1,T}$ allows the application of Lemma A10. We drop the conditioning notation in the rest of the proof to simplify notations. Then, by Markov's inequality

$$\Pr\left(\hat{k}^o_{(s)} - k - k^* > j\right) \leq \frac{E\left(\hat{k}^o_{(s)} - k - k^*\right)}{j}. \quad \text{(B.82)}$$

But

$$\begin{aligned} E\left(\hat{k}^o_{(s)}\right) &= \sum_{i=1}^{n} E\left[I_{(s)}\widehat{(\beta_i \neq 0)}\right] \\ &= \sum_{i=1}^{k+k^*} E\left[I_{(s)}\widehat{(\beta_i \neq 0)}\right] + \sum_{i=k+k^*+1}^{n} E\left[I_{(s)}\widehat{(\beta_i \neq 0)}|\theta_{i,(s)} = 0\right]. \\ &\leq k + k^* + \sum_{i=k+k^*+1}^{n} E\left[I_{(s)}\widehat{(\beta_i \neq 0)}|\theta_{i,(s)} = 0\right], \end{aligned}$$

where we have used $I_{(s)}\widehat{(\beta_i \neq 0)} \leq 1$. Moreover,

$$E\left[I_{(s)}\widehat{(\beta_i \neq 0)}|\theta_{i,(s)} = 0\right] = \Pr\left(\left|t_{\hat{\phi}_{T,i,(s)}}\right| > c_p(n,\delta_s)|\theta_{i,(s)} = 0\right),$$

for $i = k+k^*+1, k+k^*+2, ..., n$, and using (B.51) of Lemma A10, we have (for some $0 < \varkappa < 1$ and $C_0, C_1 > 0$)

$$\sup_{i>k+k^*} \Pr\left(\left|t_{\hat{\phi}_{T,i,(s)}}\right| > c_p\left(n, \delta_s\right) |\theta_{i,(s)} = 0\right) \leq \exp\left[-\frac{\varkappa c_p^2\left(n, \delta_s\right)}{2}\right] + \exp(-C_0 T^{C_1}).$$

Hence,

$$E\left(\hat{k}_{(s)}^o\right) - k - k^* \leq (n - k - k^*)\left\{\exp\left[-\frac{\varkappa c_p^2\left(n, \delta_s\right)}{2}\right] + \exp(-C_0 T^{C_1})\right\},$$

and therefore (using this result in (B.82))

$$\Pr\left(\hat{k}_{(s)}^o - k - k^* > j\right) \leq \frac{n - k - k^*}{j}\left\{\exp\left[-\frac{\varkappa c_p^2\left(n, \delta_s\right)}{2}\right] + \exp(-C_0 T^{C_1})\right\},$$

as desired. ■

Lemma A20 *Consider the data generating process (6) with k signal, k^* pseudo-signal, and $n-k-k^*$ noise variables. Let $\mathcal{T}_k$ be the event that the OCMT procedure stops after k stages or less, and suppose that conditions of Lemma A10 hold. Let $k^* = \ominus\left(n^\epsilon\right)$ for some $0 \leq \epsilon <$ $\min\{1, \kappa_1/3\}$, where κ_1 is the positive constant that defines the rate for $T = \ominus\left(n^{\kappa_1}\right)$ in Lemma A10. Moreover, let $\delta > 0$ and $\delta^* > 0$ denote the critical value exponents for stage 1 and subsequent stages of the OCMT procedure, respectively. Then,*

$$\Pr\left(\mathcal{T}_k\right) = 1 + O\left(n^{1-\nu-\varkappa\delta}\right) + O\left(n^{1-\varkappa\delta^*}\right) + O\left[n \exp\left(-C_0 n^{C_1\kappa_1}\right)\right], \tag{B.83}$$

for some $C_0, C_1 > 0$, any $\varkappa$ in $0 < \varkappa < 1$, and any ν in $\epsilon < \nu < \kappa_1/3$.

Proof. Consider the event $\mathcal{D}_{k,T} = \{\hat{k}_{(j)} \leq l_T, j = 1, 2, ..., k\}$ for $k \geq 1$, which is the event that the number of variables selected in the first k stages of OCMT is smaller than or equal to $l_T = \ominus\left(n^\nu\right)$, where ν lies in the interval $\epsilon < \nu < \kappa_1/3$. Such a ν exists since by assumption $0 \leq \epsilon < \min\{1, \kappa_1/3\}$. We have $\Pr\left(\mathcal{T}_k\right) = 1 - \Pr\left(\mathcal{T}_k^c\right)$, and

$$\begin{aligned}\Pr\left(\mathcal{T}_k^c\right) &= \Pr\left(\mathcal{T}_k^c|\mathcal{D}_{k,T}\right)\Pr\left(\mathcal{D}_{k,T}\right) + \Pr\left(\mathcal{T}_k^c|\mathcal{D}_{k,T}^c\right)\Pr\left(\mathcal{D}_{k,T}^c\right) \\ &\leq \Pr\left(\mathcal{T}_k^c|\mathcal{D}_{k,T}\right) + \Pr\left(\mathcal{D}_{k,T}^c\right),\end{aligned}$$

Therefore,

$$\Pr\left(\mathcal{T}_k\right) \geq 1 - \Pr\left(\mathcal{T}_k^c|\mathcal{D}_{k,T}\right) - \Pr\left(\mathcal{D}_{k,T}^c\right). \tag{B.84}$$

We note that

$$\Pr\left(\mathcal{D}_{k,T}\right) \geq \Pr\left[\left(\hat{k}_{(1)}^o \leq \frac{l_T}{k}\right) \cap \left(\hat{k}_{,(2)}^o \leq \frac{l_T}{k}\middle|\mathcal{D}_{1,T}\right) \cap ... \cap \left(\hat{k}_{(k)}^o \leq \frac{l_T}{k}\middle|\mathcal{D}_{k-1,T}\right)\right],$$

where $\hat{k}^o_{(s)}$ is the number of variables selected in the s-th stage of OCMT and $\mathcal{D}_{s,T} = \{\hat{k}_{(j)} \leq l_T, j = 1, 2, ..., s\}$ for $s = 1, 2, ..., k$. Hence

$$\Pr\left(\mathcal{D}^c_{k,T}\right) \leq \Pr\left\{\left[\begin{array}{c} \left(\hat{k}^o_{(1)} \leq \frac{l_T}{k}\right) \cap \left(\hat{k}^o_{(2)} \leq \frac{l_T}{k}\middle| \mathcal{D}_{1,T}\right) \cap ... \\ \cap \left(\hat{k}^o_{(k)} \leq \frac{l_T}{k}\middle| \mathcal{D}_{k-1,T}\right) \end{array}\right]^c\right\}.$$

Furthermore

$$\begin{aligned} &\Pr\left\{\left[\begin{array}{c} \left(\hat{k}^o_{(1)} \leq \frac{l_T}{k}\right) \cap \left(\hat{k}^o_{(2)} \leq \frac{l_T}{k}\middle| \mathcal{D}_{1,T}\right) \cap ... \\ \cap \left(\hat{k}^o_{(k)} \leq \frac{l_T}{k}\middle| \mathcal{D}_{k-1,T}\right) \end{array}\right]^c\right\} \\ &= \Pr\left\{\left[\begin{array}{c} \left(\hat{k}^o_{(1)} > \frac{l_T}{k}\right) \cup \left(\hat{k}^o_{(2)} > \frac{l_T}{k}\middle| \mathcal{D}_{1,T}\right) \cup ... \\ \cup \left(\hat{k}^o_{(k)} > \frac{l_T}{k}\middle| \mathcal{D}_{k-1,T}\right) \end{array}\right]\right\} \\ &\leq \Pr\left(\hat{k}^o_{(1)} > \frac{l_T}{k}\right) + \sum_{s=2}^{k} \Pr\left(\hat{k}^o_{(s)} > \frac{l_T}{k}\middle| \mathcal{D}_{s-1,T}\right). \end{aligned}$$

Since k is finite and $0 \leq \epsilon < \nu$, there exists T_0 such that for all $T > T_0$ we have $l_T/k > k + k^*$, and we can apply (B.81) of Lemma A19 (for $j = l_T/k - k - k^* > 0$), to obtain

$$\begin{aligned} \Pr\left(\hat{k}^o_{(1)} > \frac{l_T}{k}\right) &= \Pr\left(\hat{k}^o_{(1)} - k - k^* > \frac{l_T}{k} - k - k^*\right) \\ &\leq \frac{n - k - k^*}{\frac{l_T}{k} - k - k^*}\left\{\exp\left[-\frac{\varkappa c_p^2(n, \delta)}{2}\right] + \exp(-C_0 T^{C_1})\right\}, \end{aligned}$$

for some $C_0, C_1 > 0$ and any $0 < \varkappa < 1$. Noting that for $0 \leq \epsilon < \nu$,

$$\frac{n - k - k^*}{\frac{l_T}{k} - k - k^*} = \ominus\left(n^{1-\nu}\right), \tag{B.85}$$

and using also result (ii) of Lemma A2, we obtain

$$\Pr\left(\hat{k}^o_{(1)} > \frac{l_T}{k}\right) = O\left(n^{1-\nu-\varkappa\delta}\right) + O\left[n^{1-\nu}\exp\left(-C_0 T^{C_1}\right)\right].$$

Similarly,

$$\begin{aligned} \Pr\left(\hat{k}^o_{(s)} > \frac{l_T}{k}\middle| \mathcal{D}_{s-1,T}\right) &= \Pr\left(\hat{k}^o_{(s)} - k - k^* > \frac{l_T}{k} - k - k^*\middle| \mathcal{D}_{s-1,T}\right) \\ &\leq \frac{n - k - k^*}{\frac{l_T}{k} - k - k^*}\left\{\exp\left[-\frac{\varkappa c_p^2(n, \delta^*)}{2}\right] + \exp(-C_0 T^{C_1})\right\} \\ &= O\left(n^{1-\nu-\varkappa\delta^*}\right) + O\left[n^{1-\nu}\exp\left(-C_0 T^{C_1}\right)\right], \end{aligned}$$

where the critical value exponent in the higher stages ($s > 1$) of OCMT (δ^*) could differ from the one in the first stage (δ). So, overall

$$\begin{aligned} \Pr\left(\mathcal{D}^c_{k,T}\right) &\leq \Pr\left(\hat{k}^o_{(1)} > \frac{l_T}{k}\right) + \sum_{s=2}^{k} \Pr\left(\hat{k}^o_{(s)} > \frac{l_T}{k}\middle| \mathcal{D}_{s-1,T}\right) \\ &= O\left(n^{1-\nu-\varkappa\delta}\right) + O\left(n^{1-\nu-\varkappa\delta^*}\right) + O\left[n^{1-\nu}\exp\left(-C_0 T^{C_1}\right)\right], \end{aligned} \tag{B.86}$$

for some $C_0, C_1 > 0$, any $\varkappa$ in $0 < \varkappa < 1$, and any ν in $\epsilon < \nu < \kappa_1/3$. Next, consider $\Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T})$, and note that

$$\Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}) = \Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k)\Pr(\mathcal{L}_k|\mathcal{D}_{k,T}) + \Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k^c)\Pr(\mathcal{L}_k^c|\mathcal{D}_{k,T}) \\ \leq \Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k) + \Pr(\mathcal{L}_k^c|\mathcal{D}_{k,T}), \tag{B.87}$$

where $\Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k)$ is the probability that a noise variable will be selected in a stage of OCMT that includes as regressors all signal variables, conditional on the event that fewer than l_T variables are selected in the first k steps of OCMT. Note that the event $\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k$ can only occur if OCMT selects some pseudo-signal and/or some noise variables in stage $k+1$. But the net effect coefficient of signal variables in stage $k+1$ must be zero when all signal variables were selected in earlier stages ($s = 1, 2, ..., k$), namely $\theta_{i,(k+1)} = 0$ for $i = k+1, k+2, ..., k+k^*$. Moreover, $\theta_{i,(k+1)} = 0$ also for $i = k+k^*+1, k+k^*+2, ..., n$, since the net effect coefficient of noise variables is always zero (in any stage). Therefore, we have

$$\Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k) \leq \sum_{i=k+1}^{n} \Pr\left[\left|t_{\hat{\phi}_{i,(k+1)}}\right| > c_p(n, \delta^*)\,|\theta_{i,(k+1)} = 0, \mathcal{D}_{k,T}\right].$$

Note that the number of regressors in the regressions involving the t statistics $t_{\hat{\theta}_{i,(k+1)}}$, does not exceed $l_T = \ominus(n^\nu)$, for ν in the interval $0 \leq \epsilon < \nu < \kappa_1/3$ and hence $l_T = o(T^{1/3})$ as required by the conditions of Lemma A10. Using (B.51) of Lemma A10, we have

$$\Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k) \leq (n-k)\exp\left[\frac{-\varkappa c_p^2(n, \delta^*)}{2}\right] \\ + (n-k)\exp\left(-C_0 T^{C_1}\right). \tag{B.88}$$

for some $C_0, C_1 > 0$ and any $0 < \varkappa < 1$. By Lemma A2, $\exp\left[-\varkappa c_p^2(n, \delta^*)/2\right] = \ominus\left(n^{-\varkappa\delta^*}\right)$, for any $0 < \varkappa < 1$, and noting that $n - k \leq n$ we obtain

$$\Pr(\mathcal{T}_k^c|\mathcal{D}_{k,T}, \mathcal{L}_k) = O\left(n^{1-\varkappa\delta^*}\right) + O\left[n\exp\left(-C_0 T^{C_1}\right)\right]. \tag{B.89}$$

Consider next the second term of (B.87), $\Pr(\mathcal{L}_k^c|\mathcal{D}_{k,T})$, and recall that $\mathcal{L}_k = \cap_{i=1}^k \mathcal{L}_{i,k}$ where $\mathcal{L}_{i,k} = \cup_{j=1}^k \mathcal{B}_{i,j}$, $i = 1, 2, ..., k$. Hence $\mathcal{L}_{i,k}^c = \cap_{j=1}^k \mathcal{B}_{i,j}^c$, and

$$\Pr\left(\mathcal{L}_{i,k}^c\middle|\,\mathcal{T}_k, \mathcal{D}_{k,T}\right) = \Pr\left(\cap_{j=1}^k \mathcal{B}_{i,j}^c\middle|\,\mathcal{T}_k, \mathcal{D}_{k,T}\right) = \\ \Pr\left(\mathcal{B}_{i,1}^c\middle|\,\mathcal{T}_k, \mathcal{D}_{k,T}\right)\Pr\left(\mathcal{B}_{i,2}^c\middle|\,\mathcal{B}_{i,1}^c, \mathcal{T}_k, \mathcal{D}_{k,T}\right) \\ \Pr\left(\mathcal{B}_{i,3}^c\middle|\,\mathcal{B}_{i,2}^c \cap \mathcal{B}_{i,1}^c, \mathcal{T}_k, \mathcal{D}_{k,T}\right) \times ... \times \\ \Pr\left(\mathcal{B}_{i,k}^c\middle|\,\mathcal{B}_{i,k-1}^c \cap ... \cap \mathcal{B}_{i,1}^c, \mathcal{T}_k, \mathcal{D}_{k,T}\right).$$

But by Proposition 1 we are guaranteed that for some $1 \leq j \leq k$, $\theta_{i,(j)} \neq 0$. Therefore,

$$\Pr\left(\mathcal{B}_{i,j}^c\middle|\,\mathcal{B}_{i,j-1}^c \cap ... \cap \mathcal{B}_{i,1}^c, \mathcal{T}_k, \mathcal{D}_{k,T}\right) = \Pr\left(\mathcal{B}_{i,j}^c\middle|\,\mathcal{B}_{i,j-1}^c \cap ... \cap \mathcal{B}_{i,1}^c, \theta_{i,(j)} \neq 0, \mathcal{T}_k, \mathcal{D}_{k,T}\right),$$

and by (B.52) of Lemma A10,

$$\Pr\left(\mathcal{B}_{i,j}^c\middle|\mathcal{B}_{i,j-1}^c\cap...\cap\mathcal{B}_{i,1}^c,\theta_{i,(j)}\neq 0,\mathcal{T}_k,\mathcal{D}_{k,T}\right)=O\left[\exp\left(-C_0T^{C_1}\right)\right],$$

for some $C_0, C_1 > 0$. Therefore, for some $j \in \{1, 2, ..., k\}$ and $C_0, C_1 > 0$,

$$\begin{aligned}\Pr\left(\mathcal{L}_{i,k}^c\middle|\mathcal{T}_k,\mathcal{D}_{k,T}\right)&\leq\Pr\left(\mathcal{B}_{i,j}^c\middle|\mathcal{B}_{i,j-1}^c\cap...\cap\mathcal{B}_{i,1}^c,\theta_{i,(j)}\neq 0,\mathcal{T}_k,\mathcal{D}_{k,T}\right)\\&=O\left[\exp\left(-C_0T^{C_1}\right)\right].\end{aligned}\tag{B.90}$$

Noting that k is finite and

$$\begin{aligned}\Pr\left(\mathcal{L}_k^c\middle|\mathcal{T}_k,\mathcal{D}_{k,T}\right)&=\Pr\left(\cup_{i=1}^k\mathcal{L}_{ik}^c\middle|\mathcal{T}_k,\mathcal{D}_{k,T}\right)\\&\leq\sum_{i=1}^k\Pr\left(\mathcal{L}_{ik}^c\middle|\mathcal{T}_k,\mathcal{D}_{k,T}\right),\end{aligned}$$

it follows, using (B.90), that

$$\Pr\left(\mathcal{L}_k^c\middle|\mathcal{T}_k,\mathcal{D}_{k,T}\right)=O\left[\exp\left(-C_0T^{C_1}\right)\right],\tag{B.91}$$

for some $C_0, C_1 > 0$. Using (B.89) and (B.91) in (B.87) now gives[5]

$$\Pr\left(\mathcal{T}_k^c|\mathcal{D}_{k,T}\right)=O\left(n^{1-\varkappa\delta^*}\right)+O\left[n\exp\left(-C_0T^{C_1}\right)\right].\tag{B.92}$$

Using (B.86) and (B.92) in (B.84), yields

$$\begin{aligned}\Pr\left(\mathcal{T}_k\right)=&1+O\left(n^{1-\nu-\varkappa\delta}\right)+O\left(n^{1-\nu-\varkappa\delta^*}\right)+O\left[n^{1-\nu}\exp\left(-C_0T^{C_1}\right)\right]\\&+O\left(n^{1-\varkappa\delta^*}\right)+O\left[n\exp\left(-C_2T^{C_3}\right)\right],\end{aligned}$$

for some $C_0, C_1, C_2, C_3 > 0$ and any $\varkappa$ in $0 < \varkappa < 1$, and any ν in $\epsilon < \nu < \kappa_1/3$. But $O\left(n^{1-\nu-\varkappa\delta^*}\right)$ is dominated by $O\left(n^{1-\varkappa\delta^*}\right)$, and $O\left[n^{1-\nu}\exp\left(-C_0T^{C_1}\right)\right]$ is dominated by $O\left[n\exp\left(-C_2T^{C_3}\right)\right]$, since $\nu > \epsilon \geq 0$. Hence,

$$\Pr\left(\mathcal{T}_k\right)=1+O\left(n^{1-\nu-\varkappa\delta}\right)+O\left(n^{1-\varkappa\delta^*}\right)+O\left[n\exp\left(-C_0T^{C_1}\right)\right],$$

for some $C_0, C_1 > 0$, any $\varkappa$ in $0 < \varkappa < 1$, and any ν in $\epsilon < \nu < \kappa_1/3$. This result in turn establishes (B.83), noting that $T = \ominus\left(n^{\kappa_1}\right)$. ■

Lemma A21 *Suppose that the data generating process (DGP) is given by*

$$\underset{T\times 1}{\boldsymbol{y}}=\underset{T\times k+1}{\mathbf{X}}\cdot\underset{k+1\times 1}{\boldsymbol{\beta}}+\underset{T\times 1}{\boldsymbol{u}},\tag{B.93}$$

where $\boldsymbol{u} = (u_1, u_2, ..., u_T)'$, $E(\boldsymbol{u}) = \mathbf{0}$, $E(\boldsymbol{u}\boldsymbol{u}') = \sigma^2\mathbf{I}_T$, $0 < \sigma^2 < \infty$, $\mathbf{I}_T$ *is a* $T \times T$ *identity matrix,* $\mathbf{X} = (\boldsymbol{\tau}_T, \mathbf{X}_k) = (\boldsymbol{\tau}_T, \boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k)$ *includes a* $T \times 1$ *column of ones,* $\boldsymbol{\tau}_T$, *and* $T \times 1$

[5] We have dropped the term $O\left[\exp\left(-C_0T^{C_1}\right)\right]$, which is dominated by $O\left[n\exp\left(-C_0T^{C_1}\right)\right]$.

vectors of observations, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$*, on the signal variables* $i = 1, 2, ..., k$*, and the elements of* $\boldsymbol{\beta}$ *are bounded. Consider the regression model*

$$\underset{T\times 1}{\boldsymbol{y}} = \underset{T\times l_T}{\mathbf{S}} \cdot \underset{l_T\times 1}{\boldsymbol{\delta}} + \underset{T\times 1}{\boldsymbol{\varepsilon}}, \tag{B.94}$$

where $\mathbf{S} = (s_{it}) = (\boldsymbol{s}_1, \boldsymbol{s}_2 ..., \boldsymbol{s}_{l_T})$*, with* $\boldsymbol{s}_j = (s_{j1}, s_{j2}, ..., s_{jT})'$*, for* $j = 1, 2, ..., l_T$*, Denote the least squares estimator of* $\boldsymbol{\delta}$ *in the regression model (B.94), by* $\hat{\boldsymbol{\delta}}$ *, and the associated* $T \times 1$ *vector of least squares residuals, by* $\tilde{\mathbf{u}} = \boldsymbol{y} - \mathbf{S}\hat{\boldsymbol{\delta}}$*, and set* $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}', \mathbf{0}'_{l_T-k-1}\right)'$*. Denote the eigenvalues of* $\boldsymbol{\Sigma}_{ss} = E\left(T^{-1}\mathbf{S}'\mathbf{S}\right)$ *by* $\mu_1 \leq \mu_2 \leq ... \leq \mu_{l_T}$*, and assume that the following conditions hold:*

i. $\mu_i = O(l_T)$*,* $i = l_T - M + 1, l_T - M + 2, ..., l_T$*, for some finite* M*,* $\sup_{1\leq i\leq l_T-M} \mu_i < C_0 < \infty$*, for some* $C_0 > 0$*, and* $\inf_{1\leq i<l_T} \mu_i > C_1 > 0$*, for some* $C_1 > 0$*.*

ii. Regressors are uncorrelated with the errors, $E(s_{jt}u_t) = 0 = E(x_{it}u_t)$*, for all* $t = 1, 2..., T$*,* $i = 1, 2, ..., k$*, and* $j = 1, 2, ..., l_T$*,* s_{it} *have finite* 8^{th} *order moments, and* $z_{ij,t} = s_{it}s_{jt} - E(s_{it}s_{jt})$ *satisfies conditions (B.76) and (B.78) of Lemma A18. Moreover,* $z^*_{ij,t} = s_{it}x_{jt} - E(s_{it}x_{jt})$ *satisfies condition (B.76) of Lemma A18.*

Suppose that $l_T^3/T \to 0$*, as* l_T *and* $T \to \infty$*, Then, if* $\mathbf{S}$ *contains* $\mathbf{X}$

$$F_{\tilde{u}} = T^{-1} \left\|\tilde{\mathbf{u}}\right\|^2 = \sigma^2 + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T^3}{T^{3/2}}\right) + O_p\left(\frac{l_T^{3/2}}{T}\right), \tag{B.95}$$

and

$$\left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right) + O_p\left(\frac{l_T^{5/2}}{T}\right). \tag{B.96}$$

But if one or more columns of $\mathbf{X}$ *are not contained in* $\mathbf{S}$*, then*

$$F_{\tilde{u}} = \sigma^2 + O_p(1), \tag{B.97}$$

and

$$\left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0\right\|_F = O(l_T) + O_p\left(\frac{l_T^{5/2}}{T}\right) + O_p\left(\frac{l_T^{5/2}}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{\sqrt{T}}\right). \tag{B.98}$$

Proof. Let $\hat{\boldsymbol{\Sigma}}_{ss} = \mathbf{S}'\mathbf{S}/T$, and recall that by assumption matrices $\boldsymbol{\Sigma}_{ss} = E\left(T^{-1}\mathbf{S}'\mathbf{S}\right)$ and $\hat{\boldsymbol{\Sigma}}_{ss}$ are positive definite. Let $\hat{\boldsymbol{\Delta}}_{ss} = \hat{\boldsymbol{\Sigma}}_{ss}^{-1} - \boldsymbol{\Sigma}_{ss}^{-1}$ and using (2.15) of Berk (1974), note that

$$\left\|\hat{\boldsymbol{\Delta}}_{ss}\right\|_F \leq \frac{\|\boldsymbol{\Sigma}_{ss}^{-1}\|_F^2 \left\|\hat{\boldsymbol{\Sigma}}_{ss} - \boldsymbol{\Sigma}_{ss}\right\|_F}{1 - \|\boldsymbol{\Sigma}_{ss}^{-1}\|_F \left\|\hat{\boldsymbol{\Sigma}}_{ss} - \boldsymbol{\Sigma}_{ss}\right\|_F}. \tag{B.99}$$

We focus on the individual terms on the right side of (B.99) to establish a bound, in probability, for $\left\|\hat{\boldsymbol{\Delta}}_{ss}\right\|_F$. The assumptions on eigenvalues of $\boldsymbol{\Sigma}_{ss}$ in this lemma are the same as in Lemma A14 with the only exception that $O(.)$ terms are used instead of $\ominus(.)$. Using the same arguments as in the proof of (B.64) and (B.65) of Lemma A14, it follows that

$$\|\mathbf{\Sigma}_{ss}\|_F = O(l_T), \tag{B.100}$$

and

$$\|\mathbf{\Sigma}_{ss}^{-1}\|_F = O\left(\sqrt{l_T}\right). \tag{B.101}$$

Moreover, note that (i,j)-th element of $\left(\hat{\mathbf{\Sigma}}_{ss} - \mathbf{\Sigma}_{ss}\right)$, $z_{ijt} = s_{it}s_{jt} - E(s_{it}s_{jt})$, satisfies the conditions of Lemma A18, which establishes

$$E\left(\left\|\hat{\mathbf{\Sigma}}_{ss} - \mathbf{\Sigma}_{ss}\right\|_F^2\right) = O\left(\frac{l_T^2}{T}\right), \tag{B.102}$$

and therefore, using $E\left\|\hat{\mathbf{\Sigma}}_{ss} - \mathbf{\Sigma}_{ss}\right\|_F \leq \left[E\left(\left\|\hat{\mathbf{\Sigma}}_{ss} - \mathbf{\Sigma}_{ss}\right\|_F^2\right)\right]^{1/2}$, and the fact that L_1-convergence implies convergence in probability, we have.

$$\left\|\hat{\mathbf{\Sigma}}_{ss} - \mathbf{\Sigma}_{ss}\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right). \tag{B.103}$$

Using (B.101) and (B.103), it now follows that

$$\|\mathbf{\Sigma}_{ss}^{-1}\|_F \left\|\hat{\mathbf{\Sigma}}_{ss} - \mathbf{\Sigma}_{ss}\right\|_F = O_p\left(\frac{l_T^{3/2}}{\sqrt{T}}\right),$$

and since by assumption $\frac{l_T^{3/2}}{\sqrt{T}} \to 0$, then

$$\frac{1}{\left(1 - \|\mathbf{\Sigma}_{ss}^{-1}\|_F \left\|\hat{\mathbf{\Sigma}}_{ss} - \mathbf{\Sigma}_{ss}\right\|_F\right)^2} = O_p(1). \tag{B.104}$$

Now using (B.103), (B.104), and (B.101) in (B.99), we have

$$\left\|\hat{\mathbf{\Delta}}_{ss}\right\|_F = O(l_T)\, O_p\left(\frac{l_T}{\sqrt{T}}\right) O_p(1) = O_p\left(\frac{l_T^2}{\sqrt{T}}\right), \tag{B.105}$$

and hence

$$\left\|\left(\frac{\mathbf{S}'\mathbf{S}}{T}\right)^{-1}\right\|_F = \left\|\hat{\mathbf{\Sigma}}_{ss}^{-1}\right\|_F \leq \left\|\hat{\mathbf{\Delta}}_{ss}\right\|_F + \|\mathbf{\Sigma}_{ss}^{-1}\|_F = O_p\left(\frac{l_T^2}{\sqrt{T}}\right) + O_p\left(\sqrt{l_T}\right). \tag{B.106}$$

Further, since by the assumption $E(\boldsymbol{s}_t u_t) = 0$, then $\left\|\frac{\mathbf{S}'\boldsymbol{u}}{T}\right\|_F^2 = O_p\left(\frac{l_T}{T}\right)$, and

$$\left\|\frac{\mathbf{S}'\boldsymbol{u}}{T}\right\|_F = O_p\left(\sqrt{\frac{l_T}{T}}\right). \tag{B.107}$$

Consider now the $T \times 1$ vector of residuals, $\tilde{\mathbf{u}}$ from the regression model (B.94) and note that under (B.93) it can be written as

$$\tilde{\mathbf{u}} = \mathbf{M}_s \boldsymbol{y} = \mathbf{M}_s \boldsymbol{u} + \mathbf{M}_s \mathbf{X}\beta, \text{ where } \mathbf{M}_s = \mathbf{I}_T - \mathbf{S}\left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'. \tag{B.108}$$

In the case where $\mathbf{X}$ is a sub-set of $\mathbf{S}$, $\mathbf{M}_s\mathbf{X}\beta = \mathbf{0}$, and

$$F_{\tilde{u}} = T^{-1}\left\|\tilde{\mathbf{u}}\right\|^2 = T^{-1}\boldsymbol{u}'\mathbf{M}_s\boldsymbol{u} = T^{-1}\boldsymbol{u}'\boldsymbol{u} - \left(T^{-1}\boldsymbol{u}'\mathbf{S}\right)\left(T^{-1}\mathbf{S}'\mathbf{S}\right)^{-1}\left(T^{-1}\mathbf{S}'\boldsymbol{u}\right). \tag{B.109}$$

Also since u_t are serially uncorrelated with zero means and variance σ^2, we have

$$T^{-1}\boldsymbol{u}'\boldsymbol{u} = \sigma^2 + O_p\left(T^{-1/2}\right),$$

and

$$\left\|\left(T^{-1}\boldsymbol{u}'\mathbf{S}\right)\left(T^{-1}\mathbf{S}'\mathbf{S}\right)^{-1}\left(T^{-1}\mathbf{S}'\boldsymbol{u}\right)\right\|_F \leq \left\|\frac{\mathbf{S}'\boldsymbol{u}}{T}\right\|_F^2 \left\|\left(\frac{\mathbf{S}'\mathbf{S}}{T}\right)^{-1}\right\|_F,$$

which in view of (B.106) and (B.107) yields

$$\left(T^{-1}\boldsymbol{u}'\mathbf{S}\right)\left(T^{-1}\mathbf{S}'\mathbf{S}\right)^{-1}\left(T^{-1}\mathbf{S}'\boldsymbol{u}\right) = O_p\left(\frac{l_T^3}{T^{3/2}}\right) + O_p\left(\frac{l_T^{3/2}}{T}\right).$$

The result (B.95) now follows using the above results in (B.109). Now consider the case where $\mathbf{S}$ does not contain $\mathbf{X}$, and note from (B.108) that

$$F_{\tilde{u}} = T^{-1}\boldsymbol{u}'\mathbf{M}_s\boldsymbol{u} + T^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_s\mathbf{X}\boldsymbol{\beta} + 2T^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_s\boldsymbol{u}. \tag{B.110}$$

Since $\mathbf{M}_s$ is an idempotent matrix then

$$\left\|T^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_s\mathbf{X}\boldsymbol{\beta}\right\|_F \leq \boldsymbol{\beta}'\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)\boldsymbol{\beta} = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} + O_p\left(T^{-1/2}\right) = O_p(1).$$

Similarly,

$$\begin{aligned} T^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_s\boldsymbol{u} &= T^{-1}\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{u} - \left(T^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{S}\right)\left(T^{-1}\mathbf{S}'\mathbf{S}\right)^{-1}\left(T^{-1}\mathbf{S}'\boldsymbol{u}\right) \\ &= O_p\left(T^{-1/2}\right) + O_p\left(\frac{l_T}{\sqrt{T}}\right) + O_p\left(\frac{l_T^{5/2}}{T}\right). \end{aligned}$$

The result (B.97) now follows if we use the above results in (B.110) and recalling that the probability order of $T^{-1}\boldsymbol{u}'\mathbf{M}_s\boldsymbol{u}$ is given by (B.95). Consider now the least squares estimator of $\hat{\boldsymbol{\delta}}$ and note that under (B.93) it can be written as

$$\hat{\boldsymbol{\delta}} = \left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{y} = \left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{u}. \tag{B.111}$$

Suppose that $\mathbf{X}$ is included as the first $k+1$ columns of $\mathbf{S}$, and denote the remaining $l_T - k - 1$ columns of $\mathbf{S}$ by $\mathbf{W}$. Also partition $\hat{\boldsymbol{\delta}}$ accordingly as $\left(\hat{\boldsymbol{\delta}}_x', \hat{\boldsymbol{\delta}}_w'\right)'$, where $\hat{\boldsymbol{\delta}}_x$ is the $(k+1)\times 1$ vector of estimated coefficients associated with $\mathbf{X}$. Note also that in this case $\mathbf{S}\left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{X} = \mathbf{X}$, and we have

$$\mathbf{S}\hat{\boldsymbol{\delta}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{S}\left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{u},$$

or

$$\mathbf{X}\left(\hat{\boldsymbol{\delta}}_x - \boldsymbol{\beta}\right) + \mathbf{W}\left(\hat{\boldsymbol{\delta}}_w - \mathbf{0}_{l_T-k-1}\right) = \boldsymbol{S}\left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{u},$$

which can be written more compactly as $\mathbf{S}\left(\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0\right) = \boldsymbol{S}\left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{u}$, where $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}', \mathbf{0}'_{l_T-k-1}\right)'$. Premultiplying both sides by $\mathbf{S}'$, and noting that $\mathbf{S}'\mathbf{S}$ is invertible yields

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0 = \left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{u},$$

with the norm of $\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0$ given by

$$\left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0\right\|_F = \left\|\left(\frac{\mathbf{S}'\mathbf{S}}{T}\right)^{-1}\left(\frac{\mathbf{S}'\boldsymbol{u}}{T}\right)\right\|_F \leq \left\|\left(\frac{\mathbf{S}'\mathbf{S}}{T}\right)^{-1}\right\|_F \left\|\left(\frac{\mathbf{S}'\boldsymbol{u}}{T}\right)\right\|_F.$$

Now using (B.106) and (B.107) it readily follows that

$$\left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right) + O_p\left(\frac{l_T^{5/2}}{T}\right), \tag{B.112}$$

as required. Finally, in the case where one or more columns of $\mathbf{X}$ are not included in $\mathbf{S}$, consider the decomposition

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0 = \left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_*\right) + \left(\boldsymbol{\delta}_* - \boldsymbol{\beta}_0\right), \tag{B.113}$$

where $\boldsymbol{\delta}_* = \boldsymbol{\Sigma}_{ss}^{-1}\boldsymbol{\Sigma}_{sx}\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}_{sx} = E\left(T^{-1}\mathbf{S}'\mathbf{X}\right)$. When at least one of the columns of $\mathbf{X}$ does not belong to $\mathbf{S}$, then $\boldsymbol{\delta}_* \neq \boldsymbol{\beta}_0$. To investigate the probability order of the first term of the above, using (B.111), we note that

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_* = \left(\hat{\boldsymbol{\Sigma}}_{ss}^{-1}\hat{\boldsymbol{\Sigma}}_{sx} - \boldsymbol{\Sigma}_{ss}^{-1}\boldsymbol{\Sigma}_{sx}\right)\boldsymbol{\beta} + \left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{u},$$

where $\hat{\boldsymbol{\Sigma}}_{sx} = T^{-1}\mathbf{S}'\mathbf{X}$. But $\hat{\boldsymbol{\Sigma}}_{ss}^{-1}\hat{\boldsymbol{\Sigma}}_{sx} - \boldsymbol{\Sigma}_{ss}^{-1}\boldsymbol{\Sigma}_{sx} = \hat{\boldsymbol{\Delta}}_{ss}\hat{\boldsymbol{\Delta}}_{sx} + \hat{\boldsymbol{\Delta}}_{ss}\boldsymbol{\Sigma}_{sx} + \boldsymbol{\Sigma}_{ss}^{-1}\hat{\boldsymbol{\Delta}}_{sx}$, where $\hat{\boldsymbol{\Delta}}_{sx} = \hat{\boldsymbol{\Sigma}}_{sx} - \boldsymbol{\Sigma}_{sx}$, and, as before, $\hat{\boldsymbol{\Delta}}_{ss} = \hat{\boldsymbol{\Sigma}}_{ss}^{-1} - \boldsymbol{\Sigma}_{ss}^{-1}$. Hence

$$\begin{aligned}\left\|\left(\hat{\boldsymbol{\Sigma}}_{ss}^{-1}\hat{\boldsymbol{\Sigma}}_{sx} - \boldsymbol{\Sigma}_{ss}^{-1}\boldsymbol{\Sigma}_{sx}\right)\boldsymbol{\beta}\right\|_F \leq{}& \left\|\hat{\boldsymbol{\Delta}}_{ss}\right\|_F \left\|\hat{\boldsymbol{\Delta}}_{sx}\right\|_F \left\|\boldsymbol{\beta}\right\| + \left\|\hat{\boldsymbol{\Delta}}_{ss}\right\|_F \left\|\boldsymbol{\Sigma}_{sx}\right\|_F \left\|\boldsymbol{\beta}\right\| \\ &+ \left\|\boldsymbol{\Sigma}_{ss}^{-1}\right\|_F \left\|\hat{\boldsymbol{\Delta}}_{sx}\right\|_F \left\|\boldsymbol{\beta}\right\|\end{aligned}$$

Using Lemma A18 by setting $\mathbf{S}_a = \mathbf{S}$ ($l_{a,T} = l_T$) and $\mathbf{S}_b = \mathbf{X}$ ($l_{b,T} = k+1$), we also have, by (B.77),

$$\left\|\hat{\boldsymbol{\Delta}}_{sx}\right\|_F = \left\|\hat{\boldsymbol{\Sigma}}_{sx} - \boldsymbol{\Sigma}_{sx}\right\|_F = O_p\left(\sqrt{\frac{l_T}{T}}\right). \tag{B.114}$$

Also $\left\|\hat{\mathbf{\Delta}}_{ss}\right\|_F = O_p\left(l_T^2/\sqrt{T}\right)$ by (B.105), $\|\mathbf{\Sigma}_{ss}^{-1}\|_F = O\left(\sqrt{l_T}\right)$, by (B.101), $\|\mathbf{\Sigma}_{sx}\|_F = O\left(\sqrt{l_T}\right)$, $\|\boldsymbol{\beta}\| = O(1)$. Therefore

$$\left\|\left(\hat{\mathbf{\Sigma}}_{ss}^{-1}\hat{\mathbf{\Sigma}}_{sx} - \mathbf{\Sigma}_{ss}^{-1}\mathbf{\Sigma}_{sx}\right)\boldsymbol{\beta}\right\|_F = O_p\left(l_T^2/\sqrt{T}\right) O_p\left(\sqrt{\frac{l_T}{T}}\right) + O_p\left(l_T^2/\sqrt{T}\right) O\left(\sqrt{l_T}\right) + O\left(\sqrt{l_T}\right) O_p\left(\sqrt{\frac{l_T}{T}}\right)$$

$$= O_p\left(\frac{l_T^{5/2}}{T}\right) + O_p\left(\frac{l_T^{5/2}}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{\sqrt{T}}\right).$$

Therefore, also using (B.112), overall we have

$$\left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_*\right\|_F = O_p\left(\frac{l_T^{5/2}}{T}\right) + O_p\left(\frac{l_T^{5/2}}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{\sqrt{T}}\right).$$

Finally, using (B.113)

$$\left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0\right\|_F \leq \left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_*\right\|_F + \|\boldsymbol{\delta}_*\|_F + \|\boldsymbol{\beta}_0\|_F,$$

where $\|\boldsymbol{\beta}_0\| = O(1)$, since $\boldsymbol{\beta}_0$ contains finite $(k+1)$ number of bounded nonzero elements, and

$$\begin{aligned}\|\boldsymbol{\delta}_*\|_F &= \left\|\mathbf{\Sigma}_{ss}^{-1}\mathbf{\Sigma}_{sx}\right\|_F \\ &\leq \left\|\mathbf{\Sigma}_{ss}^{-1}\right\|_F \|\mathbf{\Sigma}_{sx}\|_F.\end{aligned}$$

$\|\mathbf{\Sigma}_{ss}^{-1}\|_F = O\left(\sqrt{l_T}\right)$ by (B.101), and $\|\mathbf{\Sigma}_{sx}\|_F = O\left(\sqrt{l_T}\right)$. Hence, in the case where at least one of the columns of $\mathbf{X}$ does not belong to $\mathbf{S}$, we have

$$\left\|\hat{\boldsymbol{\delta}} - \boldsymbol{\beta}_0\right\|_F = O(l_T) + O_p\left(\frac{l_T^{5/2}}{T}\right) + O_p\left(\frac{l_T^{5/2}}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{\sqrt{T}}\right).$$

which completes the proof of (B.98). ■

B. Proof of Theorem 3

We proceed as in the proof of (B.52) in Lemma A10. We have that

$$\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{y}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{x}_i}{T}\right)}}\right| > c_p(n,\delta)\right] \leq \Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x_i,(T)}} + \frac{T^{1/2}\theta_i}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right).$$

We distinguish two cases: $\frac{T^{1/2}|\theta_i|}{\sigma_{e,(T)}\sigma_{x_i,(T)}} > \frac{c_p(n,\delta)}{1+d_T}$ and $\frac{T^{1/2}|\theta_i|}{\sigma_{e,(T)}\sigma_{x_i,(T)}} \leq \frac{c_p(n,\delta)}{1+d_T}$. If $\frac{T^{1/2}|\theta_i|}{\sigma_{e,(T)}\sigma_{x_i,(T)}} > \frac{c_p(n,\delta)}{1+d_T}$,

$$\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x_i,(T)}} + \frac{T^{1/2}\theta_i}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) =$$

$$1 - \Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x_i,(T)}} + \frac{T^{1/2}\theta_i}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right| \leq \frac{c_p(n,\delta)}{1+d_T}\right),$$

and, by Lemma A12

$$\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T}-\theta\right)}{\sigma_{e,(T)}\sigma_{x_i,(T)}}+\frac{T^{1/2}\theta_i}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right|\leq\frac{c_p\left(n,\delta\right)}{1+d_T}\right)$$
$$\leq\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T}-\theta\right)}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right|>\frac{T^{1/2}\left|\theta_i\right|}{\sigma_{e,(T)}\sigma_{x_i,(T)}}-\frac{c_p\left(n,\delta\right)}{1+d_T}\right)$$

while, if $\frac{T^{1/2}|\theta_i|}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\leq\frac{c_p(n,\delta)}{1+d_T}$, by (B.150) of Lemma F4,

$$\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T}-\theta\right)}{\sigma_{e,(T)}\sigma_{x_i,(T)}}+\frac{T^{1/2}\theta_i}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right|>\frac{c_p\left(n,\delta\right)}{1+d_T}\right)$$
$$\leq\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T}-\theta\right)}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right|>\frac{c_p\left(n,\delta\right)}{1+d_T}-\frac{T^{1/2}\left|\theta_i\right|}{\sigma_{e,(T)}\sigma_{x_i,(T)}}\right)$$

We further note that since $c_p\left(n,\delta\right)\to\infty$, $\frac{T^{1/2}|\theta_i|}{\sigma_{e,(T)}\sigma_{x_i,(T)}}>\frac{c_p(n,\delta)}{1+d_T}$ implies $T^{1/2}\left|\theta_i\right|>C_2$, for some $C_2>0$. Then, noting that $\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{\eta}}{T}-\theta$ is the average of a martingale difference process, by Lemma A6, for some positive constants, C_1,C_2,C_3,C_4,C_5, and, for any $\psi>0$, we have

$$\sum_{i=k+1}^{n}\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(\boldsymbol{e}'\boldsymbol{e}/T\right)\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{x}_i}{T}\right)}}\right|>c_p\left(n,\delta\right)\right]\leq C_1\sum_{i=k+1}^{n}I\left(\sqrt{T}\theta_i>C_2\right)$$
$$+C_3\sum_{i=k+1}^{n}I\left(\sqrt{T}\theta_i\leq C_4\right)\exp\left[-\ln(n)^{C_5}\right],$$
$$=C_1\sum_{i=k+1}^{n}I\left(\sqrt{T}\theta_i>C_2\right)+o(n^{1-\psi})+O\left[\exp(-CT^{C_5})\right],\tag{B.115}$$

since $\exp\left[-\ln(n)^{C_5}\right]=o(n^{\psi})$, which follows by noting that $C_0\ln(n)^{1/2}=o\left(C_1\ln(n)\right)$, for any $C_0,C_1>0$. As a result, the crucial term for the behaviour of $FPR_{n,T}$ is the first term on the RHS of (B.115). Consider now the above probability bound under the two specifications assumed for θ_i as given by (4) and (5). Under (4), for any $\psi>0$,

$$\sum_{i=k+1}^{n}\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(\boldsymbol{e}'\boldsymbol{e}/T\right)\left(\frac{\boldsymbol{x}_i'\mathbf{M}_q\boldsymbol{x}_i}{T}\right)}}\right|>c_p\left(n,\delta\right)\right]\leq C_0\sum_{i=k+1}^{n}I\left(\sqrt{T}\varrho^i>C_i\right)+o(n^{1-\psi}).$$

for some $C_0,C_i>0$, $i=k+1,...,n$. So we need to determine the limiting property of $\sum_{i=k+1}^{n}I\left(\sqrt{T}\varrho^i>C_i\right)$. Then, without loss of generality, consider $i=[n^{\zeta}]$, $T=n^{\kappa_1}$, $\zeta\in[0,1]$,

$\kappa_1 > 0$. Then, $\sqrt{T}\varrho^i = \sqrt{T}\varrho^{T^{(1/\kappa_1)\zeta}} = o(1)$ for all $\kappa_1, \zeta > 0$. Therefore,

$$C_a \sum_{i=k+1}^{n} I\left(\sqrt{T}\varrho^i > C_b/C_i\right) = o(n^{\zeta}),$$

for all $\zeta > 0$. This implies that under (4), $\theta_i = C_i\varrho^i$, $|\varrho| < 1$, and $c_p(n,\delta) = O\left[\ln(n)^{1/2}\right]$, we have

$$E\left|FPR_{n,T}\right| = o(n^{\zeta-1}) + O\left[\exp(-n^{C_0})\right],$$

for all $\zeta > 0$. Similarly, under (5), $\theta_i = C_i i^{-\gamma}$, and setting $i = [n^{\zeta}]$, $T = n^{\kappa_1}$, $\zeta, \kappa_1 > 0$, we have $\sqrt{T}\theta_i = T^{-(1/\kappa_1)\zeta\gamma+1/2}$. We need $-(1/\kappa_1)\zeta\gamma + 1/2 < 0$ or $\zeta > \frac{1}{2\kappa_1^{-1}\gamma}$. Then,

$$\frac{C_a}{n} \sum_{i=k+1}^{n} I\left(\sqrt{T}\theta_i > C_b/C_i\right) = O\left(T^{\frac{1}{2\kappa_1^{-1}\gamma} - \kappa_1^{-1}}\right) = O\left(n^{\frac{1}{2\kappa_1^{-2}\gamma} - 1}\right)$$

So

$$E\left|FPR_{n,T}\right| = o(1), \tag{B.116}$$

as long as $2\kappa_1^{-2}\gamma > 1$ or if $\gamma > \frac{1}{2\kappa_1^{-2}}$.

Remark B1 *Note that if $\kappa_1 = 1$, then the condition for (B.116) requires that $\gamma > \frac{1}{2}$.*

C. Some results for the case where either noise variables are mixing, or both signal/pseudo-signal and noise variables are mixing

When only noise variables are mixing, all the results of the main paper go through since we can use the results obtained under (D1)-(D3) of Lemma D2 to replace Lemma A6.

As discussed in Section 4.2, some weak results can be obtained if both signal/pseudo-signal and noise variables are mixing processes, but only if $c_p(n)$ is allowed to grow faster than under the assumption of a martingale difference. This case is covered under (D4) of Lemma D2 and (B.140)-(B.141) of Lemma D3. There, it is shown that, for sufficiently large constants $C_0 - C_3$ for Assumption 4, the martingale difference bound which is given by $\exp\left[-\frac{1}{2}\varkappa c_p^2(n)\right]$ in Lemma A6 is replaced by the bound $\exp\left[-C_4 c_p(n)^{s/(s+2)}\right]$, for some $C_4 > 0$, where s is the exponent in the probability tail in Assumption 4. It is important to note here that this bound seems to be relatively sharp (see, e.g., Roussas (1996)), under our assumptions, and so we need to understand its implications for our analysis. We abstract from the constant C_4 which can further deteriorate rates. Given (see result (i) of Lemma A2),

$$c_p(n) = O\left\{\left[\ln\left(\frac{f(n)}{2p}\right)\right]^{1/2}\right\},$$

it follows that

$$\exp\left[-c_p(n)^{s/(s+2)}\right] = O\left[\exp\left\{-\left[\ln\left(\frac{f(n)}{2p}\right)\right]^{s/2(s+2)}\right\}\right]$$

Let $f(n) = 2p\exp(n^{a_n})$. Then,

$$\exp\left\{-\left[\ln\left(\frac{f(n)}{2p}\right)\right]^{s/2(s+2)}\right\} = \exp\left[-n^{a_n s/2(s+2)}\right]$$

To obtain the same bound as for the martingale difference case, we need to find a sequence $\{a_n\}$, such that $n^{Ca_n} = O(\ln(n))$. Setting $n^{Ca_n} = \ln(n)$, it follows that $a_n = \ln(\ln(n))/C\ln n$. Further, setting $C = s/2(s+2)$, we have $a_n = \frac{2(s+2)\ln(\ln(n))}{s\ln n}$, which leads to the following choice for $f(n)$

$$f(n) = 2p\exp\left(n^{\frac{2(s+2)\ln(\ln(n))}{s\ln n}}\right) \sim 2p\exp\left(\ln(n)^{\frac{2(s+2)}{s}}\right).$$

Then,

$$c_p(n) = O\left[\ln\left(\exp\left(\ln(n)^{\frac{2(s+2)}{s}}\right)\right)\right] = O\left(\ln(n)^{\frac{2(s+2)}{s}}\right),$$

which for $n = O\left(T^{C_1}\right)$, $C_1 > 0$, implies that $c_p(n) = O\left(\ln(T)^{\frac{2(s+2)}{s}}\right)$, and so, $c_p(n) = o\left(T^{C_2}\right)$, for all $C_2 > 0$, as long as $s > 0$.

We need to understand the implications of this result. For example, setting $s = 2$ which corresponds to the normal case gives $\exp(\ln(n)^4)$ which makes the calculation of $\Phi^{-1}\left(1 - \frac{p}{2f(n)}\right)$ numerically problematic for $n > 25$. The fast rate at which $f(n)$ grows basically implies that we need $s \to \infty$ which corresponds to $f(n) = 2p\exp(\ln(n)^2)$. Even then, the analysis becomes problematic for large n. $s \to \infty$ corresponds for all practical purposes to assuming boundedness for x_{it}. As a result, while the case of mixing x_{it} can be analysed theoretically, its practical implications are limited. On the other hand our Monte Carlo study in Section 5 suggests that setting $f(n) = f(n,\delta) = n^{\delta}$, $\delta \geq 1$ provides quite good results for autoregressive x_{it} in small samples.

D. Lemmas for mixing results

We consider the following assumptions that replace Assumption 3.

Assumption D1 *x_{it}, $i = 1,2,...,k+k^*$, are martingale difference processes with respect to $\mathcal{F}_{t-1}^{xs} \cup \mathcal{F}_{t}^{xn}$, where $\mathcal{F}_{t-1}^{xs}$ and $\mathcal{F}_{t}^{xn}$ are defined in Assumption 3. x_{it}, $i = 1,2,...,k+k^*$ are independent of x_{it}, $i = k+k^*+1, k+k^*+2,...,n$. $E\left(x_{it}x_{jt} - E\left(x_{it}x_{jt}\right)\left|\mathcal{F}_{t-1}^{xs}\right.\right) = 0$, $i,j = 1,2,...,k+k^*$. x_{it}, $i = k+k^*+1, k+k^*+2,...,n$, are heterogeneous strongly mixing processes with mixing coefficients given by $\alpha_{i\ell} = C_{i\ell}\xi^{\ell}$ for some $C_{i\ell}$ such that $\sup_{i,\ell} C_{i\ell} < \infty$ and some $0 < \xi < 1$. $E\left[x_{it}u_t\left|\mathcal{F}_{t-1}\right.\right] = 0$, for $i = 1,2,...,n$, and all t.*

Assumption D2 *x_{it}, $i = 1,2,...,k+k^*$ are independent of x_{it}, $i = k+k^*+1, k+k^*+2,...,n$. x_{it}, $i = 1,2,...,n$, are heterogeneous strongly mixing processes with mixing coefficients given by $\alpha_{i\ell} = C_{i\ell}\xi^{\ell}$ for some $C_{i\ell}$ such that $\sup_{i,\ell} C_{i\ell} < \infty$ and some $0 < \xi < 1$. $E\left[x_{it}u_t\left|\mathcal{F}_{t-1}\right.\right] = 0$, for $i = 1,2,...,n$, and all t.*

Lemma D1 *Let ξ_t be a sequence of zero mean, mixing random variables with exponential mixing coefficients given by $\phi_k = a_{0k}\varphi^k$, $0 < \varphi < 1$, $a_{0k} < \infty$, $k = 1, \ldots$. Assume, further, that* $\Pr(|\xi_t| > \alpha) \leq C_0 \exp[-C_1 \alpha^s]$, $s \geq 1$. *Then, for some $C_2, C_3 > 0$, each $0 < \delta < 1$ and $v_T \geq \epsilon T^{\lambda}$, $\lambda > (1+\delta)/2$,*

$$\Pr\left(\left|\sum_{t=1}^{T} \xi_t\right| > v_T\right) \leq C_2 \exp\left[-\left(C_3 v_T T^{-(1+\delta)/2}\right)^{s/(s+1)}\right]$$

Proof. We reconsider the proof of Theorem 3.5 of White and Wooldridge (1991) relaxing the assumption of stationarity. Define $w_t = \xi_t I(z_t \leq D_T)$ and $v_t = \xi_t - w_t$ where D_T will be defined below. Using Theorem 3.4 of White and Wooldridge (1991), which does not assume stationarity, we have that constants C_0 and C_1 in the statement of the present Lemma can be chosen sufficiently large such that

$$\Pr\left(\left|\sum_{t=1}^{T} w_t - E(w_t)\right| > v_T\right) \leq C_4 \exp\left[\frac{-C_5 v_T T^{-(1+\delta)/2}}{D_T}\right] \tag{B.117}$$

for some $C_4, C_5 > 0$, rather than

$$\Pr\left(\left|\sum_{t=1}^{T} w_t - E(w_t)\right| > v_T\right) \leq C_6 \exp\left[\frac{-C_7 v_T T^{-1/2}}{D_T}\right]$$

for some $C_6, C_7 > 0$, which uses Theorem 3.3 of White and Wooldridge (1991). We explore the effects this change has on the final rate. We revisit the analysis of the bottom half of page 489 of White and Wooldridge (1991). We need to determine D_T such that

$$v_T^{-1} T \left[\exp\left(-C_1 \left(\frac{D_T}{2}\right)^s\right)\right]^{1/q} \leq \exp\left[\frac{-C v_T T^{-(1+\delta)/2}}{D_T}\right]$$

for some $C > 0$. Take logs and we have

$$\ln\left(v_T^{-1} T\right) - \left(\frac{1}{q}\right) C_1 \left(\frac{D_T}{2}\right)^s \leq \frac{-C v_T T^{-(1+\delta)/2}}{D_T}$$

or

$$D_T^s \geq 2^p \left(\frac{q}{C_1}\right) \ln\left(v_T^{-1} T\right) + \frac{2^s q C v_T}{C_1 T^{(1+\delta)/2} D_T}$$

For this it suffices that

$$\frac{2^s q C v_T}{T^{(1+\delta)/2} D_T} \geq 2^p q \ln\left(v_T^{-1} T\right) \tag{B.118}$$

and

$$D_T^s \geq \frac{2^s q C v_T}{C_1 T^{(1+\delta)/2} D_T}. \tag{B.119}$$

Set

$$D_T = \left(\frac{2^s q C v_T}{C_1 T^{(1+\delta)/2}}\right)^{1/(s+1)},$$

so that (B.119) holds with equality. But since $v_T \geq \epsilon T^{\lambda}$, $\lambda > (1+\delta)/2$, (B.118) holds. Therefore,

$$\frac{2^s qCv_T}{C_1 T^{(1+\delta)/2} D_T} = \left(\frac{2^s qCv_T}{C_1 T^{(1+\delta)/2}} \right)^{s/(s+1)},$$

and the desired result follows. ■

Remark D1 *The above lemma shows how one can relax the boundedness assumption in Theorem 3.4 of White and Wooldridge (1991) to obtain an exponential inequality for heterogeneous mixing processes with exponentially declining tail probabilities. Note that neither Theorem 3.4 of White and Wooldridge (1991) which deals with heterogeneity nor Theorem 3.5 of White and Wooldridge (1991) which deals with stationary mixing processes is sufficient for handling the heterogeneous mixing processes we consider.*

Remark D2 *It is important for the rest of the lemmas in this supplement, and in particular, the results obtained under (D4) of Lemma D2, to also note that Lemma 2 of Dendramis et al. (2015) provides the result of Lemma D1 when $\delta = 0$.*

Lemma D2 *Let x_t, $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$, and u_t be sequences of random variables and suppose that there exist finite positive constants C_0 and C_1, and $s > 0$ such that $\sup_t \Pr(|x_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, $\sup_{i,t} \Pr(|q_{i,t}| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, and $\sup_t \Pr(|u_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^s)$, for all $\alpha > 0$. Let $\boldsymbol{\Sigma}_{qq} = \frac{1}{T}\sum_{t=1}^{T} E(\boldsymbol{q}_{\cdot t}\boldsymbol{q}'_{\cdot t})$ be a nonsingular matrix such that $0 < \left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F$. Suppose that Assumption 5 holds for x_t and $\boldsymbol{q}_{\cdot t}$, and denote the corresponding projection residuals defined by (11) as $u_{x,t} = x_t - \boldsymbol{\gamma}'_{qx,T}\boldsymbol{q}_{\cdot t}$. Let $\hat{\boldsymbol{u}}_x = (\hat{u}_{x,1}, \hat{u}_{x,2}, ..., \hat{u}_{x,T})'$ denote the $T \times 1$ LS residual vector of the regression of x_t on $\boldsymbol{q}_{\cdot t}$. Let $\mathcal{F}_t = \mathcal{F}_t^x \cup \mathcal{F}_t^u$, $\mathcal{F}_t^q = \sigma\left(\{\boldsymbol{q}_{\cdot t}\}_{s=1}^{t}\right)$ and assume either (D1) $E\left(u_{x,t}u_t - \mu_{xu,t}|\mathcal{F}_{t-1} \cup \mathcal{F}_{t-1}^q\right) = 0$, where $\mu_{xu,t} = E(u_{x,t}u_t)$, x_t and u_t are martingale difference processes, $\boldsymbol{q}_{\cdot t}$ is an exponentially mixing process, and $\zeta_T = o(T^{\lambda})$, for all $\lambda > 1/2$, or (D2) $E\left(u_{x,t}u_t - \mu_{xu,t}|\mathcal{F}_{t-1} \cup \mathcal{F}_{t-1}^q\right) = 0$, where $\mu_{xu,t} = E(u_{x,t}u_t)$, u_t is a martingale difference processes, x_t and $\boldsymbol{q}_{\cdot t}$ are exponentially mixing processes, and $\zeta_T = o(T^{\lambda})$, for all $\lambda > 1/2$, or (D3) x_t, u_t and $\boldsymbol{q}_{\cdot t}$ are exponentially mixing processes, and $\zeta_T = o(T^{\lambda})$, for all $\lambda > 1$, or (D4) x_t, u_t and $\boldsymbol{q}_{\cdot t}$ are exponentially mixing processes, and $\zeta_T = o(T^{\lambda})$, for all $\lambda > 1/2$. Then, we have the following. If (D1) or (D2) hold, then, for any π in the range $0 < \pi < 1$, there exist finite positive constants C_0 and C_1, such that*

$$\Pr\left(\left| \sum_{t=1}^{T} x_t u_t - E(x_t u_t) \right| > \zeta_T \right) \leq \exp\left[\frac{-(1-\pi)^2 \zeta_T^2}{2T\omega_{xu,1,T}^2} \right] + \exp\left[-C_0 T^{C_1}\right] \tag{B.120}$$

and

$$\Pr\left(\left| \sum_{t=1}^{T} \hat{u}_{x,t} u_t - \mu_{xu,t} \right| > \zeta_T \right) \leq \exp\left[\frac{-(1-\pi)^2 \zeta_T^2}{2T\omega_{xu,T}^2} \right] + \exp\left[-C_0 T^{C_1}\right], \tag{B.121}$$

as long as $l_T = o(T^{1/3})$*, where* $\omega^2_{xu,1,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[(x_t u_t - E(x_t u_t))^2\right]$, $\omega^2_{xu,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[(u_{x,t} u_t - \mu_{xu,t})^2\right]$. *If (D3) holds*

$$\Pr\left(\left|\sum_{t=1}^{T} x_t u_t - E(x_t u_t)\right| > \zeta_T\right) \leq \exp\left[-C_0 T^{C_1}\right], \tag{B.122}$$

for some $C_0, C_1 > 0$*, and*

$$\Pr\left(\left|\sum_{t=1}^{T} \hat{u}_{x,t} u_t - \mu_{xu,t}\right| > \zeta_T\right) \leq \exp\left[-C_0 T^{C_1}\right], \tag{B.123}$$

for some $C_0, C_1 > 0$*, as long as* $l_T = o(T^{1/3})$*. Finally, if (D4) holds,*

$$\Pr\left(\left|\sum_{t=1}^{T} x_t u_t - E(x_t u_t)\right| > \zeta_T\right) \leq C_1 \exp\left[-C_0 \left(\zeta_T T^{-1/2}\right)^{s/(s+2)}\right], \tag{B.124}$$

for some $C_0, C_1 > 0$*, and*

$$\Pr\left(\left|\sum_{t=1}^{T} \hat{u}_{x,t} u_t - \mu_{xu,t}\right| > \zeta_T\right) \leq C_2 \exp\left[-C_3 \left(\zeta_T T^{-1/2}\right)^{s/(s+2)}\right] + \exp\left[-C_0 T^{C_1}\right], \tag{B.125}$$

for some $C_0, C_1, C_2, C_3 > 0$*, as long as* $l_T = o(T^{1/3})$*.*

Proof. We first prove the lemma under (D1) and then modify the derivations to establish that the results also hold under (D2)-(D4). The assumptions of the lemma state that there exists a regression model underlying $\hat{u}_{x,t}$ which is denoted by

$$x_t = \boldsymbol{\beta}_q' \boldsymbol{q}_{\cdot t} + u_{x,t}$$

for some $l \times 1$ vector, $\boldsymbol{\beta}_q$. Denoting $\boldsymbol{u}_x = (u_{x,1}, u_{x,2}, ..., u_{x,T})'$, $\boldsymbol{u} = (u_1, u_2, ..., u_T)'$, $\hat{\boldsymbol{\Sigma}}_{qq} = T^{-1}(\mathbf{Q}'\mathbf{Q})$, $\mathbf{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_l)$, and $\boldsymbol{q}_{i\cdot} = (q_{i1}, q_{i2}, ..., q_{iT})'$, we have

$$\begin{aligned}\hat{\boldsymbol{u}}_x'\boldsymbol{u} = \boldsymbol{u}_x'\boldsymbol{u} - \left(T^{-1}\boldsymbol{u}_x'\mathbf{Q}\right)\hat{\boldsymbol{\Sigma}}_{qq}^{-1}\left(\mathbf{Q}'\boldsymbol{u}\right) = \boldsymbol{u}_x'\,\boldsymbol{u} - \left(T^{-1}\boldsymbol{u}_x'\mathbf{Q}\right)\left(\hat{\boldsymbol{\Sigma}}_{qq}^{-1} - \boldsymbol{\Sigma}_{qq}^{-1}\right)\left(\mathbf{Q}'\boldsymbol{u}\right) + \\ \left(T^{-1}\boldsymbol{u}_x'\mathbf{Q}\right)\boldsymbol{\Sigma}_{qq}^{-1}\left(\mathbf{Q}'\boldsymbol{u}\right)\end{aligned}$$

Noting that, since u_t is a martingale difference process with respect to $\sigma\left(\{u_s\}_{s=1}^{t-1}, \{u_{x,s}\}_{s=1}^{t}, \{q_s\}_{s=1}^{t}\right)$, by Lemma A4,

$$\Pr\left(\left|\boldsymbol{u}_x'\boldsymbol{u}\right| > \zeta_T\right) \leq \exp\left[\frac{-(1-\pi)^2\zeta_T^2}{2T\omega^2_{xu,T}}\right]. \tag{B.126}$$

It therefore suffices to show that

$$\Pr\left(\left|\left(\frac{1}{T}\boldsymbol{u}_x'\mathbf{Q}\right)\left(\hat{\boldsymbol{\Sigma}}_{qq}^{-1} - \boldsymbol{\Sigma}_{qq}^{-1}\right)\left(\mathbf{Q}'\boldsymbol{u}\right)\right| > \zeta_T\right) \leq \exp\left[-C_0 T^{C_1}\right] \tag{B.127}$$

and

$$\Pr\left(\left|\left(\frac{1}{T}\boldsymbol{u}_x'\mathbf{Q}\right)\boldsymbol{\Sigma}_{qq}^{-1}\left(\mathbf{Q}'\boldsymbol{u}\right)\right| > \zeta_T\right) \leq \exp\left[-C_0 T^{C_1}\right] \tag{B.128}$$

We explore (B.126) and (B.127). We start with (B.126). We have by Lemma A11 that, for some sequence δ_T,[6]

$$\Pr\left(\left|\left(\frac{1}{T}\boldsymbol{u}_x'\mathbf{Q}\right)\left(\hat{\boldsymbol{\Sigma}}_{qq}^{-1}-\boldsymbol{\Sigma}_{qq}^{-1}\right)(\mathbf{Q}'\boldsymbol{u})\right|>\zeta_T\right)\leq$$

$$\Pr\left(\left\|\frac{1}{T}\boldsymbol{u}_x'\mathbf{Q}\right\|\left\|\left(\hat{\boldsymbol{\Sigma}}_{qq}^{-1}-\boldsymbol{\Sigma}_{qq}^{-1}\right)\right\|\left\|\mathbf{Q}'\boldsymbol{u}\right\|_F>\zeta_T\right)\leq\Pr\left(\left\|\left(\hat{\boldsymbol{\Sigma}}_{qq}^{-1}-\boldsymbol{\Sigma}_{qq}^{-1}\right)\right\|>\frac{\zeta_T}{\delta_T}\right)+$$

$$\Pr\left(\left\|\boldsymbol{u}_x'\mathbf{Q}\right\|_F\left\|\mathbf{Q}'\boldsymbol{u}\right\|_F>\delta_T T\right) \tag{B.130}$$

We consider the first term of the RHS of (B.130). Note that for all $1\leq i,j\leq l$.

$$\Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}\left[q_{it}q_{jt}-E(q_{it}q_{jt})\right]\right|>\zeta_T\right)\leq\exp(-C_0\left(T^{1/2}\zeta_T\right)^{s/(s+2)}), \tag{B.131}$$

since $q_{it}q_{jt}-E(q_{it}q_{jt})$ is a mixing process and $\sup_i\Pr\left(|q_{i,t}|>\alpha\right)\leq C_0\exp\left(-C_1\alpha^s\right)$, $s>0$. Then, by Lemma F3,

$$\Pr\left(\left\|\left(\hat{\boldsymbol{\Sigma}}_{qq}^{-1}-\boldsymbol{\Sigma}_{qq}^{-1}\right)\right\|>\frac{\zeta_T}{\delta_T}\right)\leq l_T^2\exp\left(\frac{-C_0T^{s/2(s+2)}\zeta_T^{s/(s+2)}}{\delta_T^{s/(s+2)}l_T^{s/(s+2)}\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{s/(s+1)}\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F+\frac{\zeta_T}{\delta_T}\right)^{s/(s+1)}}\right)+$$

$$l_T^2\exp\left(-C_0\frac{T^{s/2(s+2)}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F^{s/(s+2)}l_T^{s/(s+2)}}\right)=$$

$$l_T^2\exp\left(-C_0\left(\frac{T^{1/2}\zeta_T}{\delta_T l_T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F+\frac{\zeta_T}{\delta_T}\right)}\right)^{s/(s+2)}\right)+$$

$$l_T^2\exp\left(-C_0\left(\frac{T^{1/2}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T}\right)^{s/(s+2)}\right).$$

We now consider the second term of the RHS of (B.130). By (B.12), we have

$$\Pr\left(\left\|\boldsymbol{u}_x'\mathbf{Q}\right\|_F\left\|\mathbf{Q}'\boldsymbol{u}\right\|_F>\delta_T T\right)\leq\Pr\left(\left\|\boldsymbol{u}_x'\mathbf{Q}\right\|_F>\delta_T^{1/2}T^{1/2}\right)+\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}\right\|_F>\delta_T^{1/2}T^{1/2}\right).$$

[6] In what follows we use

$$\Pr\left(|AB|>c\right)\leq\Pr\left(|A|\,|B|>c\right) \tag{B.129}$$

where A and B are random variables. To see this note that $|AB|\leq|A|\,|B|$. Further note that for any random variables $A_1>0$ and $A_2>0$ for which $A_2>A_1$ the occurrence of the event $\{A_1>c\}$, for any constant $c>0$, implies the occurrence of the event $\{A_2>c\}$. Therefore, $\Pr\left(A_2>c\right)\geq\Pr\left(A_1>c\right)$ proving the result.

Note that $\|\mathbf{Q}'\boldsymbol{u}\|_F^2 = \sum_{j=1}^{l_T} \left(\sum_{t=1}^{T} q_{jt} u_t \right)^2$, and

$$\Pr\left(\|\mathbf{Q}'\boldsymbol{u}\|_F > (\delta_T T)^{1/2} \right) = \Pr\left(\|\mathbf{Q}'\boldsymbol{u}\|_F^2 > \delta_T T \right)$$
$$\leq \sum_{j=1}^{l_T} \Pr\left[\left(\sum_{t=1}^{T} q_{jt} u_t \right)^2 > \frac{\delta_T T}{l_T} \right]$$
$$= \sum_{j=1}^{l_T} \Pr\left[\left| \sum_{t=1}^{T} q_{jt} u_t \right| > \left(\frac{\delta_T T}{l_T} \right)^{1/2} \right].$$

Noting further that $q_{it} u_t$ and $q_{it} u_{xt}$ are martingale difference processes satisfying a result of the usual form we obtain

$$\Pr\left(\|\boldsymbol{u}_x' \mathbf{Q}\|_F > \delta_T^{1/2} T^{1/2} \right) \leq l_T \Pr\left(|\boldsymbol{u}_x' \mathbf{q}_i| > \frac{\delta_T^{1/2} T^{1/2}}{l_T^{1/2}} \right) \leq l_T \exp\left(\frac{-C\delta_T}{l_T} \right),$$

or

$$\Pr\left(\|\boldsymbol{u}_x' \mathbf{Q}\|_F > \delta_T^{1/2} T^{1/2} \right) \leq l_T \Pr\left(|\boldsymbol{u}_x' \mathbf{q}_i| > \frac{\delta_T^{1/2} T^{1/2}}{l_T^{1/2}} \right) \leq l_T \exp\left(\left(\frac{-\delta_T T}{l_T} \right)^{s/2(s+2)} \right),$$

depending on the order of magnitude of $\frac{\delta_T^{1/2} T^{1/2}}{l_T^{1/2}}$, and a similar result for $\Pr\left(\|\mathbf{Q}'\boldsymbol{u}\|_F > \delta_T^{1/2} T^{1/2} \right)$. Therefore,

$$\Pr\left(\|\boldsymbol{u}_x' \mathbf{Q}\|_F \|\mathbf{Q}'\boldsymbol{u}\|_F > \delta_T T \right) \leq \exp\left[-C_0 T^{C_1} \right]. \tag{B.132}$$

We wish to derive conditions for l_T under which $\frac{T^{1/2}\zeta_T}{\delta_T l_T \left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F \left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \frac{\zeta_T}{\delta_T} \right)}$, $\frac{T^{1/2}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T}$, and $\frac{\delta_T}{l_T}$ are of larger, polynomial in T, order than $\frac{\zeta_T^2}{T}$. Then, the factors in l_T in (B.26) and (B.132) are negligible. We let $\zeta_T = T^\lambda$, $l_T = T^d$, $\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F = l_T^{1/2} = T^{d/2}$ and $\delta_T = T^\alpha$, where $\alpha \geq 0$, can be chosen freely. This is a complex analysis and we simplify it by considering relevant values for our setting and, in particular, $\lambda \geq 1/2$, $\lambda < 1/2 + c$, for all $c > 1/2$, and $d < 1$. We have

$$\frac{T^{1/2}\zeta_T}{\delta_T l_T \left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F \left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F + \frac{\zeta_T}{\delta_T} \right)} = O\left(T^{1/2+\lambda-\alpha-2d} \right) + O\left(T^{1/2-3d/2} \right), \tag{B.133}$$

$$\frac{T^{1/2}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T} = O\left(T^{1/2-3d/2} \right), \tag{B.134}$$

$$\frac{\delta_T}{l_T} = O\left(T^{\alpha-d} \right), \tag{B.135}$$

and

$$\frac{\zeta_T^2}{T} = O\left(T^{2\lambda-1} \right) = O\left(c\ln T \right). \tag{B.136}$$

Clearly $d < 1/3$. Setting $\alpha = 1/3$, ensures all conditions are satisfied. Since $\boldsymbol{\Sigma}_{qq}^{-1}$ is of lower norm order than $\hat{\boldsymbol{\Sigma}}_{qq}^{-1} - \boldsymbol{\Sigma}_{qq}^{-1}$, (B.128) follows similarly proving the result under (D1). For (D2) and (D3) we proceed as follows. Under (D3), noting that u_t is a mixing process, then by Lemma D1, we have that (B.126) is replaced by

$$\Pr\left(\left|\boldsymbol{u}_x'\boldsymbol{u}\right| > \zeta_T\right) \leq \exp\left[-C_0\left(T^{-(1+\vartheta)/2}\zeta_T\right)^{s/(s+2)}\right], \tag{B.137}$$

else, under (D2), we have again that (B.126) holds. Further, by a similar analysis to that above, it is easily seen that, under (D2),

$$\Pr\left(\left\|\boldsymbol{u}_x'\mathbf{Q}\right\|_F\left\|\mathbf{Q}'\boldsymbol{u}\right\|_F > \delta_T T\right) \leq l_T\exp\left(\frac{-C\delta_T}{l_T}\right) + l_T\exp\left[-C_0\left(\frac{T^{-\vartheta/2}\delta_T^{1/2}}{l_T^{1/2}}\right)^{s/(s+2)}\right],$$

and under (D3),

$$\Pr\left(\left\|\boldsymbol{u}_x'\mathbf{Q}\right\|_F\left\|\mathbf{Q}'\boldsymbol{u}\right\|_F > \delta_T T\right) \leq 2l_T\exp\left[-C_0\left(\frac{T^{-\vartheta/2}\delta_T}{l_T}\right)^{s/2(s+2)}\right].$$

Under (D2), we wish to derive conditions for l_T under which $\frac{T^{1/2}\zeta_T}{\delta_T l_T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F+\frac{\zeta_T}{\delta_T}\right)}$, $\frac{T^{1/2}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T}$, and $\frac{\delta_T}{l_T}$ are of larger, polynomial in T, order than $\frac{\zeta_T^2}{T}$. But this is the same requirement to that under (D1). Under (D3), we wish to derive conditions for l_T under which $\frac{T^{1/2}\zeta_T}{\delta_T l_T\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F+\frac{\zeta_T}{\delta_T}\right)}$, $\frac{T^{1/2}}{\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F l_T}$, $\frac{\delta_T}{l_T}$ and $\left(T^{-1/2}\zeta_T\right)^{s/(s+2)}$ are of positive polynomial in T, order. But again the same conditions are needed as for (D1) and (D2). Finally, we consider (D4). But, noting Remark D2, the only difference to (D3) is that $\zeta_T \geq T^{1/2}$, rather than $\zeta_T \geq T$. Then, as long as $\left(T^{-1/2}\zeta_T\right)^{s/(s+2)} \to \infty$ the result follows. ■

Lemma D3 *Let y_t, for $t = 1, 2, ..., T$, be given by the data generating process (6) and suppose that u_t and $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$ satisfy Assumptions 2-4. Let $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ contain a constant and a subset of $\boldsymbol{x}_{nt}$, and let $\eta_t = \boldsymbol{x}_{b,t}'\boldsymbol{\beta}_b + u_t$, where $\boldsymbol{x}_{b,t}$ is $k_b \times 1$ dimensional vector of signal variables that do not belong to $\boldsymbol{q}_{\cdot t}$, with the associated coefficients, $\boldsymbol{\beta}_b$. Assume that $\boldsymbol{\Sigma}_{qq} = \frac{1}{T}\sum_{t=1}^{T} E\left(\boldsymbol{q}_{\cdot t}\boldsymbol{q}_{\cdot t}'\right)$ and $\hat{\boldsymbol{\Sigma}}_{qq} = \mathbf{Q}'\mathbf{Q}/T$ are both invertible, where $\mathbf{Q} = (\boldsymbol{q}_{1\cdot}, \boldsymbol{q}_{2\cdot}, ..., \boldsymbol{q}_{l_T\cdot})$ and $\boldsymbol{q}_{i\cdot} = (q_{i1}, q_{i2}, ..., q_{iT})'$, for $i = 1, 2, ..., l_T$. Moreover, let $l_T = o(T^{1/4})$ and suppose that Assumption 5 holds for x_t and $\boldsymbol{q}_{\cdot t}$, where x_t is a generic element of $\{x_{1t}, x_{2t}, ..., x_{nt}\}$ that does not belong to $\boldsymbol{q}_{\cdot t}$. Denote the corresponding projection residuals defined by (11) as $u_{x,t} = x_t - \boldsymbol{\gamma}_{qx,T}'\boldsymbol{q}_{\cdot t}$, and the projection residuals of y_t on $(\boldsymbol{q}_{\cdot t}', x_t)'$ as $e_t = y_t - \boldsymbol{\gamma}_{yqx,T}'(\boldsymbol{q}_{\cdot t}', x_t)'$. Define $\boldsymbol{x} = (x_1, x_2, ..., x_T)'$, $\boldsymbol{y} = (y_1, y_2, ..., y_T)'$, $\boldsymbol{e} = (e_1, e_2, ..., e_T)'$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, and $\theta = E\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\right)\boldsymbol{\beta}_b$, where $\mathbf{X}_b$ is $T \times k_b$ matrix of observations on $\boldsymbol{x}_{b,t}$. Finally, $c_p(n, \delta)$ is such that $c_p(n, \delta) = o\left(\sqrt{T}\right)$. Then, under Assumption D1,for any π in the range $0 < \pi < 1$, $d_T > 0$ and bounded in T, and for some $C_i, c > 0$ for $i = 0, 1$,*

$$\Pr\left[\left|t_x\right| > c_p(n, \delta)\left|\theta = 0\right.\right] \leq \exp\left[\frac{-(1-\pi)^2\sigma_{e,(T)}^2\sigma_{x,(T)}^2 c_p^2(n, \delta)}{2(1+d_T)^2\omega_{xe,T}^2}\right] + \exp\left(-C_0T^{C_1}\right), \tag{B.138}$$

where

$$t_x = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}},$$

$$\sigma^2_{e,(T)} = E\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right),\ \sigma^2_{x,(T)} = E\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}\right),$$

and

$$\omega^2_{xe,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[(u_{x,t}\eta_t)^2\right].$$

Under $\sigma_t^2 = \sigma^2$ *and/or* $E\left(u^2_{x,t}\right) = \sigma^2_{xt} = \sigma^2_x$*, for all* $t = 1, 2, ..., T$*,*

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] \leq \exp\left[\frac{-(1-\pi)^2 c_p^2(n,\delta)}{2(1+d_T)^2}\right] + \exp\left(-C_0 T^{C_1}\right).$$

In the case where $\theta > 0$*, and assuming that there exists* T_0 *such that for all* $T > T_0$*,* $\lambda_T - c_p(n,\delta)/\sqrt{T} > 0$*, where* $\lambda_T = \theta/\left(\sigma_{x,(T)}\sigma_{e,(T)}\right)$*, then for* $d_T > 0$ *and bounded in* T *and some* $C_i > 0$*,* $i = 0, 1, 2$*, we have*

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta \neq 0\right] > 1 - \exp\left(-C_0 T^{C_1}\right). \tag{B.139}$$

Under Assumption D2, for some $C_0, C_1, C_2 > 0$*,*

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] \leq \exp\left[-C_2 c_p(n,\delta)^{s/(s+2)}\right] + \exp\left(-C_0 T^{C_1}\right), \tag{B.140}$$

and

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta \neq 0\right] > 1 - \exp\left(-C_0 T^{C_1}\right). \tag{B.141}$$

Proof. We start under Assumption D1 and in the end note the steps that differ under Assumption D2. We recall that the DGP, given by (7), can be written as

$$\boldsymbol{y} = a\boldsymbol{\tau}_T + \mathbf{X}_k\boldsymbol{\beta} + \boldsymbol{u} = a\boldsymbol{\tau}_T + \mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}$$

where $\mathbf{X}_a$ is a subset of $\mathbf{Q}$. Recall that $\mathbf{Q}_x = (\mathbf{Q}, \boldsymbol{x})$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, $\mathbf{M}_{qx} = \mathbf{I}_T - \mathbf{Q}_x(\mathbf{Q}_x'\mathbf{Q}_x)^{-1}\mathbf{Q}_x'$. Then, $\mathbf{M}_q\mathbf{X}_a = \mathbf{0}$, and let $\mathbf{M}_q\mathbf{X}_b = (\boldsymbol{x}_{bq,1}, \boldsymbol{x}_{bq,2}, ..., \boldsymbol{x}_{bq,T})'$. Then,

$$t_x = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}} = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}} + \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}.$$

Let $\theta = E\left(T^{-1}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\right)\boldsymbol{\beta}_b$, $\boldsymbol{\eta} = \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}$, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_T)'$, and write (B.53) as

$$t_x = \frac{\sqrt{T}\theta}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}} + \frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}.$$

First consider the case where $\theta = 0$, and note that in this case

$$t_x = \frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)^{-1/2}\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)}}.$$

Now by (B.46) of Lemma A9 and (B.121) of Lemma D2, we have

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] = \Pr\left[\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)^{-1/2}\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)}}\right| > c_p(n,\delta)\,|\theta = 0\right] \leq \tag{B.142}$$

$$\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)^{-1/2}\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T}}{\sigma_{e,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left(-C_0T^{C_1}\right).$$

Then, by Lemma F1, under Assumption D1 and defining $\boldsymbol{\alpha}(\boldsymbol{X}_T) = \left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)^{-1/2}\boldsymbol{x}'\mathbf{M}_q$ where $\boldsymbol{\alpha}(\boldsymbol{X}_T)$ is exogenous to y_t, $\boldsymbol{\alpha}(\boldsymbol{X}_T)'\boldsymbol{\alpha}(\boldsymbol{X}_T) = 1$ and by (B.121) of Lemma D2, we have,

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] \leq \exp\left[\frac{-(1-\pi)^2\sigma^2_{e,(T)}\sigma^2_{x,(T)}c_p^2(n,\delta)}{2(1+d_T)^2\omega^2_{xe,T}}\right] + \exp\left(-C_0T^{C_1}\right) \tag{B.143}$$

where

$$\omega^2_{xe,T} = \frac{1}{T}\sum_{t=1}^{T}E\left[(u_{x,t}\eta_t)^2\right] = \frac{1}{T}\sum_{t=1}^{T}E\left[u^2_{x,t}\left(\boldsymbol{x}'_{b,t}\boldsymbol{\beta}_b + u_t\right)^2\right],$$

and $u_{x,t}$, being the error in the regression of x_t on $\mathbf{Q}$, is defined by (11). Since by assumption u_t are distributed independently of $u_{x,t}$ and $\boldsymbol{x}_{b,t}$, then

$$\omega^2_{xe,T} = \frac{1}{T}\sum_{t=1}^{T}E\left[u^2_{x,t}\left(\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b\right)^2\right] + \frac{1}{T}\sum_{t=1}^{T}E\left(u^2_{xt}\right)E\left(u^2_t\right),$$

where $\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b$ is the t-th element of $\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b$. Furthermore $E\left[u^2_{x,t}\left(\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b\right)^2\right] = E\left(u^2_{x,t}\right)E\left(\boldsymbol{x}'_{bq,t}\boldsymbol{\beta}_b\right)^2 = E\left(u^2_{x,t}\right)\boldsymbol{\beta}'_bE\left(\boldsymbol{x}_{bq,t}\boldsymbol{x}'_{bq,t}\right)\boldsymbol{\beta}_b$, noting that under $\theta = 0$, $u_{x,t}$ and $\boldsymbol{x}_{b,t}$ are independently distributed. Hence

$$\omega^2_{xe,T} = \frac{1}{T}\sum_{t=1}^{T}E\left(u^2_{x,t}\right)\boldsymbol{\beta}'_bE\left(\boldsymbol{x}_{bq,t}\boldsymbol{x}'_{bq,t}\right)\boldsymbol{\beta}_b + \frac{1}{T}\sum_{t=1}^{T}E\left(u^2_{xt}\right)E\left(u^2_t\right)$$

Similarly

$$\begin{aligned}\sigma^2_{e,(T)} &= E\left(T^{-1}\boldsymbol{e}'\boldsymbol{e}\right) = E\left(T^{-1}\boldsymbol{\eta}'\mathbf{M}_{qx}\boldsymbol{\eta}\right) = E\left[T^{-1}\left(\mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}\right)'\mathbf{M}_{qx}\left(\mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}\right)\right] \\ &= \boldsymbol{\beta}'_bE\left(T^{-1}\mathbf{X}'_b\mathbf{M}_{qx}\mathbf{X}_b\right)\boldsymbol{\beta}_b + \frac{1}{T}\sum_{t=1}^{T}E\left(u^2_t\right),\end{aligned}$$

and since under $\theta = 0$, $\boldsymbol{x}$ being a noise variable will be distributed independently of $\mathbf{X}_b$, then $E\left(T^{-1}\mathbf{X}_b'\mathbf{M}_{qx}\mathbf{X}_b\right) = E\left(T^{-1}\mathbf{X}_b'\mathbf{M}_q\mathbf{X}_b\right)$, and we have

$$\begin{aligned}\sigma^2_{e,(T)} &= \boldsymbol{\beta}_b' E\left(T^{-1}\mathbf{X}_b'\mathbf{M}_q\mathbf{X}_b\right)\boldsymbol{\beta}_b + \frac{1}{T}\sum_{t=1}^{T} E\left(u_t^2\right) \\ &= \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\beta}_b' E\left(\boldsymbol{x}_{bq,t}\boldsymbol{x}_{bq,t}'\right)\boldsymbol{\beta}_b + \frac{1}{T}\sum_{t=1}^{T} E\left(u_t^2\right).\end{aligned}$$

Using (B.55) and (B.56), it is now easily seen that if either $E\left(u_{x,t}^2\right) = \sigma^2_{ux}$ or $E\left(u_t^2\right) = \sigma^2$, for all t, then we have $\omega^2_{xe,T} = \sigma^2_{e,(T)}\sigma^2_{x,(T)}$, and hence

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] \le \exp\left[\frac{-(1-\pi)^2 c_p^2(n,\delta)}{2\left(1+d_T\right)^2}\right] + \exp\left(-C_0 T^{C_1}\right).$$

giving a rate that does not depend on error variances. Next, we consider $\theta \neq 0$. By (B.45) of Lemma A9, for $d_T > 0$,

$$\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(\boldsymbol{e}'\boldsymbol{e}/T\right)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}\right| > c_p(n,\delta)\right] \le \Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left(-C_0 T^{C_1}\right).$$

We then have

$$\begin{aligned}\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sigma_{e,(T)}\sigma_{x,(T)}} &= \frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}} \\ &= \frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}.\end{aligned}$$

Then

$$\begin{aligned}&\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) \\ &= 1 - \Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| \le \frac{c_p(n,\delta)}{1+d_T}\right).\end{aligned}$$

We note that, by Lemma A12,

$$\begin{aligned}&\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| \le \frac{c_p(n,\delta)}{1+d_T}\right) \\ &\le \Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T} - \theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{T^{1/2}|\theta|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p(n,\delta)}{1+d_T}\right).\end{aligned}$$

But $(T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}-\theta)$ is the average of a martingale difference process and so

$$\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{T}-\theta\right)}{\sigma_{e,(T)}\sigma_{x,(T)}}\right|>\frac{T^{1/2}\left|\theta\right|}{\sigma_{e,(T)}\sigma_{x,(T)}}-\frac{c_p\left(n,\delta\right)}{1+d_T}\right) \tag{B.144}$$
$$\leq \exp\left[-C_1\left(T^{1/2}\left(\frac{T^{1/2}\left|\theta\right|}{\sigma_{e,(T)}\sigma_{x,(T)}}-\frac{\theta c_p\left(n,\delta\right)}{1+d_T}\right)\right)^{s/(s+2)}\right].$$

So overall

$$\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(\boldsymbol{e}'\boldsymbol{e}/T\right)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}\right|>c_p\left(n,\delta\right)\right]>1-\exp\left(-C_0T^{C_1}\right)$$
$$-\exp\left[-C_1\left(T^{1/2}\left(\frac{T^{1/2}\left|\theta\right|}{\sigma_{e,(T)}\sigma_{x,(T)}}-\frac{\theta c_p\left(n,\delta\right)}{1+d_T}\right)\right)^{s/(s+2)}\right].$$

Finally, we note the changes needed to the above arguments when Assumption D2 holds, rather than D1. (B.140) follows if in (B.142) we use (B.125) of Lemma D2 rather than (B.121) and, in (B.143), we use Lemma F2 rather than Lemma F1 and, again, we use (B.125) of Lemma D2 rather than (B.121). (B.140) follows again by using (B.125) of Lemma D2 rather than (B.121). ■

Remark D3 *We note that the above proof makes use of Lemmas F1 and F2. Alternatively one can use (B.45) of Lemma A9 in (B.142)-(B.143), rather that (B.46) of Lemma A9 and use the same line of proof as that provided in Lemma A10. However, we consider this line of proof as Lemmas F1 and F2 are of independent interest.*

E. Lemmas for the deterministic case

Lemmas E1 and E2 provide the necessary justification for the case where x_{it} are bounded deterministic sequences, by replacing Lemmas A6 and A10.

Lemma E1 *Let x_{it}, $i=1,2,...,n$, be a set of bounded deterministic sequences and u_t satisfy Assumption 2 and condition (10) of Assumption 4, and consider the data generating process (6) with k signal variables $x_{1t},x_{2t},...,x_{kt}$. Let $\boldsymbol{q}_{\cdot t}=(q_{1,t},q_{2,t},...,q_{l_T,t})'$ contain a constant and a subset of $\boldsymbol{x}_{nt}=(x_{1t},x_{2t},...,x_{nt})'$. Let $\eta_t=\boldsymbol{x}_{b,t}\boldsymbol{\beta}_b+u_{\eta,t}$, where $\boldsymbol{x}_{b,t}$ contains all signals that do not belong to $\boldsymbol{q}_{\cdot t}$. Let $\boldsymbol{\Sigma}_{qq}=\mathbf{Q}'\mathbf{Q}/T$ be invertible for all T, and $\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_{FF}=O\left(\sqrt{l_T}\right)$, where $\mathbf{Q}=(\boldsymbol{q}_{1\cdot},\boldsymbol{q}_{2\cdot},...,\boldsymbol{q}_{l_T\cdot})$ and $\boldsymbol{q}_{i\cdot}=(q_{i1},q_{i2},...,q_{iT})'$, for $i=1,2,...,l_T$. Suppose that Assumption 5 holds for x_{it} and $\boldsymbol{q}_{\cdot t}$, and u_t and $\boldsymbol{q}_{\cdot t}$. Let $u_{x_i,T}$ be as in (11), such that $\sup_{i,j}\lim_{T\to\infty}\frac{\left\|\mathbf{q}_i'\boldsymbol{u}_{x_j,T}\right\|}{T^{1/2}}<C<\infty$, and let $\hat{\boldsymbol{u}}_{x_i}=(\hat{u}_{x_i,1},\hat{u}_{x_i,2},...,\hat{u}_{x_i,T})'=\mathbf{M}_q\boldsymbol{x}_i$, $\boldsymbol{x}_i=(x_{i1},x_{i2},...,x_{iT})'$, $\hat{\boldsymbol{u}}_\eta=(\hat{u}_{\eta,1},\hat{u}_{\eta,2},...,\hat{u}_{\eta,T})'=\mathbf{M}_q\boldsymbol{\eta}$, $\boldsymbol{\eta}=(\eta_1,\eta_2,...,\eta_T)'$, $\mathbf{M}_q=\mathbf{I}_T-\mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}$,*

$\mathcal{F}_t = \mathcal{F}_t^x \cup \mathcal{F}_t^u$, $\mu_{x_i\eta,t} = E\left(u_{x_i,t}u_{\eta,t}\left|\mathcal{F}_{t-1}\right.\right)$, $\omega^2_{x_i\eta,1,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[\left(x_{it}\eta_t - E\left(x_{it}\eta_t\left|\mathcal{F}_{t-1}\right.\right)\right)^2\right]$ and $\omega^2_{x_i\eta,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[\left(u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t}\right)^2\right]$. Then, for any π in the range $0 < \pi < 1$, we have, under Assumption 3,

$$\Pr\left(\left|\sum_{t=1}^{T} x_{it}\eta_t - E\left(x_{it}\eta_t\left|\mathcal{F}_{t-1}\right.\right)\right| > \zeta_T\right) \leq \exp\left[\frac{-\left(1-\pi\right)^2\zeta_T^2}{2T\omega^2_{x_i\eta,1,T}}\right], \tag{B.145}$$

where $\zeta_T = O\left(T^{\lambda}\right)$, and $(s+1)/(s+2) \geq \lambda$. If $(s+1)/(s+2) < \lambda$,

$$\Pr\left(\left|\sum_{t=1}^{T} x_{it}\eta_t - E\left(x_{it}\eta_t\left|\mathcal{F}_{t-1}\right.\right)\right| > \zeta_T\right) \leq \exp\left[-C_0\zeta_T^{s/(s+1)}\right],$$

for some $C_0 > 0$. If it is further assumed that $l_T = O\left(T^d\right)$, for some λ and d such that $d < 1/3$, and $1/2 \leq \lambda \leq (s+1)/(s+2)$, then

$$\Pr\left(\left|\sum_{t=1}^{T}\left(\hat{u}_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t}\right)\right| > \zeta_T\right) \leq C_2\exp\left[\frac{-\left(1-\pi\right)^2\zeta_T^2}{2T\omega^2_{x_i\eta,T}}\right] + \exp\left(-C_0T^{C_1}\right).$$

for some $C_0, C_1, C_2 > 0$. Otherwise, if $\lambda > (s+1)/(s+2)$,

$$\Pr\left(\left|\sum_{t=1}^{T}\left(\hat{u}_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t}\right)\right| > \zeta_T\right) \leq \exp\left[-C_2\zeta_T^{s/(s+1)}\right] + \exp\left(-C_0T^{C_1}\right),$$

for some $C_0, C_1, C_2 > 0$.

Proof. Note that all results used in this proof hold both for sequences and triangular arrays. (B.145) follows immediately given our assumptions and Lemma A3. We proceed to prove the rest of the lemma. Note that now $\hat{\boldsymbol{u}}_{x_i}$ is a bounded deterministic vector and $\boldsymbol{u}_{x_i} = (u_{x_i,1}, u_{x_i,2}, ..., u_{x_i,T})'$ a segment of dimension T of its limit. We first note that

$$\begin{aligned}\sum_{t=1}^{T}\left(\hat{u}_{x_i,t}\hat{u}_{\eta,t} - \mu_{x_i\eta,t}\right) &= \hat{\boldsymbol{u}}'_{x_i}\hat{\boldsymbol{u}}_{\eta} - \sum_{t=1}^{T}\mu_{x_i\eta,t} = \boldsymbol{u}'_{x_i}\mathbf{M}_q\boldsymbol{u}_{\eta} - \sum_{t=1}^{T}\mu_{x_i\eta,t} \\ &= \sum_{t=1}^{T}\left(u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t}\right) - \left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)\boldsymbol{\Sigma}_{qq}^{-1}\left(\mathbf{Q}'\boldsymbol{u}_{\eta}\right),\end{aligned}$$

where $\boldsymbol{u}_x = (u_{x,1}, u_{x,2}, ..., u_{x,T})'$ and $\boldsymbol{u}_{\eta} = (u_{\eta,1}, u_{x,2}, ..., u_{\eta,T})'$. By (B.59) and for any $0 < \pi_i < 1$ such that $\sum_{i=1}^{2}\pi_i = 1$,we have

$$\begin{aligned}\Pr\left(\left|\sum_{t=1}^{T}\left(\hat{u}_{x_i,t}\hat{u}_{\eta,t} - \mu_{x_i\eta,t}\right)\right| > \zeta_T\right) &\leq \Pr\left(\left|\sum_{t=1}^{T}\left(u_{x_i,t}u_{\eta,t} - \mu_{x_i\eta,t}\right)\right| > \pi_1\zeta_T\right) \\ &+ \Pr\left(\left|\left(T^{-1}\boldsymbol{u}'_{x_i}\mathbf{Q}\right)\boldsymbol{\Sigma}_{qq}^{-1}\left(\mathbf{Q}'\boldsymbol{u}_{\eta}\right)\right| > \pi_2\zeta_T\right).\end{aligned}$$

Also applying (B.60) to the last term of the above we obtain

$$\begin{aligned}
&\Pr\left(\left|\left(T^{-1}\boldsymbol{u}_{x_i}'\mathbf{Q}\right)\boldsymbol{\Sigma}_{qq}^{-1}\left(\mathbf{Q}'\boldsymbol{u}_{\eta}\right)\right|>\pi_2\zeta_T\right)\\
&\leq \Pr\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F\left\|T^{-1}\boldsymbol{u}_{x_i}'\mathbf{Q}\right\|_F\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F>\pi_2\zeta_T\right)\\
&\leq \Pr\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F>\frac{\pi_2\zeta_T}{\delta_T}\right)+\Pr\left(T^{-1}\left\|\boldsymbol{u}_{x_i}'\mathbf{Q}\right\|_F\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F>\pi_2\delta_T\right)\\
&\leq \Pr\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F>\frac{\pi_2\zeta_T}{\delta_T}\right)+\Pr\left(\left\|\boldsymbol{u}_{x_i}'\mathbf{Q}\right\|_F>\left(\pi_2\delta_T T\right)^{1/2}\right)\\
&+\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F>\left(\pi_2\delta_T T\right)^{1/2}\right),
\end{aligned}$$

where $\delta_T>0$ is a deterministic sequence. In what follows we set $\delta_T=O\left(\zeta_T^{\alpha}\right)$, with $0<\alpha<\lambda$, so that ζ_T/δ_T is rising in T. Overall

$$\begin{aligned}
&\Pr\left(\left|\sum_{t=1}^{T}\left(\hat{u}_{x,t}u_{\eta,t}-\mu_{x\eta,t}\right)\right|>\zeta_T\right) \qquad \text{(B.146)}\\
&\leq \Pr\left(\left|\sum_{t=1}^{T}\left(u_{x,t}u_{\eta,t}-\mu_{x\eta,t}\right)\right|>\pi_1\zeta_T\right)+\Pr\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F>\frac{\pi_2\zeta_T}{\delta_T}\right)\\
&+\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F>\left(\pi_2\delta_T T\right)^{1/2}\right)+\Pr\left(\left\|\boldsymbol{u}_{x}'\mathbf{Q}\right\|_F>\left(\pi_2\delta_T T\right)^{1/2}\right).
\end{aligned}$$

We consider the four terms of the above, and note that since by assumption $\{q_{it}u_{\eta,t}\}$ are martingale difference sequences and satisfy the required probability bound conditions of Lemma A4, and $\{q_{it}u_{x_i,t}\}$ are bounded sequences, then for some $C,c>0$ we have[7]

$$\sup_i \Pr\left(\left\|\mathbf{q}_i'\boldsymbol{u}_{\eta}\right\|>\left(\pi_2\delta_T T\right)^{1/2}\right)\leq \exp\left(-C_0T^{C_1}\right)$$

and as long as $l_T=o\left(\delta_T\right)$,

$$\Pr\left(\left\|\boldsymbol{u}_{x}'\mathbf{Q}\right\|_F>\left(\pi_2\delta_T T\right)^{1/2}\right)=0$$

Also, since $\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F^2=\sum_{j=1}^{l_T}\left(\sum_{t=1}^{T}q_{jt}u_t\right)^2$,

$$\begin{aligned}
&\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F>\left(\pi_2\delta_T T\right)^{1/2}\right)\\
&=\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F^2>\pi_2\delta_T T\right)\\
&\leq \sum_{j=1}^{l_T}\Pr\left[\left(\sum_{t=1}^{T}q_{jt}u_{\eta,t}\right)^2>\frac{\pi_2\delta_T T}{l_T}\right]\\
&=\sum_{j=1}^{l_T}\Pr\left[\left|\sum_{t=1}^{T}q_{jt}u_{\eta,t}\right|>\left(\frac{\pi_2\delta_T T}{l_T}\right)^{1/2}\right],
\end{aligned}$$

[7] The required probability bound on u_{xt} follows from the probability bound assumptions on x_t and on q_{it}, for $i=1,2,...,l_T$, even if $l_T\to\infty$. See also Lemma A5.

which upon using (B.22) yields (for some $C, c > 0$)

$$\Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{\eta}\right\|_F > (\pi_2\delta_T T)^{1/2}\right) \leq l_T \exp\left(-CT^c\right), \ \Pr\left(\left\|\mathbf{Q}'\boldsymbol{u}_{x}\right\| > (\pi_2\delta_T T)^{1/2}\right) = 0.$$

Further, it is easy to see that

$$\Pr\left(\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F > \frac{\pi_2\zeta_T}{\delta_T}\right) = 0$$

as long as $\frac{\zeta_T}{\delta_T l_T^{1/2}} \to \infty$. But as long as $l_T = o\left(T^{1/3}\right)$, there exists a sequence δ_T such that $\zeta_T/\delta_T \to \infty$, $l_T = o\left(\delta_T\right)$ and $\frac{\zeta_T}{\delta_T l_T^{1/2}} \to \infty$ as required, establishing the required result. ■

Lemma E2 *Let y_t, for $t = 1, 2, ..., T$, be given by the data generating process (6) and suppose that $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$ are bounded deterministic sequences, and u_t satisfy Assumption 2 and condition (10) of Assumption 4. Let $\boldsymbol{q}_{\cdot t} = (q_{1,t}, q_{2,t}, ..., q_{l_T,t})'$ contain a constant and a subset of $\boldsymbol{x}_{nt} = (x_{1t}, x_{2t}, ..., x_{nt})'$, and let $\eta_t = \boldsymbol{x}_{b,t}\boldsymbol{\beta}_b + u_t$, where $\boldsymbol{x}_{b,t}$ is $k_b \times 1$ dimensional vector of signal variables that do not belong to $\boldsymbol{q}_{\cdot t}$. Assume that $\boldsymbol{\Sigma}_{qq} = \mathbf{Q}'\mathbf{Q}/T$ is invertible for all T, and $\left\|\boldsymbol{\Sigma}_{qq}^{-1}\right\|_F = O\left(\sqrt{l_T}\right)$, where $\mathbf{Q} = (\boldsymbol{q}_{1\cdot}, \boldsymbol{q}_{2\cdot}, ..., \boldsymbol{q}_{l_T\cdot})$ and $\boldsymbol{q}_{i\cdot} = (q_{i1}, q_{i2}, ..., q_{iT})'$, for $i = 1, 2, ..., l_T$. Moreover, let $l_T = o(T^{1/4})$ and suppose that Assumption 5 holds for x_{it} and $\boldsymbol{q}_{\cdot t}$, and u_t and $\boldsymbol{q}_{\cdot t}$. Define $\boldsymbol{x} = (x_1, x_2, ..., x_T)'$, $\boldsymbol{y} = (y_1, y_2, ..., y_T)'$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, and $\theta = T^{-1}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b$, where $\mathbf{X}_b$ is $T \times k_b$ matrix of observations on $\boldsymbol{x}_{b,t}$. Let $u_{x_i,T}$ be as in (11), such that $\sup_{i,j}\lim_{T\to\infty}\frac{\left\|\mathbf{q}_i'\boldsymbol{u}_{x_j,T}\right\|}{T^{1/2}} < C < \infty$. Let $\boldsymbol{e} = (e_1, e_2, ..., e_T)'$ be the $T \times 1$ vector of residuals in the linear regression model of y_t on $\boldsymbol{q}_{\cdot t}$ and x_t. Then, for any π in the range $0 < \pi < 1$, $d_T > 0$ and bounded in T, and for some $C_i > 0$ for $i = 0, 1$,*

$$\Pr\left[|t_x| > c_p\left(n, \delta\right)|\theta = 0\right] \leq \exp\left[\frac{-\left(1-\pi\right)^2\sigma_{u,(T)}^2\sigma_{x,(T)}^2 c_p^2\left(n,\delta\right)}{2\left(1+d_T\right)^2\omega_{xu,T}^2}\right] + \exp\left(-C_0T^{C_1}\right),$$

where

$$t_x = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(\boldsymbol{e}'\boldsymbol{e}/T\right)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}},$$

$\sigma_{u,(T)}^2$ and $\sigma_{x,(T)}^2$ are defined by (B.39) and (B.34), and

$$\omega_{xu,T}^2 = \frac{1}{T}\sum_{t=1}^{T}\sigma_{xt}^2\sigma_t^2,$$

Under $\sigma_t^2 = \sigma^2$ and/or $\sigma_{xt}^2 = \sigma_x^2$ for all $t = 1, 2, ..., T$,

$$\Pr\left[|t_x| > c_p\left(n, \delta\right)|\theta = 0\right] \leq \exp\left[\frac{-\left(1-\pi\right)^2 c_p^2\left(n,\delta\right)}{2\left(1+d_T\right)^2}\right] + \exp\left(-C_0T^{C_1}\right).$$

for some $C_0, C_1 > 0$. In the case where $\theta > 0$, and assuming that $c_p(n,\delta) = o(\sqrt{T})$, then for $d_T > 0$ and some $C_i > 0$, $i = 0, 1$, we have

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta \neq 0\right] > 1 - \exp\left(-C_0 T^{C_1}\right).$$

Proof. The model for $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = a\boldsymbol{\tau}_T + \mathbf{X}_k\boldsymbol{\beta} + \boldsymbol{u} = a\boldsymbol{\tau}_T + \mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}$$

where $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of ones, $\mathbf{X}_a$ is a subset of $\mathbf{Q}$. Let $\mathbf{Q}_x = (\mathbf{Q}, \boldsymbol{x})$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, $\mathbf{M}_{qx} = \mathbf{I}_T - \mathbf{Q}_x(\mathbf{Q}_x'\mathbf{Q}_x)^{-1}\mathbf{Q}_x'$. Then, $\mathbf{M}_q\mathbf{X}_a = \mathbf{0}$. $\mathbf{M}_q\mathbf{X}_b = (\boldsymbol{x}_{bq,1}, \boldsymbol{x}_{bq,2}, ..., \boldsymbol{x}_{bq,T})'$. Then,

$$t_x = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}} = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}} + \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}.$$

Let

$$\boldsymbol{\eta} = \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{u}, \quad \boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_T)'$$

$$\theta = T^{-1}\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b,$$
$$\sigma^2_{e,(T)} = E(\boldsymbol{e}'\boldsymbol{e}/T) = E\left(\frac{\boldsymbol{\eta}'\mathbf{M}_{qx}\boldsymbol{\eta}}{T}\right), \ \sigma^2_{x,(T)} = E\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right),$$

and write (B.53) as

$$t_x = \frac{\sqrt{T}\theta}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}} + \frac{T^{-1/2}\left[\boldsymbol{x}'\mathbf{M}_q\,\boldsymbol{\eta} - E(\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta})\right]}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}.$$

$$\boldsymbol{x}'\mathbf{M}_q\,\boldsymbol{\eta} - E(\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}) = \left[\boldsymbol{x}'\mathbf{M}_q\,\boldsymbol{u} - E(\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u})\right],$$
$$\frac{(\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b)'(\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b)}{T} = \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{x}'_{bq,1}\boldsymbol{\beta}_b\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\sigma^2_{xbt} = \sigma^2_{b,(T)}.$$

Then, we consider two cases: $\frac{\boldsymbol{x}'\mathbf{M}_q\mathbf{X}_b\boldsymbol{\beta}_b}{T} := \theta = 0$ and $\theta \neq 0$. We consider each in turn. First, we consider $\theta = 0$ and note that

$$t_x = \frac{T^{-1/2}\left[\boldsymbol{x}'\mathbf{M}_q\,\boldsymbol{u} - E(\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u})\right]}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}.$$

By Lemma A9, we have

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] = \Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}\right| > c_p(n,\delta)\,|\theta = 0\right] \leq$$
$$\Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{\eta}}{\sigma_{x,(T)}\sigma_{e,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left(-C_0T^{C_1}\right).$$

By Lemma E1, it then follows that,

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] \leq \exp\left[\frac{-(1-\pi)^2\,\sigma^2_{e,(T)}\sigma^2_{x,(T)}c_p^2(n,\delta)}{2(1+d_T)^2\,\omega^2_{xe,T}}\right] + \exp\left(-C_0T^{C_1}\right)$$

where $\omega^2_{xe,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[(u_{x,t}\eta_t)^2\right]$. Note that, by independence of u_t with $u_{x,t}$ and $\boldsymbol{x}_{bq,t}$ we have

$$\omega^2_{xe,T} = \frac{1}{T}\sum_{t=1}^{T} E\left[(u_{x,t}\eta_t)^2\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[u^2_{x,t}\left(\boldsymbol{x}'_{bq,1}\boldsymbol{\beta}_b\right)^2\right] + E\left(u^2_{xt}\right)E\left(u^2_t\right).$$

By the deterministic nature of x_{it}, and under homoscedasticity for η_t, it follows that $\sigma^2_{e,(T)}\sigma^2_{x,(T)} = \omega^2_{xe,T}$, and so

$$\Pr\left[|t_x| > c_p(n,\delta)\,|\theta = 0\right] \leq \exp\left[\frac{-(1-\pi)^2\,c_p^2(n,\delta)}{2(1+d_T)^2}\right] + \exp\left(-C_0T^{C_1}\right).$$

giving a rate that does not depend on variances. Next, we consider $\theta \neq 0$. By Lemma A9, for $d_T > 0$,

$$\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{(\boldsymbol{e}'\boldsymbol{e}/T)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}\right| > c_p(n,\delta)\right] \leq \Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) + \exp\left(-C_0T^{C_1}\right).$$

We then have

$$\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sigma_{e,(T)}\sigma_{x,(T)}} = \frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}$$

Then,

$$\Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{c_p(n,\delta)}{1+d_T}\right) = 1 - \Pr\left(\left|\frac{T^{1/2}T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| \leq \frac{c_p(n,\delta)}{1+d_T}\right).$$

We note that

$$\Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sigma_{e,(T)}\sigma_{x,(T)}} + \frac{T^{1/2}\theta}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| \leq \frac{c_p(n,\delta)}{1+d_T}\right) \leq \Pr\left(\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{T^{1/2}|\theta|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p(n,\delta)}{1+d_T}\right).$$

But $T^{-1}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}$ is the average of a martingale difference process and so

$$\Pr\left(\left|\frac{T^{1/2}\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{u}}{T}\right)}{\sigma_{e,(T)}\sigma_{x,(T)}}\right| > \frac{T^{1/2}\left|\theta\right|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p\left(n,\delta\right)}{1+d_T}\right)$$
$$\leq \exp\left(-C_0T^{C_1}\right) + \exp\left[-C\left(T^{1/2}\left(\frac{T^{1/2}\left|\theta\right|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p\left(n,\delta\right)}{1+d_T}\right)\right)^{s/(s+2)}\right].$$

So overall,

$$\Pr\left[\left|\frac{T^{-1/2}\boldsymbol{x}'\mathbf{M}_q\boldsymbol{y}}{\sqrt{\left(\boldsymbol{e}'\boldsymbol{e}/T\right)\left(\frac{\boldsymbol{x}'\mathbf{M}_q\boldsymbol{x}}{T}\right)}}\right| > c_p\left(n,\delta\right)\right] > 1-\exp\left(-C_0T^{C_1}\right)$$
$$-\exp\left[-C\left(T^{1/2}\left(\frac{T^{1/2}\left|\theta\right|}{\sigma_{e,(T)}\sigma_{x,(T)}} - \frac{c_p\left(n,\delta\right)}{1+d_T}\right)\right)^{s/(s+2)}\right].$$

■

F. Supplementary lemmas for Sections B and C of the online theory supplement

Lemma F1 *Suppose that u_t, $t=1,2,...,T$, is a martingale difference process with respect to $\mathcal{F}_{t-1}^u$ and with constant variance σ^2, and there exist constants $C_0, C_1 > 0$ and $s>0$ such that $\Pr\left(|u_t|>\alpha\right) \leq C_0\exp\left(-C_1\alpha^s\right)$, for all $\alpha>0$. Let $\boldsymbol{X}_T = (\boldsymbol{x}_{l_T,1}, \boldsymbol{x}_{l_T,2},,.., \boldsymbol{x}_{l_T,T})$, where $\boldsymbol{x}_{l_T,t}$ is an $l_T \times 1$ dimensional vector of random variables, with probability measure given by $P(\boldsymbol{X}_T)$, and assume*

$$E\left(u_t\left|\mathcal{F}_T^x\right.\right) = 0, \text{ for all } t=1,2,...,T, \tag{B.147}$$

where $\mathcal{F}_T^x = \sigma\left(\boldsymbol{x}_{l_T,1}, \boldsymbol{x}_{l_T,2},,.., \boldsymbol{x}_{l_T,T}\right)$. Further assume that there exist functions $\boldsymbol{\alpha}(\boldsymbol{X}_T) = [\alpha_1(\boldsymbol{X}_T), \alpha_2(\boldsymbol{X}_T), ..., \alpha_T(\boldsymbol{X}_T)]'$ such that $0 < \sup_{\boldsymbol{X}_T}\boldsymbol{\alpha}(\boldsymbol{X}_T)'\boldsymbol{\alpha}(\boldsymbol{X}_T) \leq g_T$, for some sequence $g_T>0$. Then,

$$\Pr\left(\left|\sum_{t=1}^{T}\alpha_t(\boldsymbol{X}_T)u_t\right| > \zeta_T\right) \leq \exp\left(\frac{-\zeta_T^2}{2g_T\sigma^2}\right).$$

Proof. Define $\mathcal{A}_T$=$\left\{\left|\sum_{t=1}^{T}\alpha_t(\boldsymbol{X}_T)u_t\right| > \zeta_T\right\}$. Then,

$$\Pr\left(\mathcal{A}_T\right) = \int_{\boldsymbol{X}_T}\Pr\left(\mathcal{A}_T\left|\mathcal{F}_T^x\right.\right)P(\boldsymbol{X}_T) \leq \sup_{\boldsymbol{X}_T}\Pr\left(\mathcal{A}_T\left|\mathcal{F}_T^x\right.\right)\int_{\boldsymbol{X}_T}P(\boldsymbol{X}_T) = \sup_{\boldsymbol{X}_T}\Pr\left(\mathcal{A}_T\left|\mathcal{F}_T^x\right.\right)$$

■

But, by (B.147) and Lemma A3

$$\Pr\left(\mathcal{A}_T\left|\mathcal{F}_T^x\right.\right) \leq \exp\left(\frac{-\zeta_T^2}{2\sigma^2\sum_{t=1}^{T}\alpha_t^2(\boldsymbol{X}_T)}\right)$$

But

$$\sup_{\boldsymbol{X}_T} \exp\left(\frac{-\zeta_T^2}{2\sigma^2 \sum_{t=1}^T \alpha_t^2(\boldsymbol{X}_T)}\right) \leq \exp\left(\frac{-\zeta_T^2}{2g_T\sigma^2}\right),$$

proving the result.

Lemma F2 *Suppose that u_t, $t = 1, 2, ..., T$, is a zero mean mixing random variable with exponential mixing coefficients given by $\phi_k = a_{0k}\varphi^k$, $0 < \varphi < 1$, $a_{0k} < \infty$, $k = 1, ...,$ with constant variance σ^2, and there exist sufficiently large constants $C_0, C_1 > 0$ and $s > 0$ such that $\Pr(|u_t| > \alpha) \leq C_0 \exp(-C_1\alpha^s)$, for all $\alpha > 0$. Let $\boldsymbol{X}_T = (\boldsymbol{x}_{l_T,1}, \boldsymbol{x}_{l_T,2}, , .., \boldsymbol{x}_{l_T,T})$, where $\boldsymbol{x}_{l_T,t}$ is an $l_T \times 1$ dimensional vector of random variables, with probability measure given by $P(\boldsymbol{X}_T)$. Further assume that there exist functions $\boldsymbol{\alpha}(\boldsymbol{X}_T) = [\alpha_1(\boldsymbol{X}_T), \alpha_2(\boldsymbol{X}_T), ..., \alpha_T(\boldsymbol{X}_T)]'$ such that $0 < \sup_{\boldsymbol{X}_T} \boldsymbol{\alpha}(\boldsymbol{X}_T)'\boldsymbol{\alpha}(\boldsymbol{X}_T) \leq g_T$, for some sequence $g_T > 0$. Then,*

$$\Pr\left(\left|\sum_{t=1}^T \alpha_t(\boldsymbol{X}_T)u_t\right| > \zeta_T\right) \leq \exp\left(-\left(\frac{\zeta_T}{g_T^{1/2}\sigma}\right)^{s/(s+1)}\right).$$

Proof. Define $\mathcal{A}_T = \left\{\left|\sum_{t=1}^T \alpha_t(\boldsymbol{X}_T)u_t\right| > \zeta_T\right\}$ and consider $\mathcal{F}_T^x = \sigma(\boldsymbol{x}_{l_T,1}, \boldsymbol{x}_{l_T,2}, , .., \boldsymbol{x}_{l_T,T})$. Then,

$$\Pr(\mathcal{A}_T) = \int_{\boldsymbol{X}_T} \Pr(\mathcal{A}_T | \mathcal{F}_T^x) P(\boldsymbol{X}_T) \leq \sup_{\boldsymbol{X}_T} \Pr(\mathcal{A}_T | \mathcal{F}_T^x) \int_{\boldsymbol{X}_T} P(\boldsymbol{X}_T) = \sup_{\boldsymbol{X}_T} \Pr(\mathcal{A}_T | \mathcal{F}_T^x)$$

But, using Lemma 2 of Dendramis et al. (2015) we can choose C_0, C_1 such that

$$\Pr(\mathcal{A}_T | \mathcal{F}_T^x) \leq \exp\left[-\left(\frac{-\zeta_T}{\sigma\sqrt{\sum_{t=1}^T \alpha_t^2(\boldsymbol{X}_T)}}\right)^{s/(s+1)}\right],$$

and

$$\sup_{\boldsymbol{X}_T} \exp\left[-\left(\frac{-\zeta_T}{\sigma\sqrt{\sum_{t=1}^T \alpha_t^2(\boldsymbol{X}_T)}}\right)^{s/(s+1)}\right] \leq \exp\left[-\left(\frac{\zeta_T}{g_T^{1/2}\sigma}\right)^{s/(s+1)}\right],$$

thus establishing the desired result. ■

Lemma F3 *Let $\boldsymbol{A}_T = (a_{ij,T})$ be a $l_T \times l_T$ matrix and $\hat{\boldsymbol{A}}_T = (\hat{a}_{ij,T})$ be an estimator of $\boldsymbol{A}_T$. Let $\left\|\boldsymbol{A}_T^{-1}\right\|_F > 0$ and suppose that for some $s > 0$, any $b_T > 0$ and $C_0 > 0$*

$$\sup_{i,j} \Pr(|\hat{a}_{ij,T} - a_{ij,T}| > b_T) \leq \exp\left(-C_0 \left(T^{1/2}b_T\right)^{s/(s+2)}\right).$$

Then

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\| > b_T\right) \leq l_T^2 \exp\left(\frac{-C_0\left(T^{1/2}b_T\right)^{s/(s+2)}}{l_T^{s/(s+2)}\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)}\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)^{s/(s+2)}}\right) \tag{B.148}$$

$$+ l_T^2 \exp\left(-C_0 \frac{T^{s/2(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right),$$

where $\|\mathbf{A}\|$ *denotes the Frobenius norm of* $\mathbf{A}$*.*

Proof. First note that since $b_T > 0$, then

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > b_T\right) = \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F^2 > b_T^2\right)$$

$$= \Pr\left(\left[\sum_{j=1}^{l_T}\sum_{i=1}^{l_T}\left(\hat{a}_{ij,T} - a_{ij,T}\right)^2 > b_T^2\right]\right),$$

and using the probability bound result, (B.59), and setting $\pi_i = 1/l_T$, we have

$$\Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > b_T\right) \leq \sum_{j=1}^{l_T}\sum_{i=1}^{l_T}\Pr\left(\left|\hat{a}_{ij,T} - a_{ij,T}\right|^2 > l_t^{-2}b_T^2\right) \tag{B.149}$$

$$= \sum_{j=1}^{l_T}\sum_{i=1}^{l_T}\Pr\left(\left|\hat{a}_{ij,T} - a_{ij,T}\right| > l_t^{-1}b_T\right)$$

$$\leq l_T^2 \sup_{ij}\left[\Pr\left(\left|\hat{a}_{ij,T} - a_{ij,T}\right| > l_t^{-1}b_T\right)\right] = l_T^2 \exp\left(-C_0 T^{s/2(s+1)}\frac{b_T^{s/(s+2)}}{l_t^{s/(s+2)}}\right).$$

To establish (B.148) define the events

$$\mathcal{A}_T = \left\{\left\|\boldsymbol{A}_T^{-1}\right\|_F\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F < 1\right\} \text{ and } \mathcal{B}_T = \left\{\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\| > b_T\right\}$$

and note that by (2.15) of Berk (1974) if $\mathcal{A}_T$ holds we have

$$\left\|\hat{\boldsymbol{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\| \leq \frac{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}{1 - \left\|\boldsymbol{A}_T^{-1}\right\|_F\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}.$$

Hence

$$\Pr\left(\mathcal{B}_T\left|\mathcal{A}_T\right.\right) \leq \Pr\left(\frac{\left\|\boldsymbol{A}_T^{-1}\right\|_F^2\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F}{1 - \left\|\boldsymbol{A}_T^{-1}\right\|_F\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F} > b_T\right)$$

$$= \Pr\left(\left\|\hat{\boldsymbol{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right).$$

Note also that

$$\Pr\left(\mathcal{B}_T\right) = \Pr\left(\left\{\mathcal{B}_T \cap \mathcal{A}_T\right\} \cup \left\{\mathcal{B}_T \cap \mathcal{A}_T^C\right\}\right) = \Pr\left(\mathcal{B}_T|\mathcal{A}_T\right)\Pr\left(\mathcal{A}_T\right) + \Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right)\Pr\left(\mathcal{A}_T^C\right).$$

Furthermore

$$\begin{aligned}\Pr\left(\mathcal{A}_T^C\right) &= \Pr\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F\left\|\hat{\mathbf{A}}_T - \boldsymbol{A}_T\right\|_F > 1\right)\\ &= \Pr\left(\left\|\hat{\mathbf{A}}_T - \boldsymbol{A}_T\right\|_F > \left\|\boldsymbol{A}_T^{-1}\right\|_F^{-1}\right),\end{aligned}$$

and by (B.149) we have

$$\Pr\left(\mathcal{A}_T^C\right) \leq l_T^2 \exp\left(-C_0 T^{s/2(s+2)}\frac{b_T^{s/(s+2)}}{l_t^{s/(s+2)}}\right) = \exp\left(-C_0\frac{T^{s/2(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right).$$

Using the above result, we now have

$$\begin{aligned}\Pr\left(\mathcal{B}_T\right) \leq{}& \Pr\left(\left\|\hat{\mathbf{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right)\Pr\left(\mathcal{A}_T\right)\\ &+ \Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right)\exp\left(-C_0\frac{T^{s/2(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right).\end{aligned}$$

Furthermore, since $\Pr\left(\mathcal{A}_T\right) \leq 1$ and $\Pr\left(\mathcal{B}_T|\mathcal{A}_T^C\right) \leq 1$ then

$$\begin{aligned}\Pr\left(\mathcal{B}_T\right) = \Pr\left(\left\|\hat{\mathbf{A}}_T^{-1} - \boldsymbol{A}_T^{-1}\right\| > b_T\right) \leq{}& \Pr\left(\left\|\hat{\mathbf{A}}_T - \boldsymbol{A}_T\right\|_F > \frac{b_T}{\left\|\boldsymbol{A}_T^{-1}\right\|_F\left(\left\|\boldsymbol{A}_T^{-1}\right\|_F + b_T\right)}\right)\\ &+ \exp\left(-C_0\frac{T^{s/2(s+2)}}{\left\|\boldsymbol{A}_T^{-1}\right\|_F^{s/(s+2)} l_T^{s/(s+2)}}\right).\end{aligned}$$

Result (B.148) now follows if we apply (B.149) to the first term on the RHS of the above. ■

Lemma F4 *Consider the scalar random variable X_T, and the constants B and C. Then, if $C > |B| > 0$,*

$$\Pr\left(|X + B| > C\right) \leq \Pr\left(|X| > C - |B|\right). \tag{B.150}$$

Proof. The result follows by noting that $|X + B| \leq |X| + |B|$. ■

References

Bailey, N., M. H. Pesaran, and L. V. Smith (2018). A multiple testing approach to the regularisation of large sample correlation matrices. *CAFE Research Paper No. 14.05, revised 2018*.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics 2*, 489–502.

Dendramis, Y., L. Giraitis, and G. Kapetanios (2015). Estimation of random coefficient time varying covariance matrices for large datasets. *Mimeo*.

Freedman, D. A. (1975). On tail probabilities for martingales. *Annals of Probability 3*, 100–118.

Roussas, G. (1996). Exponential probability inequalities with some applications. *Statistica, Probability and Game Theory, IMS Lecture Notes - Monograph Series 30*, 303–319.

Wecker, W. E. (1978). A note on the time series which is the product of two stationary time series. *Stochastic Processes and their Applications 8*, 153–157.

White, H. and J. M. Wooldridge (1991). Some results on sieve estimation with dependent observations. In W. J. Barnett, J. Powell, and G. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 459–493. New York: Cambridge University Press.

Online Empirical Supplement to "A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models"

A. Chudik
Federal Reserve Bank of Dallas

G. Kapetanios
King's College, London

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

24 June 2018

1 Introduction

This supplement to Chudik, Kapetanios, and Pesaran (2018, hereafter CKP) provides a description of the individual methods employed in the empirical illustration, and additional empirical results. The empirical illustration is set out in Section 6 of CKP. Section 2 below describes the forecasting exercise, and Section 3 reports additional empirical results.

2 Description of the forecasting exercise

We forecast the U.S. GDP growth and CPI inflation using a set of macroeconomic variables. We use the smaller dataset considered in Stock and Watson (2012), which contains 109 series. The series are transformed by taking logarithms and/or differencing following Stock and Watson (2012).[1] After transformations, the available sample is 1960Q3:2008Q4, or $T = 194$. Let $\boldsymbol{\xi}_t = \left(\xi_{1t}, \xi_{2t}, ..., \xi_{n-1,t}\right)'$ be a vector of the 109 transformed variables. Define the $n \times 1$ vector $\mathbf{x}_t = (\boldsymbol{\xi}_t, y_t, y_{t-1}, y_{t-2}, y_{t-3})'$ considered below, where y_t is either the first-differenced log of real gross domestic product, or the second differenced log of consumer price index.

We are interested in forecasting y_{t+1} with the predictors in $\mathbf{x}_t$ and common factors $\mathbf{f}_t$ extracted from variables in $\mathbf{z}_t^s$, where $\mathbf{z}_t^s$ is the standardized $\mathbf{z}_t = \left(y_t, \boldsymbol{\xi}_t'\right)'$ (by subtracting its sample mean and dividing each series by its sample standard deviation). We consider:

(a) the AR(h) model,

$$y_t = \sum_{\ell=1}^{h} \rho_\ell y_{t-\ell} + v_t,$$

which we use as a benchmark. The lag order h is selected using the SBC criterion with the maximum number of lags set equal to $h_{\max} = 4$.

[1] For further details, see the online supplement of Stock and Watson (2012), in particular columns E and T of their Table B.1.

Data-rich forecasting methods are:

(b) The factor-augmented AR,

$$y_t = \sum_{\ell=1}^{h} \rho_\ell y_{t-\ell} + \boldsymbol{\gamma}' \mathbf{f}_{t-1} + v_t,$$

where $\mathbf{f}_t$ is $m \times 1$ vector of unobserved common factors extracted from variables in $\mathbf{z}_t^s$. We use Bai and Ng's PC_{p1} criterion to select the number of factors (m) with the maximum number of factors set to 5. The vector of unobserved factors, $\mathbf{f}_t$, is estimated using the method of principal components. Same as in the AR case, the lag order h is selected using the SBC criterion with the maximum number of lags set equal to $h_{\max} = 4$.

(c) Lasso method, implemented in the same way as described in Section 2 of the online Monte Carlo supplement of CKP using $\left(\mathbf{x}'_{t-1}, \mathbf{f}'_{t-1}\right)$ as the vector of predictors for y_t.

(d) Adaptive Lasso method, implemented in the same way as described in Section 2 of the online Monbte Carlo supplement of CKP using $\left(\mathbf{x}'_{t-1}, \mathbf{f}'_{t-1}\right)$ as the vector of predictors for y_t.

(e-g) OCMT method. We use OCMT described in CKP to select the relevant variables from the vector $\mathbf{x}_{t-1}$ to forecasts the target variable y_t. We set $p = 0.01$ (e), 0.05 (f) and 0.1 (g), and $(\delta, \delta^*) = (1, 2)$, and we always include c (intercept) , and $\mathbf{f}_{t-1}$ (lagged factors) in the testing regressions. Next, we use the selected variables together with c, and $\mathbf{f}_{t-1}$ in an ordinary least squares regression for y_t.

We use a rolling window of $T = 120$ time periods, which leaves us with the last $H = 74$ out-of-sample evaluation periods, 1990Q3-2008Q4. We also consider pre-crisis evaluation subsample, 1990Q3-2007Q2 with $H = 68$ periods, to evaluate the sensitivity of results to exclusion of the global financial crisis from the sample.

3 Results

Table 1 reports the root mean squared forecasting error (RMSFE) findings for all forecasting methods. Diebold-Mariano (DM) test statistics for testing $H_0 : E\left(\hat{v}_{ij,t}\right) = 0$, where $\hat{v}_{ij,t} = \hat{e}_{i,t}^2 - \hat{e}_{j,t}^2$ is the difference between the squared forecasting errors of methods i and j, are presented in Table 2. The DM statistics is computed assuming serially uncorrelated one-step-ahead forecasting errors. Specifically

$$DM_{ij} = \sqrt{H} \frac{\overline{\hat{v}}_{H,ij}}{\hat{\sigma}_{H,ij}}, \tag{1}$$

where $H = 68$ or 74 (depending on the evaluation period) is the length of the evaluation period, $\overline{\hat{v}}_{H,ij} = H^{-1} \sum_{t=T+1}^{T+H} \hat{v}_{ij,t}$ is the sample mean of $\hat{v}_{ij,t}$, and

$$\hat{\sigma}_{H,ij} = \sqrt{\frac{1}{H} \sum_{t=T+1}^{T+H} \hat{v}_{ij,t}^2}.$$

Table 1: RMSFE performance of the AR, factor-augmented AR, Lasso, adaptive Lasso, and OCMT methods

Evaluation sample:	Full 1990Q3-2008Q4		Pre-crisis 1990Q3-2007Q2	
	RMSFE ($\times$100)	Relative RMSFE	RMSFE ($\times$100)	Relative RMSFE
	Real output growth			
(a) AR benchmark	0.561	1.000	0.505	1.000
(b) Factor-augmented AR	0.484	0.862	0.470	0.930
(c) Lasso	0.510	0.910	0.465	0.922
(d) Adaptive Lasso	0.561	1.000	0.503	0.996
(e) OCMT, $p = 0.01$	0.495	0.881	0.479	0.948
(f) OCMT, $p = 0.05$	0.477	0.850	0.461	0.912
(g) OCMT, $p = 0.1$	0.490	0.874	0.464	0.918
	Inflation			
(a) $AR(1)$ benchmark	0.601	1.000	0.435	1.000
(b) Factor-augmented $AR(1)$	0.557	0.927	0.415	0.954
(c) Lasso	0.599	0.997	0.462	1.063
(d) Adaptive Lasso	0.715	1.190	0.524	1.205
(e) OCMT, $p = 0.01$	0.596	0.992	0.472	1.086
(f) OCMT, $p = 0.05$	0.590	0.982	0.464	1.068
(g) OCMT, $p = 0.1$	0.595	0.990	0.471	1.084

Notes: RMSFE is computed using a rolling forecasting scheme with a rolling window of 120 observations. We use the smaller dataset considered in Stock and Watson (2012) which contains 109 series. The series are transformed by taking logarithms and/or differencing following Stock and Watson (2012). The transformed series span 1960Q3 to 2008Q4 and are collected in the vector $\boldsymbol{\xi}_t$. Set of regressors in Lasso and adaptive-Lasso contains $h_{\max} = 4$ lags of y_t (lagged target variables), $\boldsymbol{\xi}_{t-1}$, and a lagged set of principal components obtained from the large dataset given by $(y_t, \boldsymbol{\xi}_t')'$. OCMT procedure is applied to regressions of y_t conditional on lagged principal components, with elements of $\boldsymbol{\xi}_{t-1}$ and $h_{\max} = 4$ lags of y_t considered one at a time. OCMT is reported for $\delta = 1$ in the first stage, and $\delta^* = 2$ in the subsequent stages of the OCMT procedure, and three choices of p, similarly to the MC section of CKP. The number of principal components in the factor-augmented AR, Lasso, adaptive Lasso, and OCMT methods is determined in a rolling scheme by using criterion PC_{p_1} of Bai and Ng (2002) (with the maximum number of PCs set to 5). See Section 2 for further details.

Table 2: DM statistics for the forecasting performance of the AR, factor-augmented AR, Lasso, adaptive Lasso, and OCMT methods

	DM_{ij} test statistics													
	Full evaluation sample: 1990Q3-2008Q4													
	Real output growth							**Inflation**						
Method pair i (below), j (on right)	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a) AR(1)	.	1.50	1.95	0.00	1.49	1.73	1.44	.	1.12	0.06	-2.55	0.12	0.28	0.14
(b) Factor-augmented $AR(1)$	-1.50	.	-0.67	-1.39	-0.59	0.43	-0.38	-1.12	.	-1.89	-2.06	-2.39	-2.07	-2.09
(c) Lasso	-1.95	0.67	.	-1.76	0.45	0.92	0.57	-0.06	1.89	.	-1.82	0.14	0.45	0.20
(d) Adaptive Lasso	0.00	1.39	1.76	.	1.29	1.56	1.31	2.55	2.06	1.82	.	1.61	1.69	1.62
(e) OCMT, $p = 0.01$	-1.49	0.59	-0.45	-1.29	.	1.32	0.24	-0.12	2.39	-0.14	-1.61	.	0.49	0.08
(f) OCMT, $p = 0.05$	-1.73	-0.43	-0.92	-1.56	-1.32	.	-1.21	-0.28	2.07	-0.45	-1.69	-0.49	.	-0.71
(g) OCMT, $p = 0.05$	-1.44	0.38	-0.57	-1.31	-0.24	1.21	.	-0.14	2.09	-0.20	-1.62	-0.08	0.71	.
	Pre-Crisis evaluation sample: 1990Q3-2007Q2													
(a) AR(1)	.	0.95	1.60	0.13	0.84	1.19	1.11	.	0.98	-1.13	-2.28	-1.54	-1.01	-1.18
(b) Factor-augmented $AR(1)$	-0.95	.	0.14	-0.88	-0.48	0.52	0.34	-0.98	.	-1.66	-2.31	-2.46	-2.21	-2.21
(c) Lasso	-1.60	-0.14	.	-1.39	-0.48	0.16	0.06	1.13	1.66	.	-1.78	-0.47	-0.07	-0.37
(d) Adaptive Lasso	-0.13	0.88	1.39	.	0.66	1.07	1.00	2.28	2.31	1.78	.	1.22	1.31	1.15
(e) OCMT, $p = 0.01$	-0.84	0.48	0.48	-0.66	.	1.22	0.82	1.54	2.46	0.47	-1.22	.	0.46	0.05
(f) OCMT, $p = 0.05$	-1.19	-0.52	-0.16	-1.07	-1.22	.	-0.33	1.01	2.21	0.07	-1.31	-0.46	.	-0.71
(g) OCMT, $p = 0.05$	-1.11	-0.34	-0.06	-1.00	-0.82	0.33	.	1.18	2.21	0.37	-1.15	-0.05	0.71	.

Notes: This table reports results for DM_{ij} statistics defined in (1). See also notes to Table 1.

References

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

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, forthcoming.

Stock, J. H. and M. W. Watson (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics 30*, 481–493.