Inference on the Change Point in High Dimensional Dynamic Graphical Models

Abhishek Kaul[a], Hongjin Zhang, Konstantinos Tsampourakis

Department of Mathematics and Statistics,
Washington State University, Pullman, WA 99164, USA.

Abstract

We propose a new estimator for the change point parameter in a dynamic high dimensional graphical model setting. We show that the proposed estimator retains sufficient adaptivity against plugin estimates of the edge structure of the underlying graphical models, in order to yield an $O(\psi^{-2})$ rate of convergence of the change point estimator in the integer scale. This rate is preserved while allowing high dimensionality as well as a diminishing jump size ψ, provided $s\log^{3/2}(p\vee T)=o\big(\surd(Tl_T)\big)$. Here s,p,T and l_T represent a sparsity parameter, model dimension, sampling period and the separation of the change point from its parametric boundary, respectively. Moreover, since the rate of convergence is free of s,p and logarithmic terms of T, it allows the existence of a limiting distribution valid in the high dimensional setting, which is then derived. The method does not assume an underlying Gaussian distribution. Theoretical results are supported numerically with monte carlo simulations.

Keywords: High dimensions, dynamic graphical models, change point, inference, limiting distribution.

1 Introduction

A large body of literature has been developed on the recovery of large network structures such as high dimensional graphical models. Such networks play a vital role in a variety of problems, for e.g., to serve as a representation of interactions between a set of nodes, to aid in classification problems, and to allow the implementation of classical dimension reduction techniques such as factor analysis amongst several other uses. Owing to their versatility, such models have been adopted in a variety of scientific fields such as in the field of neuroimaging, e.g., Cribben et al. [2012], where graphical models obtained from FMRI data are utilized to understand neurological network structures. In microbiome studies, e.g., Kaul et al. [2017], where such models have found use in geographical classification of persons based on their gut microbiome observations.

An undirected graphical model is a network where an edge between the $(i,j)^{th}$ nodes represents a non-zero $(i,j)^{th}$ entry of underlying the precision matrix, which is defined as the inverse of the covariance matrix. The reasoning for this network representation arises from the well known classical multivariate theory result, where under a Gaussian distribution, a zero valued $(i,j)^{th}$ entry of the precision matrix characterizes independence between the $(i,j)^{th}$ nodes, conditioned on all remaining nodes. Accordingly, the statistical problem in the recovery of a graphical model is equivalent to the estimation of the underlying precision matrix. In the high dimensional setting, where the dimension (p) of this matrix might diverge faster than the number of observations (T), it is now well understood that given sparsity assumptions on parameters (edge structure), these matrices can be consistently estimated by several different methods in the literature, for e.g., neighborhood selection and its variants [Meinshausen et al. [2006], Friedman et al. [2008] and Yuan [2010] among others] where edges are recovered locally for each node. Alternatively, direct global approaches to estimate the precision or covariance matrix via ℓ_1 regularization of ℓ_1 minimization [Banerjee et al. [2008], and Cai et al. [2011] among others].

[a]Email: abhishek.kaul@wsu.edu

In many problems of current scientific interest the assumption of stationarity of a network over an extended sampling period could be unrealistic and may lead to flawed inference. In the recent past dynamic graphical models capable of fitting evolving networks in a piecewise manner, characterized via one or more change points have received attention in the literature as tractable relaxation of the fairly rigid assumption of stationarity. In a Gaussian graphical model (GGM) setting, Kolar and Xing [2012] consider a fused lasso regularization together with a neighborhood selection approach. Angelosante and Giannakis [2011] propose dynamic programming in conjunction with neighborhood selection. Likelihood based approaches together with suitable regularization have been proposed in Kolar et al. [2010], Gibberd and Roy [2017], and Keshavarz et al. [2018], where the latter take a detection perspective. In similar setup, Londschien et al. [2019] provide a correction method to alleviate bias due to missingness. Atchade and Bybee [2017] provide a majorize-minimize algorithm allowing efficient implementation of the likelihood approach in comparison to a brute force search. Studies on other types of dynamic network structures are also available in the literature. Roy et al. [2017] provide a likelihood based approach for Markov random fields with a single change point. Bhattacharjee et al. [2018] and Wang et al. [2018] consider stochastic block models and provide least squares and cusum based methodologies, respectively.

In this article we consider the following change point model, where the change is in the covariance matrix or equivalently the precision matrix of a p-dimensional distribution.

$$z_t = \begin{cases} w_t, & t = 1, ..., \lfloor T\tau^0 \rfloor \\ x_t, & t = \lfloor T\tau^0 \rfloor + 1, ..., T \end{cases}, \tag{1.1}$$

here the observed variable is $z_t \in \mathbb{R}^p$, $t = 1, ..., T$. The variables $w_t, x_t \in \mathbb{R}^p$ are independent and zero mean subgaussian random variables (r.v.'s), with unknown covariance matrices Σ and Δ, respectively. These variables are not directly observed, in the sense that the change point parameter $\tau^0 \in (0, 1)$ is unknown. Thus it is apriori unknown from which of the two subgaussian distributions a realization z_t arises. We allow the dimension p to diverge potentially at an exponential rate, i.e., $\log p = o(T^\delta)$, while making a sparsity assumption to be specified in the following section. The parameters of interest here are the change point τ^0, and the unknown matrices Σ and Δ, with the former being of main interest in this article. The homogenous case of 'no change' where model (1.1) reduces to T i.i.d. observations of a subgaussian distribution with covariance Σ, occurring at $\tau^0 = 1$ is disallowed, i.e., our objective throughout the article is that of estimation and inference on τ^0 when it exists.

To aid further discussion on the main objectives of this article, we require additional notation. For any $p \times p$ matrix W, define a $(p-1)$-dimensional vector $W_{-i,j}$ as the j^{th} column of W with the i^{th} entry removed, and similarly define $W_{i,-j}$. Also define a $(p-1) \times (p-1)$ matrix $W_{-i,-j}$ as the sub-matrix of W with the i^{th} row and the j^{th} column removed. Now define the following parameter vectors in $\mathbb{R}^{p-1}$,

$$\mu^0_{(j)} = \Sigma^{-1}_{-j,-j}\Sigma_{-j,j}, \quad \text{and} \quad \gamma^0_{(j)} = \Delta^{-1}_{-j,-j}\Delta_{-j,j}, \quad \text{for each} \quad j = 1, ..., p. \tag{1.2}$$

The parameter vectors $\mu^0_{(j)}$, and $\gamma^0_{(j)}$'s play a fundamental role in neighborhood selection and the underlying graphical model structure. When $\mu^0_{(j)k} = 0$ (k^{th} component of $\mu^0_{(j)}$) $\Leftrightarrow$ the $(j, k)^{th}$ entry of the corresponding precision matrix is zero, and thus indicates the absence of an edge between these nodes in the corresponding graph. From a technical perspective, these coefficient vectors play an important role since they can be used to orthogonalize the j^{th} and the remaining components of the underlying distribution, i.e., if $w_t \in \mathbb{R}^p$ is a realization from a zero mean distribution with covariance Σ, then it is straightforward to see that the component w_{tj} and the vector $(w_{tj} - w^T_{t,-j}\mu^0_{(j)})$ are uncorrelated (or independent under Gaussianity). Given their fundamental role in characterizing the network structure, we use these coefficients to characterize the magnitude of the jump size across the two networks. Specifically, we let,

$$\eta^0_{(j)} = \mu^0_{(j)} - \gamma^0_{(j)}, \;\; j = 1, ..., p, \quad \xi_{2,2} = \Big(\sum_{j=1}^{p} \|\eta^0_{(j)}\|_2^2\Big)^{\frac{1}{2}}, \quad \text{and} \quad \psi = \frac{\xi_{2,2}}{\sqrt{p}}.^{\text{[b]}} \tag{1.3}$$

The quantities $\xi_{2,2}$, and ψ are representative of jump size, with the latter being a normalized version that plays a central role in our analysis. For $\xi_{2,2}$ or ψ to be non-zero, it is necessary that there are either edge connectivity changes in the underlying graphs or changes in magnitude of conditional dependence between nodes. An example of a structural change that the measure $\xi_{2,2}$ will be insensitive towards is the following. Let the two underlying covariance matrices be Σ and $\Delta = c\Sigma$, for any constant $0 < c < \infty$, in other words the covariance structures are different but the correlation structure is identical, thus the underlying graph structures are identical. It is straightforward to observe that in this case $\mu^0_{(j)} = \gamma^0_{(j)}$, $j = 1, ..., p$, and that the jump size $\xi_{2,2} = \psi = 0$. This characterization is somewhat similar to that of Kolar and Xing [2012], who define the jump size as $\min_j \|\eta^0_{(j)}\|_2$. The advantage of using the measure $\xi_{2,2}$ or ψ over $\min_j \|\eta^0_{(j)}\|_2$ is that the latter requires changes in each and every row and column of the precision matrix, whereas the former allows for sub-block changes of the precision matrix pre and post the change point. Other metrics of the jump size have also been utilized in the literature, e.g. $\|\Sigma - \Delta\|_F$ in Gibberd and Roy [2017], which is comparable to $\xi_{2,2}$ defined above.

The proposed estimator is described in the following. Consider any $z_t \in \mathbb{R}^p$, $t = 1, .., T$ and $z = (z_1, z_2, ..., z_T)^T \in \mathbb{R}^{T\times p}$, any $\mu_{(j)}, \gamma_{(j)} \in \mathbb{R}^{p-1}$, $j = 1, ..., p$, any $\tau \in (0, 1)$, $T \geq 2$, $p \geq 3$,[c] Let μ, and γ be the concatenation of $\mu'_{(j)}$s and $\gamma'_{(j)}$s. Then define the least squares loss,

$$Q(z, \tau, \mu, \gamma) = \frac{1}{T}\sum_{t=1}^{\lfloor T\tau \rfloor}\sum_{j=1}^{p}(z_{tj} - z^T_{t,-j}\mu_{(j)})^2 + \frac{1}{T}\sum_{t=(\lfloor T\tau \rfloor+1)}^{T}\sum_{j=1}^{p}(z_{tj} - z^T_{t,-j}\gamma_{(j)})^2.$$

Now suppose the availability of nuisance estimates $\hat\mu_{(j)}, \hat\gamma_{(j)} \in \mathbb{R}^{p-1}$, $j = 1, ..., p$ of the coefficient vectors defined in (1.2), such that the following bound is satisfied.

$$\max_{1\leq j\leq p}\Big(\|\hat\mu_{(j)} - \mu^0_{(j)}\|_2 \vee \|\hat\gamma_{(j)} - \gamma^0_{(j)}\|_2\Big) \leq c_u\sqrt{(1+\nu^2)}\frac{\sigma^2}{\kappa}\Big\{\frac{s\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}}, \tag{1.4}$$

with probability at least $1 - o(1)$. Here l_T is a sequence separating the change point parameter from the boundary of its parametric space, i.e. $\lfloor T\tau^0 \rfloor \wedge (T - \lfloor T\tau^0 \rfloor) \geq Tl_T$, (see, Condition A). The quantities σ, κ, ν, are other parameters defined in Section 2 (see, Condition B). Under this setup, define the plug in estimator $\tilde\tau$ as,

$$\tilde\tau := \tilde\tau(\hat\mu, \hat\gamma) = \arg\min_{\tau\in(0,1)} Q(z, \tau, \hat\mu, \hat\gamma). \tag{1.5}$$

Clearly, the estimator $\tilde\tau$ utilizes estimates of the potentially high dimensional edge parameters $\mu^0_{(j)}$, and $\gamma^0_{(j)}$, $j = 1, ..., p$. The inference results of Section 2 are agnostic about the choice of the estimator used to obtain these nuisance estimates, as long as they satisfy sufficient estimation properties,

[b] The jump sizes $\xi_{2,2}$, and ψ may depend on sampling period T, however this dependence is suppressed for clarity of exposition.

[c] We assume $p \geq 3$ throughout the article. This is done so that $\log(p) \geq 1$. This is not a necessary condition and is only assumed for clarity of exposition.

mainly, the ℓ_2 error bound (1.4). In Section 3 we present an estimator for these nuisance parameters which possesses the required theoretical guarantees, thereby making the methodology feasible in practice.

Section 2 studies the behavior of $\tilde{\tau}$ in a high dimensional setup $p >> T$. Our first main result obtains the rate of convergence of the proposed estimator $\tilde{\tau}$, where we show that $\big(\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big) = O_p(\psi^{-2})$, under weak conditions. It is important to note this rate is free of the dimension parameters s, p and other logarithmic terms of the sampling period T. In a change point model with a mean change of a random vector, a rate of convergence of this form is known to be optimal. Since the same is not known in a graphical model setting, thus we do not refer to the presented rate as optimal, although it is natural to suspect that it is indeed the case. Furthermore, the given rate is preserved while allowing a diminishing jump size ψ, under the sufficient condition that the model dimensions satisfy $s\log^{3/2}(p\vee T) = o\big(\surd(Tl_T)\big)$. Here s is a sparsity parameter to be defined in Section 2. To the best of our knowledge, the rate of convergence described above is sharper, and the minimum jump size assumption is significantly weaker than those available for existing estimators in the literature on dynamic high dimensional network models, for e.g., Kolar and Xing [2012] provide a rate of convergence $O_p(\psi^{-2}p\log T)$ under a minimum jump size assumption of order $O(p\log T/T)^{1/2}$, Gibberd and Roy [2017] provide a rate $O_p(\psi^{-2}p^2\log p)$ under a jump assumption of order $O\{p\surd(\log p^{\beta/2}/T)\}$, and Roy et al. [2017] provide a rate of $O_p\big(\psi^{-2}\log(pT)\big)$ under a jump assumption of order $O(\log pT)^{1/4}$. The former two articles are under a gaussian graphical model setting, whereas the latter is in a markov random field setting.

The sharper rate of convergence of the proposed estimator in the high dimenisonal setting allows for the existence of a limiting distribution. Our second main result derives this limiting distribution as the distribution of the minimizer of an asymmetric and off-center Brownian motion. This enables inference on τ^0, when it exists, i.e., to construct an asymptotically valid confidence interval for τ^0, when $\tau^0 < 1$. Here asymptotics are in the high dimensional sense, where the sample size T is diverging, and the dimension p can be fixed or diverging, at potentially an exponential rate of T. This limiting distribution has been studied in Bai [1997], where it was obtained as the limiting distribution of the change point estimate in the classical fixed p linear regression setting. The density function of this distribution is also available in the same article, thus enabling straightforward computation of quantiles, which can in turn be utilized to construct asymptotically valid confidence intervals.

An indirect but informative comparison here is with recent results of inference on regression coefficients in high dimensional regression models. For estimation of a component of the regression vector, it is known that the least squares estimator itself is not sufficiently adaptive against nuisance parameter estimates (estimates of remaining regression vector components) to allow for an optimal rate of convergence. Instead, certain corrections to the least squares loss or its first order moment equations, such as debiasing (Van de Geer et al. [2014]) or orthogonalization (Belloni et al. [2011], Chernozhukov et al. [2015],Belloni et al. [2017] and Ning et al. [2017]) induce sufficient adaptivity against nuisance estimates and thereby allow optimal estimation of the target parameter. The results of this article show that in the context of change point estimation, the plugin least squares estimator (1.5) itself possesses the required adaptivity against potentially high dimensional nuisance estimates, in order to allow for $O(\psi^{-2})$ estimation of the change point $\lfloor T\tau^0\rfloor$, provided the nuisance parameters are estimated with sufficient precision.

We conclude this section with a short note on the notations used in this article. The following section provides a description of the statistical behavior of $\tilde{\tau}$ defined above.

Notation: Throughout the paper, $\mathbb{R}$ represents the real line. For any vector δ, the norms $\|\delta\|_1$, $\|\delta\|_2$, $\|\delta\|_\infty$ represent the usual 1-norm, Euclidean norm, and sup-norm respectively. For any set of indices $U \subseteq \{1, 2, ..., p\}$, let $\delta_U = (\delta_j)_{j\in U}$ represent the subvector of δ containing the components corresponding to the indices in U. Let $|U|$ and U^c represent the cardinality and complement of U. We denote by $a \wedge b = \min\{a, b\}$, and $a \vee b = \max\{a, b\}$, for any $a, b \in \mathbb{R}$. The notation $\lfloor \cdot \rfloor$ is the usual greatest integer function. We use a generic notation $c_u > 0$ to represent universal constants that do not depend on T or any other model parameter. In the following this constant c_u may be different from one term to the next. All limits in this article are with respect to the sample size $T \to \infty$. We use $\Rightarrow$ to represent convergence in distribution.

2 Assumptions and Main Results

In this section we state all sufficient conditions assumed to obtain our main theoretical results regarding the plugin least squares estimator $\tilde{\tau}$ of (1.5). Specifically, an $O(\psi^{-2})$ rate of convergence of $\lfloor T\tilde{\tau} \rfloor$, and its limiting distribution.

Condition A (assumption on model parameters): *Let $S_{1j} = \{k;\, \mu^0_{(j)k} \neq 0\}$, and $S_{2j} = \{k;\, \gamma^0_{(j)k} \neq 0\}$, $1 \leq j \leq p$ be sets of non-zero indices.*
(i) Assume that $\max_{1\leq j\leq p} |S_{1j}| \vee |S_{2j}| = s \geq 1$.
(ii) Assume a change point exists and is sufficiently separated from the boundaries of $(0, 1)$, i.e., for some positive sequence $l_T \to 0$, we have $\big(\lfloor T\tau^0 \rfloor\big) \wedge \big(T - \lfloor T\tau^0 \rfloor\big) \geq Tl_T$.
(iii) Let $\eta^0_{(j)}, \xi_{2,2}$ and ψ be as defined in (1.3). Then assume that for an appropriately chosen small enough constant $c_{u1} > 0$, the following relations hold,

$$c_u\surd(1 + \nu^2)\frac{\sigma^2}{\kappa\psi}\Big\{\frac{s\log^{3/2}(p \vee T)}{\surd(Tl_T)}\Big\} \leq c_{u1}, \quad \text{and} \quad c_u\surd(1 + \nu^2)\frac{\sigma^2}{\psi\kappa}\left\{\frac{s\log(p \vee T)}{T^{\left(\frac{1}{2}-b\right)}\surd l_T}\right\} \leq c_{u1},$$

for some $0 < b < (1/2)$. Here σ^2, ν, κ are model parameters (constants) defined in Condition B.

Condition A provides control on the rate of divergence of s, p, and convergence of ψ, and l_T, which are model parameters that can vary with T. Condition A(iii) can be viewed from two perspectives. First, it allows a vanishing jump size, $\psi \to 0$, when $s\log^{3/2}(p \vee T) = o\big(\surd(Tl_T)\big)$. Alternatively, s, p can be allowed to diverge at an arbitrary rate provided Condition A(iii) is preserved, i.e., provided the jump size is large enough to compensate for the increasing dimensions s, p, so as to preserve Condition A(iii) (also see, Remark 2.2). To the best of our knowledge, this is the weakest condition assumed on the jump size in the dynamic networks literature, where the comparable counterpart of ψ is typically assumed to be diverging. The assumption of sparsity on the coefficient vectors $\mu^0_{(j)}$ and $\gamma^0_{(j)}$ is equivalent to assuming that both the pre and post network structures of Σ^{-1} and Δ^{-1} are such that each node has at most s connecting edges of a total of $(p - 1)$ possible edges. This is a direct extension of the same assumption in the in the static setting, see, e.g. Yuan [2010]. In the case where s, p, l_T are fixed, the rate required of the minimum jump size ψ in Part (iii) can be replaced with $T^{\left(\frac{1}{2}-b\right)}\psi \to \infty$, for some $0 < b < (1/2)$.

Condition B (assumption on the underlying distributions):
(i) The vectors $w_t = (w_{t1}, ..., w_{tp})^T$, $t = 1, .., \lfloor T\tau^0 \rfloor$, and $x_t = (x_{t1}, ..., x_{tp})^T$, $t = \lfloor T\tau^0 \rfloor + 1, ...T$, are independent subgaussian r.v's with mean vector zero, and variance proxy $\sigma^2 \leq c_u$.
(ii) The p-dimensional matrices $\Sigma := Ew_tw_t^T$ and $\Delta := Ex_tx_t^T$ have bounded eigenvalues, i.e., $0 < \kappa \leq \big\{\text{mineigen}(\Sigma) \wedge \text{mineigen}(\Delta)\big\} \leq \big\{\text{maxeigen}(\Sigma) \vee \text{maxeigen}(\Delta)\big\} \leq \phi < \infty$. Consequently, the condition numbers of Σ and Δ are also bounded above by $\nu = \phi/\kappa$.

A subgaussian assumption is a significant relaxation to assuming a Gaussian distribution, for e.g., this condition allows asymmetric distributions such as a centered mixture of two Gaussian distributions. While this assumption is fairly standard in the high dimensional regression literature, these are however much rarer for graphical models, where a Gaussian distribution has often been assumed. Our methodology allows this more general setup since $\tilde{\tau}$ is based on least squares as opposed to a likelihood based approach which are more common in the graphical models setting. More specifically, this condition serves three purposes. Firstly, it allows the residual process in the estimation of τ^0 to converge weakly to the distribution (2.3). Secondly, under a suitable choice of regularization parameters, it allows estimation of nuisance parameters at the rates of convergence presented in (1.4). Finally, in addition to other technical uses, part (ii) of this condition also provides an upper bound on the components of $\mu^0_{(j)}$ and $\gamma^0_{(j)}$, $j = 1, ..., p$, which is necessary to our analysis (Lemma F.7). For the presentation of this section we are agnostic about the choice of the estimator of nuisance parameters and instead require the following condition.

Condition C (assumption nuisance parameter estimates): *Let $\pi_T \to 0$ be a positive sequence. Then with probability $1 - \pi_T$, the following relations are assumed to hold.*
(i) The vectors $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$, $1 \le j \le p$, satisfy the bound (1.4).
(ii) The vectors $(\hat{\mu}_{(j)} - \mu^0_{(j)}) \in \mathcal{A}_{1j}$, $(\hat{\gamma}_{(j)} - \gamma^0_{(j)}) \in \mathcal{A}_{2j}$, for each $1 \le j \le p$. Here $\mathcal{A}_{ij}$, $i = 1, 2$, $j = 1, ..., p$, is a convex subset of $\mathbb{R}^{p-1}$ defined as, $\mathcal{A}_{ij} = \big\{\delta \in \mathbb{R}^{p-1};\ \|\delta_{S^c_{ij}}\|_1 \le 3\|\delta_{S_{ij}}\|_1\big\}$, with S_{ij} being the set of indices defined in Condition A(i) and S^c_{ij} being its complement set.

This condition is a mild requirement of the nuisance estimates. It allows the nuisance estimates $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$ to be irregular in the sense that they are only required to be in a $\{s \log(p \vee T)/T\}^{1/2}$ order neighborhood of the corresponding unknown vectors $\mu^0_{(j)}$ and $\gamma^0_{(j)}$, $j = 1, ..., p$, in the ℓ_2 norm. These nuisance estimates are not required to possess any oracle properties, i.e., selection mistakes in the identification of the sign of coefficient vectors do not influence the eventual change point estimate $\tilde{\tau}$ in its rate of convergence. Accordingly, we do not require assuming irrepresentable conditions on the covariance matrices Σ and Δ, such as those assumed in Kolar and Xing [2012], nor minimum magnitude conditions of the coefficient vectors $\mu^0_{(j)}$, $\gamma^0_{(j)}$, which are assumptions that typically guarantee perfect selection in the components of the vectors $\mu^0_{(j)}$, and $\gamma^0_{(j)}$, $j = 1, ..., p$.

A few more notations are necessary to proceed further. For any $\mu, \gamma \in \mathbb{R}^{p(p-1)}$, and any $\tau \in (0, 1)$ define,

$$\mathcal{U}(z, \tau, \mu, g) = Q(z, \tau, \mu, \gamma) - Q(z, \tau^0, \mu, \gamma),$$

where $\tau^0 \in (0, 1)$ is the unknown change point parameter and $Q(z, \tau, \mu, \gamma)$ is the least squares loss defined in (1.4). Also, for any non-negative sequences $0 \le v_T \le u_T \le 1$, define the collection,

$$\mathcal{G}(u_T, v_T) = \Big\{\tau \in (0, 1);\ Tv_T \le \big|\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor\big| \le Tu_T\Big\} \tag{2.1}$$

We begin with a lemma that provides a uniform lower bound on the expression $\mathcal{U}(z, \tau, \hat{\mu}, \hat{\gamma})$, over the collection $\mathcal{G}(u_T, v_T)$. This lower bound forms the basis of the argument used to obtain the desired rate of convergence for the proposed estimator.

Lemma 2.1. *Suppose Condition A, B and C hold and let $0 \le v_T \le u_T$ be any non-negative sequences. For any $0 < a < 1$, let $c_{a1} = 4 \cdot 48 c_{a2}$, with $c_{a2} \ge \sqrt{(1/a)}$, and*

$$c_{a3} = c_u \Big\{ \frac{c_{a1}(\sigma^2 \vee \phi)\sqrt{(1 + \nu^2)}}{\kappa(1 \wedge \psi)} \Big\}.$$

Additionally, let $u_T \ge c^2_{a1}\sigma^4/(T\phi^2)$, then for $T \ge 2$, we have,

$$\inf_{\substack{\tau \in \mathcal{G}(u_T, v_T);\\ \tau \ge \tau^0}} \mathcal{U}(z, \tau, \hat{\mu}, \hat{\gamma}) \ge \kappa \xi^2_{2,2} \Big[v_T - c_{a3} \max\Big\{ \Big(\frac{u_T}{T}\Big)^{\frac{1}{2}},\ \frac{u_T}{T^b} \Big\} \Big] \tag{2.2}$$

with probability at least $1 - 3a - o(1)$.

Lemma 2.1 is a tool that allows us to obtain the rate of convergence of the change point estimator $\tilde{\tau}$. An observation that provides some insight into this connection and the adaptivity property of the proposed plug in least squares estimator is as follows. Although $\mathcal{U}(z, \tau, \hat{\mu}, \hat{\gamma})$ involves the r.v.'s z_t, which are p-dimensional, and the estimates $\hat{\mu}_{(j)}$, and $\hat{\gamma}_{(j)}$ which approximate $(p-1)$-dimensional unknown parameters $\mu^0_{(j)}$ and $\gamma^0_{(j)}$, $j = 1, ..., p$, up to the rate $O\big(\sqrt{(s \log p/T)}\big)$, yet, the eventual lower bound of Lemma 2.1 is free of the dimensions s, p under the assumed conditions. In a heuristic sense, this alludes to the plugin least squares estimator of the change point behaving as if the nuisance parameters $\mu^0_{(j)}$ and $\gamma^0_{(j)}$ are known. This is indeed the property that allows for the rate of convergence presented in the following theorem to hold. Further insight on the inner working of this result is provided in Remark 2.3 stated after the following result.

Theorem 2.1. *Suppose Conditions A, B and C hold, and for any* $0 < a < 1$, *let* c_{a1}, c_{a2} *and* c_{a3} *be as defined in Lemma 2.1. Then, for* T *sufficiently large, we have the following.*
(i) When $\psi \to 0$ *we have,* $(1+\nu^2)^{-1}(\sigma^2 \vee \phi)^{-2}\kappa^2\psi^2\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \le c_u^2 c_{a1}^2$, *with probability at least* $1 - 3a - o(1)$. *Equivalently, in this case we have,* $\psi^2\big(\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big) = O_p(1)$.
(ii) When $\psi \not\to 0$, *we have,* $\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \le c_{a3}^2$, *with probability at least* $1 - 3a - o(1)$. *Equivalently, in this case we have,* $\big(\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big) = O_p(1)$.

Theorem 2.1 provides the rate of convergence of the proposed estimator $\tilde{\tau}$. The main idea of the proof of this results is to use a contradiction argument as follows. Using Lemma 2.1 recursively, we show that any value of $\lfloor T\tau\rfloor$ lying outside an $O(c_{a3}^2)$ neighborhood of $\lfloor T\tau^0\rfloor$ satisfies, $\mathcal{U}(z, \tau, \hat{\mu}, \hat{\gamma}) > 0$, with probability at least $1 - 3a - o(1)$. Upon noting that by definition of $\tilde{\tau}$, we have, $\mathcal{U}(z, \tilde{\tau}, \hat{\mu}, \hat{\gamma}) \le 0$, yields the desired result. The complete argument is provided in Appendix A. Following are two important remarks regarding this result.

Remark 2.1. Theorem 2.1 provides the rate of convergence for $\lfloor\tilde{\tau}\rfloor$ in the integer time scale. The analogous result in a continuous time scale can be obtained as follows. Note that for any $\tau \ge \tau^0$, we have the deterministic inequality, $T(\tau - \tau^0) - 1 \le \big(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor\big) \le T(\tau - \tau^0) + 1$. In the case where $\psi \to 0$, an application of this inequality together with the Part (i) of Theorem 2.1 leads to $T\psi^2(\tilde{\tau} - \tau^0) = O_p(1)$. In the case where $\psi \not\to 0$, the same inequality when used with Part (ii) of Theorem 2.1 yields $T(\tilde{\tau} - \tau^0) = O_p(1)$.

Remark 2.2. It may be observed that Theorem 2.1 is obtained without any any explicit restriction only on the rate of divergence of the sparsity s and dimension p with respect to the sampling period T. This result is true in itself for s, p diverging at an arbitrary rate with respect to T, as long as the jump size ψ is large enough to compensate in order to preserve Condition A(iii). This is however not the complete picture. Effectively, this result has passed on the burden of an additional assumption controlling the divergence of s, p to Condition C on the nuisance estimates. In order to obtain feasible estimates of the nuisance parameters we shall later require an additional assumption of the form $s \log p = o(Tl_T)$ $\big($see, Condition A$'$(i) and Theorem 3.1 of Section 3$\big)$.

Remark 2.3. Here we provide some partial technical insight as to how Lemma 2.1 and Theorem 2.1 are able to eliminate dimensional parameters s, p and other logarithmic terms of T to obtain the rate of convergence described in its result. The behavior of the estimator $\tilde{\tau}$, is in part controlled by a stochastic noise term of the form,

$$\sup_{\tau;\, \tau \ge \tau^0} \xi_{2,2}^{-1}\Big|\sum_{t=\lfloor T\tau^0\rfloor}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj} z_{t,-j}^T \hat{\eta}_{(j)}\Big|, \quad \text{where } \hat{\eta}_{(j)} = \hat{\mu}_{(j)} - \hat{\gamma}_{(j)},$$

and its mirroring counterpart. Here ε_{tj} is as defined in (2.4). Note here the need for uniformity over τ of this stochastic term, since this forms a critical part of the analysis. A large proportion of the literature upper bounds such uniform stochastic terms using usual subexponential tail bounds (or similar) and supplying uniformity over τ by means of union bounds over the at most T distinct values $\lfloor T\tau \rfloor$. This approach using union bounds forces logarithmic terms of T to necessarily be present in the upper bound for this stochastic term, which passes over to the eventual bound for the change point estimate. Additionally, dimensional parameters s, p also often show up, depending upon how one chooses to control for the nuisance estimates $\hat{\eta}_{(j)}$. Instead of following this approach, we use a novel application of the Kolmogrov's inequality (Theorem F.1) on partial sums in order to control such stochastic terms with sharper upper bounds. This is done by first using a triangle inequality,

$$\begin{aligned}\sup_{\tau;\,\tau\geq\tau^0} \xi_{2,2}^{-1}\Big|\sum_{t=\lfloor T\tau^0\rfloor}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj} z_{t,-j}^T\hat{\eta}_{(j)}\Big| \quad &\leq \quad \sup_{\tau;\,\tau\geq\tau^0} \xi_{2,2}^{-1}\Big|\sum_{t=\lfloor T\tau^0\rfloor}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj} z_{t,-j}^T\eta^0_{(j)}\Big| \\ &\quad + \sup_{\tau;\,\tau\geq\tau^0} \xi_{2,2}^{-1}\Big|\sum_{t=\lfloor T\tau^0\rfloor}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj} z_{t,-j}^T(\hat{\eta}_{(j)}-\eta^0_{(j)})\Big|.\end{aligned}$$

The first term on the rhs can now be controlled at an optimal rate $O(\sqrt{T})$, (see, Lemma C.2 and Lemma C.4) without any additional logarithmic terms of T, using the Kolmogorov's inequality. Moreover, under Condition A and C, the second term on the rhs of the above inequality can also controlled with the same upper bound, despite high dimensionality and without dimensional parameters s, p, being involved in the upper bound (see, Lemma C.3, Lemma C.6 and proof of Lemma 2.1). This provides the desired sharper control on stochastic noise terms and consequently allows for the rate of convergence presented in Theorem 2.1. This is ofcourse a simplified explanation and only meant for intuition purposes.

The availability of an $O(\psi^{-2})$ rate of the proposed change point estimator $\lfloor T\tilde{\tau}\rfloor$ from Theorem 2.1 allows the existence of a limiting distribution and thus we now shift our focus to performing inference on the change point τ^0. For this purpose, let $W_1(r)$, and $W_2(r)$ be two independent Brownian motions defined on $[0,\infty)$. For any constants $0<\sigma_1,\sigma_2,\sigma_1^*,\sigma_2^*<\infty$, define,

$$Z(r)=\begin{cases}\frac{\sigma_2^2}{\sigma_1^2}|r|-\frac{2\sigma_2^*}{\sigma_1^*}W_1(r) & \text{if } r>0,\\ 0 & \text{if } r=0,\\ |r|-2W_2(r) & \text{if } r<0.\end{cases} \tag{2.3}$$

This process is presented in Bai [1997], where it is defined in a linear regression context. It is an asymmetric variant of the well known process $\{|r|-2W(r)\}$, which often arises in the context of change point estimators, see, e.g., Yao [1987], Bai [1994], among others. In the following for any $t=1,...,T$ we define,

$$\varepsilon_{tj}=\begin{cases} z_{tj}-z_{t,-j}^T\mu^0_{(j)}, & t=1,...,\lfloor T\tau^0\rfloor\\ z_{tj}-z_{t,-j}^T\gamma^0_{(j)}, & t=\lfloor T\tau^0\rfloor+1,...,T.\end{cases} \tag{2.4}$$

We shall also require the following additional conditions to proceed further.

Condition D: *(i) Assume that the jump size is vanishing, i.e.,* $\psi\to 0$.
(ii) The covariance matrices Σ *and* Δ *satisfy,*

$$\xi_{2,2}^{-2}\sum_{j=1}^{p}\eta^{0T}_{(j)}\Sigma_{-j,-j}\eta^0_{(j)}\to\sigma_1^2, \quad \text{and} \quad \xi_{2,2}^{-2}\sum_{j=1}^{p}\eta^{0T}_{(j)}\Delta_{-j,-j}\eta^0_{(j)}\to\sigma_2^2. \tag{2.5}$$

The additional assumption (2.5) is made to ensure convergence of the variance of the proposed estimator $\tilde{\tau}$, which in turn ensures the stability of the limiting distribution. This is only a mild additional requirement, since the assumed convergence is on a positive sequence that is already guaranteed to be bounded, i.e.,

$$\kappa\xi_{2,2}^2 \leq \sum_{j=1}^{p} \eta_{(j)}^{0T}\Sigma_{-j,-j}\eta_{(j)}^0 \leq \phi\xi_{2,2}^2,$$

where the inequalities from the bounded eigenvalues of the covariance matrix Σ $\big($Condition B(ii)$\big)$, and similar for the post-change covariance matrix Δ.

Condition E: *(i) For ε_{tj}, for $t = 1, ..., T$, and $j = 1, ..., p$, as defined in (2.4), assume that,*

$$\xi_{2,2}^{-2}\mathrm{var}\Big(\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0\Big) \to \sigma_1^{*2}, \qquad \text{for } t = 1, ..., \lfloor T\tau^0\rfloor \text{ and,}$$
$$\xi_{2,2}^{-2}\mathrm{var}\Big(\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0\Big) \to \sigma_2^{*2}, \qquad \text{for } t = \lfloor T\tau^0\rfloor + 1, ..., T.$$

(ii) A functional central limit theorem holds for double array $\zeta_{tj} = \varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0$, i.e., for $r > 0$,

$$p^{-1/2}\psi\xi_{2,2}^{-1}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^0+r\psi^{-2}\rfloor}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 \Rightarrow \sigma_2^*W_1(r), \quad \text{when } \psi\to 0, \text{ and } p\to\infty.$$

Also assume a mirroring statement for $r < 0$.

The convergence in Condition E(i) is closely related to Condition D(ii). The need to assume this condition despite the availability of Condition D(ii) is due to the block dependence in the double array $\{\zeta_{tj}\}$. Although for any $t \neq t'$, we have the independence of ε_{tj}, and ε_{tk}. However, within a fixed block t, the variables ζ_{tj}, and ζ_{tk} may be correlated. It is due to the same underlying dependence that Condition E(ii) needs to be assumed. If all ζ_{tj}'s are pairwise independent, the Condition E becomes redundant. In this case, the first requirement follows from Condition D(ii) and the second requirement follows directly from the classical functional central limit theorem. In the case where p is fixed, Condition E(ii) again becomes redundant, since we have independence over the indices t and thus the assumption follows from the functional central limit theorem.

Theorem 2.2. *Suppose Conditions A, B, C, D and E hold. Additionally assume that*

$$\frac{1}{\psi}\Big\{\frac{s\log^{3/2}(p\vee T)}{\surd(Tl_T)}\Big\} = o(1). \tag{2.6}$$

Then the estimator $\tilde{\tau}$ of (1.5) obeys the following limiting distribution.

$$T(\sigma_1^*)^{-2}\sigma_1^4\psi^2(\tilde{\tau}-\tau) \Rightarrow \arg\min_r Z(r).$$

where $Z(r)$ is as define in (2.3).

The result of Theorem 2.2 establishes the second main result of this article. It provides the limiting distribution of $\tilde{\tau}$, whose density function is readily available in Bai [1997]. Thereby allowing straightforward computation of quantiles of this distribution. The only difference between the assumption (2.6) and the second rate restriction in Part (iii) of Condition A is that the rhs has been tightened to $o(1)$ from $O(1)$. This slightly stronger requirement for the existence of the limiting distribution is in coherence with classical results such as those in Bai [1994] and Bai [1997]. Feasible computation of a confidence interval requires the parameters $\sigma_1, \sigma_1^*, \sigma_2$, and σ_2^*, these computations can be carried on binary partition of data induced by $\tilde{\tau}$, the details of these estimations are provided in Section 4

Remark 2.4. The results of Theorem2.1 and Theorem 2.2 above can be viewed as stating that $\tilde{\tau}$ utilizing $2p$ estimated vectors $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$, $j = 1, ..., p$, each of dimension $p - 1$, is still behaving as if these nuisance parameters are known. This is despite allowing high dimenisonality in keeping with Condition A(iii), and a potentially diminishing jump size ψ. This is effectively the adaptation property as described in Bickel [1982], in a high dimensional setting withing a change point parameter context.

While the results of this subsection allow $\lfloor T\tilde{\tau} \rfloor$ to provide an $O(\psi^{-2})$ approximation of $\lfloor T\tau^0 \rfloor$, and in turn allows a limiting distribution to perform inference on the unknown change point. However, all results rely on the apriori availability of nuisance estimates $\hat{\mu}_{(j)}$, and $\hat{\gamma}_{(j)}$, $j = 1, ..., p$, satisfying Condition C. Without this availability, these results remain infeasible to implement in practice. In the following section we develop an algorithmic estimator to obtain these nuisance estimates theoretically guarantee Condition C for the same. Consequently, making the methodology of this section viable in practice.

3 Construction of a feasible $O(\psi^{-2})$ estimator of $\lfloor T\tau^0 \rfloor$

To discuss the methods and results of this section we require more notation. For any $\tau \in (0, 1)$, such that $\lfloor T\tau \rfloor \geq 1$, consider ordinary lasso estimates of the regression of each column of the observed variable z on the rest, for each of the two binary partitions induced by τ. Specifically, for each $j = 1, ..., p$, define

$$\begin{aligned}
\hat{\mu}_{(j)}(\tau) &= \underset{\mu_{(j)} \in \mathbb{R}^{p-1}}{\arg\min} \Big\{ \frac{1}{\lfloor T\tau \rfloor} \sum_{t=1}^{\lfloor T\tau \rfloor} \big(z_{tj} - z_{t,-j}^T \mu_{(j)} \big)^2 + \lambda_j \|\mu_{(j)}\|_1 \Big\}, \quad \text{and} \qquad (3.1)\\
\hat{\gamma}_{(j)}(\tau) &= \underset{\gamma_{(j)} \in \mathbb{R}^{p-1}}{\arg\min} \Big\{ \frac{1}{(T - \lfloor T\tau \rfloor)} \sum_{t=\lfloor T\tau \rfloor + 1}^{T} \big(z_{tj} - z_{t,-j}^T \gamma_{(j)} \big)^2 + \lambda_j \|\gamma_{(j)}\|_1 \Big\}, \quad \lambda_j > 0.
\end{aligned}$$

To develop a feasible estimator for the τ^0, recall two aspects from Section 2. (a) The missing links required to implement the estimator $\tilde{\tau}(\hat{\mu}, \hat{\gamma})$ of Section 2 are the edge parameter vector estimates $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$, $j = 1, ..., p$. (b) These edge estimates require the sufficient Condition C to be satisfied in order to retain the results of Section 2. We shall fulfill these nuisance estimate requirements using the estimators in (3.1), implemented in a twice iterated manner. Here the iterations are between the change point parameter τ and the edge parameters $\mu_{(j)}$ and $\gamma_{(j)}$.

The twice iterative approach of the estimator to be considered is as follows. Very rough edge estimates $\check{\mu}_{(j)} = \hat{\mu}_{(j)}(\check{\tau})$, and $\check{\gamma}_{(j)} = \hat{\gamma}_{(j)}(\check{\tau})$, $j = 1, ..., p$, computed using a nearly arbitrarily chosen $\check{\tau} \in (0, 1)$ (see, initializing condition of Algorithm 1 below) possesses sufficient information so that an update of the change point parameter, i.e., $\hat{\tau} = \tilde{\tau}(\check{\mu}, \check{\gamma})$, moves the nearly arbitrary starting point $\check{\tau}$ into a near optimal neighborhood, $O_p\big(\psi^{-2} T^{-1} \log(p \vee T)\big)$, with this single step. With this availability of a near optimal estimate $\hat{\tau}$, we shall show that another update $\hat{\mu}_{(j)} = \hat{\mu}_{(j)}(\hat{\tau})$, and $\hat{\gamma}_{(j)} = \hat{\gamma}_{(j)}(\hat{\tau})$ satisfies all requirements of Condition C. This provides sufficient ingredients to now implement the

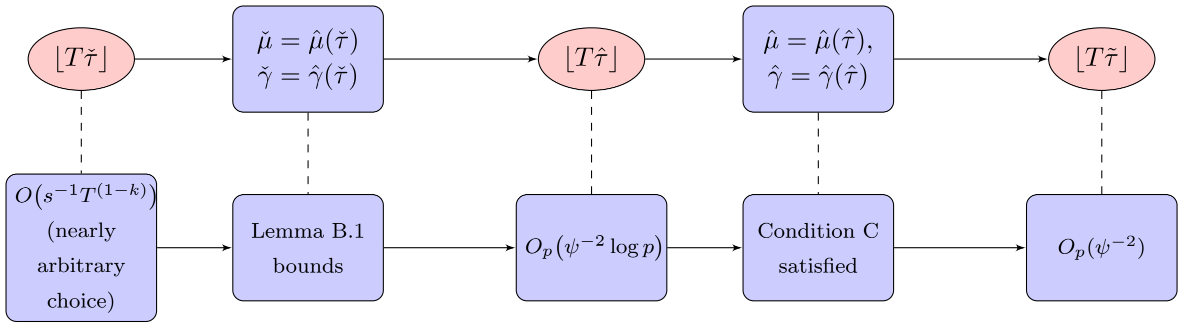

Figure 1: A schematic of the proposed Algorithm 1 and its underlying working mechanism.

estimator of Section 2, i.e., performing a second update of the change point $\tilde{\tau}=\tilde{\tau}(\hat{\mu},\hat{\gamma})$, which moves $\hat{\tau}$ from the near optimal neighborhood $O_p\big(\psi^{-2}T^{-1}\log(p\vee T)\big)$ into an $O_p\big(\psi^{-2}T^{-1}\big)$ neighborhood of τ^0. This is a direct consequence of Theorem 2.1. Additionally, Theorem 2.2 also provides the limiting distribution of this second update $\tilde{\tau}$, thereby allowing inference on τ^0. Thus, in performing these updates (two each of the change point and the mean) we have taken a $\check{\tau}$ from a nearly arbitrary neighborhood of τ^0, and deposited it in an $O_p\big(\psi^{-2}T^{-1}\big)$ neighborhood of τ^0, with an intermediate $\hat{\tau}$ that lies in a near optimal neighborhood. This process is stated as Algorithm 1 below and is described visually in Figure 1. Theorem 3.1 and the subsequent corollaries provide the precise description of the statistical performance of the proposed Algorithm 1 and the required sufficient conditions.

Algorithm 1: $O(\psi^{-2})$ estimation of $\lfloor T\tau^0\rfloor$:

(Initialize): Choose any $\check{\tau}\in(0,1)$ satisfying Condition F.

Step 1: Obtain $\check{\mu}_{(j)}=\hat{\mu}_{(j)}(\check{\tau})$, and $\check{\gamma}_{(j)}=\hat{\gamma}(\check{\tau})$, $j=1,...,p$, and update change point as,

$$\hat{\tau}=\arg\min_{\tau\in(0,1)} Q(z,\tau,\check{\mu},\check{\gamma})$$

Step 2: Obtain $\hat{\mu}_{(j)}=\hat{\mu}_{(j)}(\hat{\tau})$, and $\hat{\gamma}_{(j)}=\hat{\gamma}(\hat{\tau})$, $j=1,...,p$, and perform another update,

$$\tilde{\tau}=\arg\min_{\tau\in(0,1)} Q(z,\tau,\hat{\mu},\hat{\gamma})$$

(Output): $\tilde{\tau}$

To complete the description of Algorithm 1, we provide the weak sufficient condition required from the initializing choice $\check{\tau}$.

Condition F (initializing assumption): *Assume that the initializer $\check{\tau}$ of Algorithm 1 satisfies the following relations.*

$$(i)\ \lfloor T\check{\tau}\rfloor\wedge(T-\lfloor T\check{\tau}\rfloor)\geq c_uTl_T,\quad\text{and}\quad (ii)|\lfloor T\check{\tau}\rfloor-\lfloor T\tau^0\rfloor|\leq\frac{c_u\kappa l_T}{s(\sigma^2\vee\phi)}T^{(1-k)},$$

for any constant $k>0$. [d]

[d] Without loss of generality we assume $k<b$, where b is as defined in Condition A′.

The first requirement of Condition F is clearly innocuous, all it requires is a marginal separation of the chosen initializer from the boundaries of the parametric space of the change point. It is satisfied with any $\check{\tau} \in [c_{u1}, c_{u2}] \subset (0,1)$. While at face value the second may seem stringent, however this is not the case and this condition is satisfied by nearly any and all arbitrarily chosen $\check{\tau} \in (0,1)$, when T is large. To see this, first consider a simplistic case when $l_T \geq c_u > 0$, i.e., true change point τ^0 is in some bounded subset of $(0,1)$, and $s \leq c_u$, i.e., the sparsity parameter is bounded above by a constant. Then, requirement (ii) of Condition F, reduces to $|\lfloor T\check{\tau}\rfloor - \lfloor T\tau^0\rfloor| = O(T^{1-k})$. Recall here that constant k in Condition F can be any arbitrary value (not depending on T) close to zero. Thus the neighborhood $O(T^{(1-k)})$, of τ^0 can cover a larger and larger proportion of the entire parametric space $(0,1)$ as T increases. This can be further illustrated by noting that the disallowed case of $k = 0$ covers the entire parameteric space of τ^0. Here we also refer to Kaul et al. [2019] where a similar initializer condition has been discussed in detail. This makes it a very weak requirement and when T is large, it will be satisfied by any nearly any arbitrarily chosen value. More generally, the powerfulness of Algorithm 1 is that it starts with any $\lfloor T\check{\tau}\rfloor$ in a very wide neighborhood of $\lfloor T\tau^0\rfloor$, $\hat{\tau}$ of Step 1 then moves it into a near optimal neighborhood, and finally $\tilde{\tau}$ of Step 2 moves it into an sharper neighborhood, i.e., $O(s^{-1}T^{(1-k)})$-nbd.$\longrightarrow^{\text{Step1}}$ near optimal-nbd., $O_p(\psi^{-2}\log p)$ $\longrightarrow^{\text{Step2}}$ $O_p(\psi^{-2})$.

While the working mechanism of Algorithm 1 in its ability to provide an $O(\psi^{-2})$ rate of convergence from a nearly arbitrary neighborhood in two iterations is clear, in the following we provide further arguments regarding the initializer in case the reader remains unconvinced on the choice of the initializer from a practical perspective. In order to find a suitable initializer in an $O(s^{-1}T^(1-k))$ neighborhood, at a given sampling period T, one may consider using a preliminary equally spaced coarse grid of values in $(0,1)$, and choosing a best fitting value to the data. If the dimensionality of this preliminary grid is larger than $O(sT^k)$, then one will arrive at a theoretically valid initializer. A similar preliminary coarse grid search has also been heuristically utilized in Roy et al. [2017] in a different model setting. However, based on extensive numerical experiments, we have observed that even this preliminary coarse grid search is redundant. In Section 4 we present results with this initializer fixed at $\check{\tau} = 0.5$, irrespective of the underlying true change point τ^0, which is allowed to vary all across the parametric space $(0,1)$. Note here that in the absence of any information on τ^0, the choice $\check{\tau} = 0.5$ forms the worst or farthest initializer in a mean sense, and all other value of $\check{\tau}$ shall only serve to make the estimation easier. Despite this worst possible choice, numerical results remain indistinguishable compared to those obtained when $\check{\tau}$ is chosen with a preliminary coarse grid search. The reader may numerically confirm these observations using the simple software associated with this article. This observation while at first may be counter intuitive, however it is exactly that described in the previous paragraph, i.e., Condition F is weak enough to be satisfied with nearly any arbitrarily chosen $\check{\tau}$.

The results to follow provide a precise description of the above discussion. We begin with an additional condition that is largely a weaker version of Condition A, which is sufficient to obtain near optimality of $\hat{\tau}$ of Step 1 of Algorithm 1. Note that this near optimal rate of convergence to follow, matches the best available in the current literature in the assumed setting. However, we shall obtain this rate under a much weaker assumption on the jump size ψ, which is infact even weaker than that assumed on ψ earlier in Condition A of Section 2.

Condition A$'$(assumption on model parameters): *Let $\xi_{2,2}$, ψ, be as defined in (1.3), s, and l_T as defined in Condition A and model constants σ^2, ν, and κ as defined in Condition B.*
(i) Assume that $c_u\kappa T^{(1-b)} l_T \geq (\sigma^2 \vee \phi)s\log(p \vee T)$, where b is as given in A$'$(iii).
(ii) Assume τ^0 is separated from the parametric boundary, i.e., $(\lfloor T\tau^0\rfloor) \wedge (T - \lfloor T\tau^0\rfloor) \geq Tl_T$.

(iii) Assume that for an appropriately chosen small enough constant $c_{u1} > 0$, the following holds,

$$c_u\sqrt{(1+\nu^2)}\frac{\sigma^2 s}{\psi\kappa}\Big\{\frac{\log(p\vee T)}{T^{(1-2b)}l_T}\Big\}^{\frac{1}{2}} \le c_{u1},$$

for some $0 < b < (1/2)$.

The following theorem shows that $\lfloor T\hat{\tau}\rfloor$ of Step 1 of Algorithm 1 lies in an $O\big(\psi^{-2}\log(p\vee T)\big)$ neighborhood of τ^0. The validity of $\tilde{\tau}$ of Step 2 in terms of properties provided in Section 2 shall rely critically on this result.

Theorem 3.1. *Suppose Conditions A$'$, B and F hold. Let $\hat{\tau}$ be the change point estimate of Step 1 of Algorithm 1. Then for T sufficiently large, we have,*

$$(1\wedge\psi^2)(1+\nu^2)^{-1}(\sigma^2\vee\phi)^{-2}\kappa^2\big|\lfloor T\hat{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \le c_u\log(p\vee T) \qquad (3.2)$$

with probability $1-o(1)$. In other words, $(1\wedge\psi^2)\big(\lfloor T\hat{\tau}\rfloor - \lfloor T\tau^0\rfloor\big) = O\big(\log(p\vee T)\big)$, with probability converging to one.

Theorem 3.1 shows that $\lfloor T\hat{\tau}\rfloor$ of Step 1 of Algorithm 1 will satisfy a near optimal bound $O_p\big(\psi^{-2}\log(p\vee T)\big)$, despite the algorithm initializing with any $\lfloor T\check{\tau}\rfloor$ in a $O(s^{-1}T^{(1-k)})$ neighborhood of τ^0. This result now allows us to study the behavior of estimates of the edge parameters and the change point parameter obtained from the second iteration of Step 2 of Algorithm 1. We note here that the properties of these second iteration estimates rely solely on the bound (3.2) of $\lfloor T\hat{\tau}\rfloor$, and the availability of this bound renders no further use of the initial edge estimates $\check{\mu}_{(j)}$ and $\check{\gamma}_{(j)}$, $j=1,...,p$. This feature allows Algorithm 1 to be modular in its construction, in the sense that for Step 2 to yield an estimate $\lfloor T\tilde{\tau}\rfloor$ that is $O(\psi^{-2})$ in its rate of convergence, it does not require the estimator of Step 1 to be specifically the one that has currently been chosen in Algorithm 1. Alternatively, Step 1 of Algorithm 1 can readily be replaced with any other near optimal estimator available in the literature, i.e., satisfying a bound $O\big(\psi^{-2}\log(p\vee T)\big)$ with probability $1-o(1)$. This is described below as Algorithm 2.

Algorithm 2: $O(\psi^{-2})$ estimation of $\lfloor T\tau^0\rfloor$:

Step 1: Implement any $\hat{\tau}$ from the literature that satisfies (3.2) with probability $1-o(1)$.
Step 2: Obtain $\hat{\mu}_{(j)} = \hat{\mu}_{(j)}(\hat{\tau})$, and $\hat{\gamma}_{(j)} = \hat{\gamma}(\hat{\tau})$, $j=1,...,p$, and perform update,

$$\tilde{\tau} = \arg\min_{\tau\in(0,1)} Q(z,\tau,\hat{\mu},\hat{\gamma})$$

(Output): $\tilde{\tau}$

An estimator from the literature that can be used in Step 1 of Algorithm 2 is of Atchade and Bybee [2017], which obeys the near optimal bound of Theorem 3.1. However, their method being based on a likelihood approach would limit the algorithm to the Gaussian setting, moreover it requires stronger sufficient conditions of the minimum jump size and separation sequence l_T for analytical validity. To the best of our knowledge, there is no available estimator in the current literature that would serve as a replacement for Step 1 of Algorithm 1 under the assumptions of Condition A$'$ (or Condition A) and Condition B.

The following results present the statistical behavior of $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$, $j=1,...,p$ and $\tilde{\tau}$ obtained from Step 2 of Algorithm 1 or Algorithm 2. These results show that edge parameter updates $\hat{\mu}_{(j)}$ and

$\hat{\gamma}_{(j)}$, $j = 1, ..., p$ obtained using the near optimal $\hat{\tau}$, are of a much tighter precision than those in Step 1. In particular, these satisfy all requirements of Condition C. This in turn allows the applicability of the results of Section 2 obtaining a higher precision for $\tilde{\tau}$ in comparison to that of the change point estimate of Step 1 of Algorithm 1 or Algorithm 2.

Corollary 3.1. *Suppose Conditions A′, B and F hold. Let $\hat{\mu}_{(j)}$, and $\hat{\gamma}_{(j)}$, $j = 1, ..., p$, be the edge estimate of Step 2 of Algorithm 1 or Algorithm 2. Then the following two properties hold with probability at least $1 - o(1)$.*
(i) We have $\hat{\mu}_{(j)} - \mu^0_{(j)} \in \mathcal{A}_{1j}$, and $\hat{\gamma}_{(j)} - \gamma^0_{(j)} \in \mathcal{A}_{2j}$, $j = 1, ..., p$ where $\mathcal{A}_{ij}$ are sets as defined in Condition C.
(ii) The following bound is satisfied,

$$\max_{1 \le j \le p} \Big(\|\hat{\mu}_{(j)} - \mu^0_{(j)}\|_2 \vee \|\hat{\gamma}_{(j)} - \gamma^0_{(j)}\|_2 \Big) \le c_u \surd(1 + \nu^2) \frac{\sigma^2}{\kappa} \Big\{ \frac{s \log(p \vee T)}{T l_T} \Big\}^{\frac{1}{2}}.$$

Consequently, these second iteration edge estimates satisfy all requirements of Condition C.

Corollary 3.1 provides the feasibility of Condition C. We are now in a position to appeal to the results of Section 2 in order to obtain the statistical performance of $\tilde{\tau}$ of Step 2 of Algorithm 1 or Algorithm 2. The following corollary is a direct consequence of Theorem 2.1 and Theorem 2.2.

Corollary 3.2. *Suppose Conditions A, B and F hold and additionally assume that model dimensions are sufficiently restricted to satisfy $c_u \kappa T^{(1-b)} l_T \ge (\sigma^2 \vee \phi) s \log(p \vee T)$. Then $\tilde{\tau}$ of Algorithm 1 or Algorithm 2 satisfies the error bounds of Theorem 2.1. Additionally assuming Condition D, E and (2.6) holds, then $\tilde{\tau}$ obeys the limiting distribution of Theorem 2.2.*

Corollary 3.2 completes the description of the behavior of the proposed Algorithm 1 and Algorithm 2, which are both feasible estimators implementable in practice. This result allows for $O(\psi^{-2})$ estimation of $\lfloor T\tau^0 \rfloor$, and the ability to perform inference. It may be of interest to note that Corollary 3.2 assumes tighter restrictions on the model parameters ψ, s, p (see, Condition A(iii)) in comparison to those required for near optimality described in Theorem 3.1 (see, Condition A′(iii)). This is the only price we have paid to go from $O\big(\psi^{-2} \log(p \vee T)\big)$ to an $O(\psi^{-2})$ neighborhood of the unknown change point $\lfloor T\tau^0 \rfloor$. Similar additional sufficient restrictions to allow for an optimal rate of convergence and in turn for inference in a high dimensional setting have a precedent in the recent literature in context of inference on a regression coefficient parameter in the presence of high dimensionality, see, e.g. The debiased lasso Van de Geer et al. [2014], orthogonalized score estimator Belloni et al. [2011], Chernozhukov et al. [2015], Belloni et al. [2017] and Ning et al. [2017]. In this context, it has been shown that while the condition $s \log p = o(T)$ is sufficient for near optimal estimation, a tighter constraint $s \log p = o(\surd T)$ is sufficient for an optimal rate and in turn inference. The distinction between Condition A and Condition A′ of Section 2 and Section 3 can be viewed from a similar lens.

4 Numerical Results

This section provides an empirical validation of the proposed $\hat{\tau}$ and $\tilde{\tau}$ of Step 1 and Step 2 of Algorithm 1, respectively. We begin with the description of the simulation design. In all cases considered, the unobserved variables w_t, x_T on model (1.1) are generated as independent, p-dimensional, mean zero Gaussian r.v.'s with distinct covariance structures. More precisely, we set $w_t \sim \mathcal{N}(0, \Sigma)$, $t = 1, ..., \lfloor T\tau^0 \rfloor$ and $x_t \sim \mathcal{N}(0, \Delta)$, $t = \lfloor T\tau^0 \rfloor + 1, ..., T$. The covariance's Σ, and Δ are chosen so as

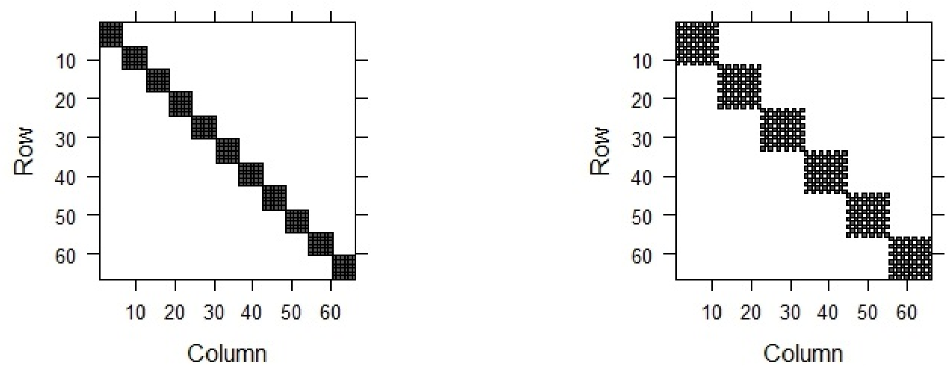

Figure 2: Adjacency matrices $|\text{sign}(\Sigma)|$ and $|\text{sign}(\Delta)|$, with $p = 66$. Here $|\cdot|$ is componentwise absolute value. These matrices represent the underlying graphical model. Each dark pixel is an edge between the corresponding nodes, i.e., a non-zero entry of the precision matrix.

to preserve the sparsity of each of these matrices, and the jump size between these two matrices, as the dimension p grows larger. The covariances Σ and Δ are constructed as follows. We begin with a toeplitz type covariance matrix Γ where the $(l, m)^{th}$ component of this matrix is chosen as $\Gamma_{(l,m)} = \rho^{|l-m|^a}$, $l, m = 1, ..., p$. We set $\rho = 0.5$ and $a = 1/4$.[e]. Then we set $\Sigma = c_1 \cdot A_1 \cdot \Gamma$ and $\Delta = c_2 \cdot A_2 \cdot \Gamma$, where $\cdot$ represents a componentwise product. Here $c_1 = 2$ and $c_1 = 1$ are constants that allow the data generating process to have differing variances pre and post the change point. Recall from the discussion in Section 1, the constants c_1 and c_2 have no impact on the jump size ψ between the matrices Σ and Δ. The matrices A_1 and A_2 are chosen as $p \times p$ matrices of signs $\{-1, 0, 1\}$. These serve to induce sparsity and a varying edge structure between Σ and Δ as follows. A_1 and A_2 are constructed as block diagonal matrices of ones consisting of $(s+1) \times (s+1)$ blocks, and $(2s+1) \times (2s+1)$ blocks, respectively. Furthermore, alternating components of A_1 and A_2 are switched to a negative sign and zero, respectively. This yields distinct positive definite matrices Σ and Δ with differing variance and edge structures. Moreover, by this construction the sparsity of each row and column of Σ and Δ is fixed at $s = 5$, i.e., $|S_{ij}| = 5$, $i = 1, 2$ and $j = 1, ..., p$ where S_{ij} are sets of indices defined in Condition A(i). The normalized jump size $\psi \approx 0.46$,[f] remains fixed when the p changes. Examples of the adjacency matrices corresponding to Σ and Δ obtained from this construction are illustrated in Figure 2.

We perform simulations for all combinations of parameters described in the following. The sampling period T in chosen in the grid $\{200, 275, 350, 425\}$, and the dimension p in $\{100, 200, 300, 400\}$. The change point τ^0 is chosen from an equally spaced grid of values $\{0.15, ..., 0.85\}_{10\times 1}$. Note that this grid contians values from nearly the entire parameteric space $(0, 1)$ of the change point τ^0. All computations are carried out in the software *R*, [R Core Team [2017]], lasso optimizations of the form (3.1) are carried out via the r-package *glmnet*, [Friedman et al. [2010]]. In all simulations the initializer for Algorithm 1 is chosen as $\check{\tau} = 0.5$, which assumes no prior knowledge of the underlying

[e]We choose the fourth root of $|l - m|$ so as to somewhat preserve the magnitude of correlations

[f]The matrices Σ and Δ are constructed so as to preserve the jump ψ, irrespective of the dimension p. However slight numerical fluctuations of order $\approx 10^{-3}$ are observed, potentially due the numerical inversion of large matrices undertaken to calculate ψ.

change point parameter.

The regularizers λ_j, $j = 1, ..., p$ for the implementation of lasso optimizations of Step 1 and Step 2 of Algorithm 1 are chosen via a five fold cross validation from a set grid of values, carried out internally in *glmnet*. This grid is chosen as $\Big\{c\{\log(p \vee T)/T\}^{1/2}, ...2\Big\}$, with $c = 1.75$, with fifty equally spaced grid points. The separation of this grid from zero is done to avoid overfitting and is in coherence with the choice of regularizer prescribed in Theorem B.1.

For the purposes of calculation of the standard error and quantiles, which in turn require $\sigma_1, \sigma_2, \sigma_1^*$, and σ_2^* defined in Codition D and Condition E of Section 2, we assume the covariances Σ and Δ to be known. In practice these can be replaced with estimates of Σ and Δ reconstructed from the edge estimates $\hat{\mu}_{(j)}$, and $\hat{\gamma}_{(j)}$ of Step 2 using for e.g. neighborhood selection (Meinshausen et al. [2006], Yuan [2010]) on each binary partition yielded by $\tilde{\tau}$. The parameters σ_1, σ_2 are obtained as their respective finite sample approximations from their defining expressions in Condition D. A further approximation is made in order to calculate σ_1^*, and σ_2^*. These are obtained as,

$$\begin{aligned} &\sigma_1^{*2} \approx \xi_{2,2}^{-2} \sum_{j=1}^{p} \sigma_{(j)}^2 \eta_{(j)}^{0T} \Sigma \eta_{(j)}^0, \quad \text{where,} \\ &\sigma_{(j)}^2 = \mathrm{var}(\varepsilon_{tj}) = \Sigma_{j,j} - \mu_{(j)}^{0T} \Sigma_{-j,j},\ t \le \lfloor T\tau^0 \rfloor,\ j = 1, ..., p, \end{aligned} \tag{4.1}$$

and similar for the calculation of σ_2^{*2}, where Σ replaced with Δ, and $\mu_{(j)}^0$ replaced with $\gamma_{(j)}^0$, $j = 1, ..., p$. This calculation ignores the blockwise dependence structure between $\varepsilon_{tj} z_{t,-j}$ and $\varepsilon_{tk} z_{t,-k}$, $j \ne k$, i.e., fourth order interactions between the components of z are not taken into account for this approximation.

To report our results we provide the following metrics which are approximated based on 200 monte carlo replications: $\big($bias, $|E(\hat{\tau} - \tau^0)|\big)$, and root mean squared error $\big($rmse, $E^{\frac{1}{2}}(\hat{\tau} - \tau^0)^2\big)$ and corresponding metrics for $\tilde{\tau}$. We also report the standard error and quantile associated with the estimator $T\tilde{\tau}$, and the limiting distribution of Theorem 2.2, respectively. In particular, the former is given as $\mathrm{SE}(T\tilde{\tau}) = \sigma_1^{*2}/(\psi^2 \sigma_1^4)$. The latter, c_α is obtained as a symmetric quantile at a $(1-\alpha) = 0.95$ confidence level. This computation in turn requires the ratio's σ_2^2/σ_1^2 and $\sigma_2^{*2}/\sigma_1^{*2}$, and the cumulative distribution function of the limiting distribution, which is provided in Bai [1997]. Since the standard error and quantiles are computed based on parameters which are assumed to be known, these are not ifluenced by the sampling period T, the change point parameter τ^0, or the monte carlo replications. Furthermore, by the construction of Σ and Δ, these quantites stay roughly the same accross the dimension p.

Partial results of the numerical experiments are provided in Table 1, Table 2, and Figure 3. Most notable observation here is the uniform improvement in bias and rmse provided by the Step 2 estimate $\tilde{\tau}$ over the Step 1 estimate $\hat{\tau}$. Moreover, this improvement is more pronounced when the sampling period is larger. This indicative of the sharper rate of convergence of $\tilde{\tau}$ (Theorem 2.1) over the near optimal rate of $\hat{\tau}$ (Theorem 3.1). A clear consistency trend of improved estimation with a large sampling period with an expected deterioration of estimation precision with increased dimenisonality p is observed. The numerical results of all other cases of τ^0 mimic these trends. The results for $\tau^0 = 0.69$ and $\tau^0 = 0.77$ are provided in Table 3 and Table 4 in Appendix G of the supplementary materials, the remainder are ommitted. Coverage of confidence intervals constructed over replications using the corresponding standard error and quantile was also evaluated. This was found to be conservative and near one in all case with $T \ge 275$, and $0.21 \le \tau^0 \le 0.77$. This is likely due to the nature of result of Theorem 2.2, which is obtained under the regime $\psi \to 0$. Thus under any finite ψ setup, these intervals are expected to be conservative, when T is large. This conservative behavior of confidence intervals has also been mentioned in the seminal work of Bai [1994], where

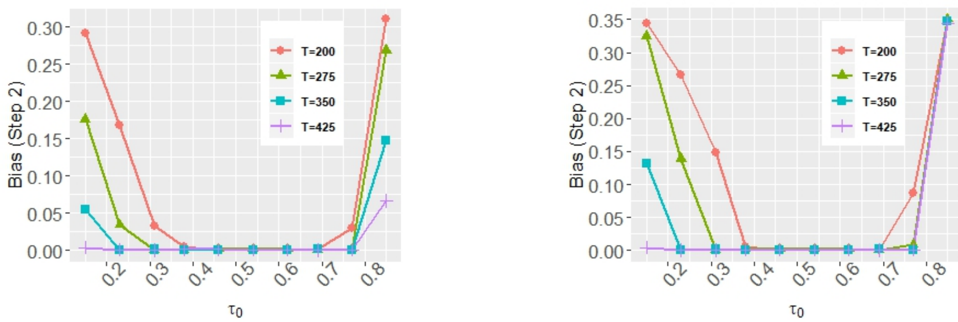

Figure 3: Iluustration of numerical results of bias of $\tilde{\tau}$ of Step 2 of Algorithm 1 over several values of the change point parameter τ^0. **Left panel:** results at $p = 100$, **right panel:** results at $p = 400$.

a similar limiting distribution was first presented in context of estimation of a change point in the mean of a random variable.

The U-shaped trends observed in the left and right panels of Figure 3 are indicative of any one of two underlying effects. First, the boundary effect, i.e., as the change point moves closer to the boundary of $(0, 1)$ the effective sample size on one of the induced binary segements is reduced thereby causing the diminished precision. Second, is the intializer effect, i.e., these U-shaped trends are potentially indicative of the reach of the initializing choice $\check{\tau}$ of Algorithm which is set to 0.5. Upon observing the distinctions between the left and right panel of Figure 3, these effects are observed to be further compounded (as expected) with an increase in dimensionality. These concerns are however alleviated by observing their diminishing effects over an increasing sampling period T, thereby providing numerical evidence for its statistical validity.

References

Daniele Angelosante and Georgios B Giannakis. Sparse graphical modeling of piecewise-stationary time series. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1960–1963. IEEE, 2011.

Yves Atchade and Leland Bybee. A scalable algorithm for gaussian graphical models with change-points. *arXiv preprint arXiv:1707.04306*, 2017.

Jushan Bai. Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 15(5):453–472, 1994.

Jushan Bai. Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4):551–563, 1997.

Onureena Banerjee, Laurent El Ghaoui, and Alexandre dAspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine learning research*, 9(Mar):485–516, 2008.

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*, 2011.

Alexandre Belloni, Victor Chernozhukov, and Abhishek Kaul. Confidence bands for coefficients in high dimensional linear models with error-in-variables. *arXiv preprint arXiv:1703.00469*, 2017.

Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Change point estimation in a dynamic stochastic block model. *arXiv preprint arXiv:1812.03090*, 2018.

T	p	bias ($\hat{\tau}$) ($\times 10^2$)	bias ($\tilde{\tau}$) ($\times 10^2$)	rmse ($\hat{\tau}$) ($\times 10^2$)	rmse ($\tilde{\tau}$) ($\times 10^2$)	SE ($T\tilde{\tau}$)	c_α
200	100	30.26	29.18	31.47	30.76	2.27	10.22
200	200	33.81	33.58	33.94	33.75	2.27	10.22
200	300	34.37	34.33	34.43	34.39	2.26	10.22
200	400	34.48	34.45	34.57	34.55	2.27	10.22
275	100	18.72	17.62	23.48	22.91	2.27	10.22
275	200	26.14	25.43	28.76	28.23	2.27	10.22
275	300	30.27	30.15	31.49	31.39	2.26	10.22
275	400	32.52	32.47	32.99	32.94	2.27	10.22
350	100	7.43	5.54	13.66	12.39	2.27	10.22
350	200	7.55	6.44	13.18	12.53	2.27	10.22
350	300	13.37	12.3	18.9	18.14	2.26	10.22
350	400	14.28	13.25	19.76	18.95	2.27	10.22
425	100	1.74	0.37	4.94	2.60	2.27	10.22
425	200	1.20	0.25	3.32	1.08	2.27	10.22
425	300	1.01	0.23	2.32	1.16	2.26	10.22
425	400	0.97	0.42	3.05	2.44	2.27	10.22

Table 1: Summary of numerical results at $\tau^0 = 0.15$. The estimates $\hat{\tau}$, and $\tilde{\tau}$ are those of Step 1 and Step 2 of Algorithm 1, respectively.

T	p	bias ($\hat{\tau}$) ($\times 10^2$)	bias ($\tilde{\tau}$) ($\times 10^2$)	rmse ($\hat{\tau}$) ($\times 10^2$)	rmse ($\tilde{\tau}$) ($\times 10^2$)	SE ($\lfloor T\tilde{\tau} \rfloor$)	c_α
200	100	18.27	16.76	20.82	20.04	2.27	10.22
200	200	24.2	23.38	24.86	24.32	2.27	10.22
200	300	25.89	25.77	26.13	26.04	2.26	10.22
200	400	26.59	26.54	26.62	26.57	2.27	10.22
275	100	5.89	3.44	10.34	8.54	2.27	10.22
275	200	7.31	4.60	11.61	9.97	2.27	10.22
275	300	12.03	9.52	15.92	14.41	2.26	10.22
275	400	15.82	13.88	18.79	17.83	2.27	10.22
350	100	0.77	0.03	2.51	1.41	2.27	10.22
350	200	0.22	0.12	0.89	0.15	2.27	10.22
350	300	0.24	0.10	0.87	0.19	2.26	10.22
350	400	0.27	0.12	1.07	0.15	2.27	10.22
425	100	0.30	0.09	0.58	0.15	2.27	10.22
425	200	0.14	0.07	0.25	0.09	2.27	10.22
425	300	0.16	0.07	0.30	0.08	2.26	10.22
425	400	0.11	0.06	0.18	0.06	2.27	10.22

Table 2: Summary of numerical results at $\tau^0 = 0.23$. The estimates $\hat{\tau}$, and $\tilde{\tau}$ are those of Step 1 and Step 2 of Algorithm 1, respectively.

Peter J Bickel. On adaptive estimation. *The Annals of Statistics*, pages 647–671, 1982.

Tony Cai, Weidong Liu, and Xi Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

Victor Chernozhukov, Christian Hansen, and Martin Spindler. Valid post-selection and post-regularization inference: An elementary, general approach. *Annu. Rev. Econ.*, 7(1):649–688, 2015.

Ivor Cribben, Ragnheidur Haraldsdottir, Lauren Y Atlas, Tor D Wager, and Martin A Lindquist. Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage*, 61(4):907–920, 2012.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

Alex J. Gibberd and Sandipan Roy. Multiple changepoint estimation in high-dimensional gaussian graphical models. *arXiv preprint arXiv:1712.05786*, 2017.

J Hájek and A Rényi. Generalization of an inequality of kolmogorov. *Acta Mathematica Hungarica*, 6(3-4):281–283, 1955.

Abhishek Kaul, Ori Davidov, and Shyamal D. Peddada. Structural zeros in high-dimensional data with applications to microbiome studies. *Biostatistics*, 18(3):422–433, 2017.

Abhishek Kaul, Venkata K Jandhyala, and Stergios B Fotopoulos. An efficient two step algorithm for high dimensional change point regression models without grid search. *Journal of Machine Learning Research*, 20(111):1–40, 2019.

Hossein Keshavarz, George Michailidis, and Yves Atchadé. Sequential change-point detection in high-dimensional gaussian graphical models. *arXiv preprint arXiv:1806.07870*, 2018.

Mladen Kolar and Eric P Xing. Estimating networks with jumps. *Electronic journal of statistics*, 6: 2069, 2012.

Mladen Kolar, Le Song, Amr Ahmed, Eric P Xing, et al. Estimating time-varying networks. *The Annals of Applied Statistics*, 4(1):94–123, 2010.

Po-Ling Loh and Martin J. Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.*, 40(3):1637–1664, 06 2012. doi: 10.1214/12-AOS1018. URL `https://doi.org/10.1214/12-AOS1018`.

Malte Londschien, Solt Kovács, and Peter Bühlmann. Change point detection for graphical models in presence of missing values. *arXiv preprint arXiv:1907.05409*, 2019.

Nicolai Meinshausen, Peter Bühlmann, et al. High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, 34(3):1436–1462, 2006.

Yang Ning, Han Liu, et al. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45(1):158–195, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

Philippe Rigollet. 18. s997: High dimensional statistics. *Lecture Notes), Cambridge, MA, USA: MIT Open-CourseWare*, 2015.

Sandipan Roy, Yves Atchadé, and George Michailidis. Change point estimation in high dimensional markov random-field models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1187–1206, 2017.

Aad W Vaart and Jon A Wellner. *Weak convergence and empirical processes: with applications to statistics*. Springer, 1996.

Sara Van de Geer, Peter Bühlmann, Yaacov Ritov, Ruben Dezeure, et al. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.

Roman Vershynin. *High-Dimensional Probability.* Cambridge, UK: Cambridge University Press, 2019. URL `https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf`.

Daren Wang, Yi Yu, and Alessandro Rinaldo. Optimal change point detection and localization in sparse dynamic networks. *arXiv preprint arXiv:1809.09602*, 2018.

Yi-Ching Yao. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics*, 15(3):1321–1328, 1987.

Ming Yuan. High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, 11(Aug):2261–2286, 2010.

Supplementary materials: Inference on the Change Point in High Dimensional Dynamic Graphical Models

A Proofs of results of Section 2

The following notations are required for readability of this section. In addition to $\xi_{2,2}$ defined in the $\ell_{2,2}$ norm in Condition A, we also define $\xi_{2,1} = \sum_{j=1}^{p} \|\eta^0_{(j)}\|_2$ in the $\ell_{2,1}$ norm. Also, in all to follow we denote as $\hat\eta_{(j)} = \hat\mu_{(j)} - \hat\gamma_{(j)}$, $j = 1, ..., p$.

Proof of Lemma 2.1. For any fixed $\tau \geq \tau^0$ consider,

$$\begin{aligned}
\mathcal{U}(z,\tau,\hat\mu,\hat\gamma) &= Q(z,\tau,\hat\mu,\hat\gamma) - Q(z,\tau^0,\hat\mu,\hat\gamma)\\
&= \frac{1}{T}\sum_{t=1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat\mu_{(j)})^2 + \frac{1}{T}\sum_{t=\lfloor T\tau\rfloor+1}^{T}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat\gamma_{(j)})^2\\
&\quad -\frac{1}{T}\sum_{t=1}^{\lfloor T\tau^0\rfloor}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat\mu_{(j)})^2 - \frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{T}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat\gamma_{(j)})^2\\
&= \frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat\mu_{(j)})^2 - \frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat\gamma_{(j)})^2\\
&= \frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\big(z_{t,-j}^T\hat\eta_{(j)}\big)^2 - \frac{2}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\hat\eta_{(j)}\\
&\quad +\frac{2}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(\hat\gamma_{(j)} - \gamma^0_{(j)})^T z_{t,-j}z_{t,-j}^T\hat\eta_{(j)}. \qquad\text{(A.1)}
\end{aligned}$$

The expansion in (A.1) provides the following relation,

$$\begin{aligned}
\inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \mathcal{U}(z,\tau,\hat\mu,\hat\gamma) &\geq \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\big(z_{t,-j}^T\hat\eta_{(j)}\big)^2\\
&\quad -2\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\hat\eta_{(j)}\Big|\\
&\quad -2\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(\hat\gamma_{(j)} - \gamma^0_{(j)})^T z_{t,-j}z_{t,-j}^T\hat\eta_{(j)}\Big|\\
&= R1 - R2 - R3 \qquad\text{(A.2)}
\end{aligned}$$

Bounds for terms $R1, R2$ and $R3$ have been provided in Lemma C.6 and Lemma C.7. In particular,

$$\begin{aligned}
R1 &\geq \kappa\xi_{2,2}^2\Big[v_T - \frac{c_{a1}\sigma^2}{\kappa}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}} - c_u(\sigma^2\vee\phi)\frac{u_T}{\kappa\xi_{2,2}}\Big\{s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)} - \eta^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}}\Big]\\
&\geq \kappa\xi_{2,2}^2\Big[v_T - \frac{c_{a1}\sigma^2}{\kappa}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}} - c_{u1}(\sigma^2\vee\phi)\frac{u_T}{\kappa T^b}\Big],
\end{aligned}$$

with probability at least $1-a-o(1)$. Here the first inequality follows from Lemma C.6 and the final inequality follows by using the bounds of Lemma C.7. Next we obtain upper bounds for the terms $R2\big/\kappa\xi_{2,2}^2$ and $R3\big/\kappa\xi_{2,2}^2$. For this purpose, first note that $(\xi_{2,1}/\xi_{2,2})\le\sqrt{p}$, consequently $(\xi_{2,1}/\xi_{2,2}^2)\le 1/\psi$. Now consider,

$$\begin{aligned}\frac{R2}{\kappa\xi_{2,2}^2} &\le c_{a1}\surd(1+\nu^2)\frac{\sigma^2\xi_{2,1}}{\kappa\xi_{2,2}^2}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}+c_u\surd(1+\nu^2)\frac{\sigma^2}{\kappa\xi_{2,2}^2}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta^0_{(j)}\|_1\\ &\le c_{a1}\surd(1+\nu^2)\frac{\sigma^2}{\kappa\psi}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}+\Big\{\surd(1+\nu^2)\frac{\sigma^2}{\kappa\psi}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\Big\}\Big\{c_u\surd(1+\nu^2)\frac{\sigma^2}{\kappa\psi}\frac{s\log^{3/2}(p\vee T)}{\surd(Tl_T)}\Big\}\\ &\le c_uc_{a1}\surd(1+\nu^2)\frac{\sigma^2}{\kappa\psi}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\end{aligned}$$

with probability at least $1-a-o(1)$. As before, the first inequality follows from Lemma C.6 and the final inequality follows by using the bounds of Lemma C.7. Similarly we can also obtain,

$$\begin{aligned}\frac{R3}{\kappa\xi_{2,2}^2} &\le c_u(\sigma^2\vee\phi)\frac{u_T}{\kappa\xi_{2,2}}\Big\{s\log(p\vee T)\sum_{j=1}^{p}\|\hat\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}}\Big[1+\frac{1}{\xi_{2,2}}\Big\{s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}}\Big]\\ &\le c_{u1}(\sigma^2\vee\phi)\frac{u_T}{\kappa T^b}\end{aligned}$$

with probability at least $1-a-o(1)$. Substituting these bounds in (A.2) and applying a union bound over these three events yields the bound (2.2) uniformly over the set $\{\mathcal{G}(u_T,v_T);\,\tau\ge\tau^0\}$. The mirroring case of $\tau\le\tau^0$ can be obtained by similar arguments. □

Proof of Theorem 2.1. To prove this result, we show that for any $0<a<1$, the bound

$$\big|\lfloor T\tilde\tau\rfloor-\lfloor T\tau^0\rfloor\big|\le c_{a3}^2, \tag{A.3}$$

holds with probability at least $1-3a-o(1)$. Note that Part (i) of this theorem is a direct consequence of the bound (A.3) in the case where $\psi\to 0$. The proof for the bound (A.3) to follow relies on a recursive argument on Lemma 2.1, where the optimal rate of convergence $O_p(1)$ is obtained by a series of recursions with the rate of convergence being sharpened at each step.

We begin by considering any $v_T>0$, and applying Lemma 2.1 on the set $\mathcal{G}(1,v_T)$ to obtain,

$$\inf_{\tau\in\mathcal{G}(1,v_T)}\mathcal{U}(z,\tau,\hat\mu,\hat\gamma)\ge\kappa\xi_{2,2}^2\Big[v_T-c_{a3}\max\Big\{\Big(\frac{1}{T}\Big)^{\frac{1}{2}},\frac{1}{T^b}\Big\}\Big]$$

with probability at least $1-3a-o(1)$. Recall by assumption $b<(1/2)$, and choose any $v_T>v_T^*=c_{a3}/T^b$. Then we have $\inf_{\tau\in\mathcal{G}(1,v_T)}\mathcal{U}(z,\tau,\hat\mu,\hat\gamma)>0$, thus implying that $\tilde\tau\notin\mathcal{G}(1,v_T)$, i.e., $\big|\lfloor T\tilde\tau\rfloor-\lfloor T\tau^0\rfloor\big|\le Tv_T^*$, with probability at least $1-3a-o(1)$[g]. Now reset $u_T=v_T^*$ and reapply Lemma 2.1 for any $v_T>0$ to obtain,

$$\inf_{\tau\in\mathcal{G}(u_T,v_T)}\mathcal{U}(z,\tau,\hat\mu,\hat\gamma)\ge\kappa\xi_{2,2}^2\Big[v_T-c_{a3}\max\Big\{\Big(\frac{c_{a3}}{T^{1+b}}\Big)^{\frac{1}{2}},\frac{c_{a3}}{T^{b+b}}\Big\}\Big]$$

Again choosing any,

$$v_T>v_T^*=\max\Big\{\frac{c_{a3}^{g_2}}{T^{u_2}},\ \frac{c_{a3}^2}{T^{v_2}}\Big\}, \tag{A.4}$$

[g]Since by construction of $\tilde\tau$ we have, $\mathcal{U}(\tilde\tau,\hat\gamma,\hat\gamma)\le 0$.

where,

$$g_2 = 1 + \frac{1}{2},\ u_2 = \frac{1}{2} + \frac{u_1}{2},\ \text{and } v_2 = b + v_1 \geq 2b,\ \text{with } u_1 = v_1 = b,$$

we obtain $\inf_{\mathcal{G}(u_T,v_T)} \mathcal{U}(z,\tau,\hat{\mu},\hat{\gamma}) > 0$, with probability at least $1-3a-o(1)$. Consequently $\tilde{\tau} \notin \mathcal{G}(u_T,v_T)$, i.e., $\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq Tv_T^*$. Note that rate of convergence of $\tilde{\tau}$ has been sharpened at the second recursion in comparison to the first. Continuing these recursions by resetting u_T to the bound of the previous recursion, and applying Lemma 2.1, we obtain for the m^{th} recursion,

$$\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq T\max\Big\{\frac{c_{a3}^{g_m}}{T^{u_m}},\ \frac{c_{a3}^{m}}{T^{v_m}}\Big\} := T\max\{R_{1m}, R_{2m}\}, \qquad \text{where,}$$

$$g_m = \sum_{k=0}^{m-1}\frac{1}{2^k},\ u_m = \frac{1}{2} + \frac{u_{m-1}}{2} = \frac{b}{m} + \sum_{k=1}^{m}\frac{1}{2^m},\ \text{and } v_m = b + v_{m-1} \geq mb,\ \text{with } u_1 = v_1 = b$$

Next, we observe that for m large enough, $R_{2m} \leq R_{1m}$. This follows since R_{2m} is faster than any polynomial rate of $1/T$.[h] Consequently for m large enough we have $\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq TR_{1m}$, with probability at least $1-3a-o(1)$. Finally, we continue these recursions an infinite number of times to obtain, $g_\infty = \sum_{k=0}^{\infty} 1/2^k$, $u_\infty = \sum_{k=1}^{\infty}(1/2^k)$, thus yielding,

$$\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq T\frac{c_{a3}^2}{T} = c_{a3}^2$$

with probability at least $1-3a-o(1)$. This proves the bound (A.3). To finish the proof, note that despite the recursions in the argument, the probability bound after every step is maintained at $1-3a-o(1)$. This follows since the probability statement of Lemma 2.1 arises from stochastic upper bounds of Lemma C.2, Lemma C.3, Lemma C.4 and Lemma E.2, applied recursively, with a tighter bound at each recursion. This yields a sequence of events such that each event is a proper subset of the event at the previous recursion. □

For a clear presentation of the proof of Theorem 2.2 we use the following additional notation. Denote by,

$$\hat{\mathcal{U}}(\tau) = \mathcal{U}(z,\tau,\hat{\mu},\hat{\gamma}), \quad \text{and} \quad \mathcal{U}(\tau) = \mathcal{U}(z,\tau,\mu^0,\gamma^0),$$

where $\mathcal{U}(z,\tau,\mu,\gamma)$ is defined in (2.1). The proof of this theorem shall also rely on the 'Argmax' theorem, see, Theorem 3.2.2 of Vaart and Wellner [1996] (reproduced as Theorem F.2).

Proof of Theorem 2.2. Under the assumed regime of $\psi \to 0$, recall from Remark 2.1 that we have $T\psi^2(\tilde{\tau}-\tau^0) = O_p(1)$. It is thus sufficient to examine the behavior of $\tilde{\tau}$, such that $\tilde{\tau} = \tau^0 + rT^{-1}\psi^{-2}$, with $r \in [-c_1, c_1]$, for a given constant $c_1 > 0$. Now in view of 'Argmax' theorem (Theorem F.2), in order to prove the statement of this theorem it is sufficient to establish the following results,

$$(i) \qquad \sup_{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big)} Tp^{-1}\big|\hat{\mathcal{U}}(\tau) - \mathcal{U}(\tau)\big| = o_p(1), \quad \text{and}$$

$$(ii) \quad Tp^{-1}\mathcal{U}(\tau^0 + r\psi^{-2}T^{-1}) \Rightarrow G(r) = \begin{cases} \sigma_2^2|r| - 2\sigma_2^* W_1(r), & \text{if } r > 0, \\ 0, & \text{if } r = 0,, \\ \sigma_1^2|r| - 2\sigma_1^* W_2(r), & \text{if } r < 0. \end{cases} \tag{A.5}$$

[h] Consider $c_1^m/T^{mb} \leq (c_1/\log T)^m(\log T/T)^{mb} \leq (1/T^{mb_1})$, for any $0 < b_1 < b$, for T sufficiently large.

Then it is straightforward to show that $\arg\min_r G(r) =^d \left(\sigma_1^{*2}/\sigma_1^4\right) \arg\min_r Z(r)$, where $Z(r)$ is as defined in (2.3) and $=^d$ represents equality in distribution, see, e.g. proof of Proposition 3 of Bai [1997]. Thereby yielding the statement of this theorem. **Step 1**, and **Step 2** below provides the results (i), and (ii) of (A.5), respectively.

Step 1: For any $\tau \geq \tau^0$, first define the following,

$$\begin{aligned} R_1 &= p^{-1} \sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor} \sum_{j=1}^{p} \|z_{t,-j}^T \hat{\eta}_{(j)}\|_2^2 - 2p^{-1} \sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor} \sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j}^T \hat{\eta}_{(j)} \\ &\quad +2p^{-1} \sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor} \sum_{j=1}^{p} (\hat{\gamma}_{(j)} - \gamma_{(j)}^0)^T z_{t,-j} z_{t,-j}^T \hat{\eta}_{(j)} = R_{11} - 2R_{12} + 2R_{13}, \\ R_2 &= p^{-1} \sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor} \sum_{j=1}^{p} \|z_{t,-j}^T \eta_{(j)}^0\|_2^2 - 2p^{-1} \sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor} \sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j}^T \eta_{(j)}^0 = R_{21} - 2R_{22}. \end{aligned}$$

Then we have the following algebraic expansion,

$$\begin{aligned} Tp^{-1}\big(\hat{\mathcal{U}}(\tau) - \mathcal{U}(\tau)\big) &= Tp^{-1}\Big(Q(z,\tau,\hat{\beta},\hat{\gamma}) - Q(z,\tau^0,\hat{\beta},\hat{\gamma})\Big) \\ &\quad -Tp^{-1}\Big(Q(z,\tau,\beta^0,\gamma^0) - Q(z,\tau^0,\beta^0,\gamma^0)\Big) \\ &= \big(R_1 - R_2\big) = \Big\{\big(R_{11} - 2R_{12} + 2R_{13}\big) - \big(R_{21} - 2R_{22}\big)\Big\}. \end{aligned} \tag{A.6}$$

Lemma C.8 shows that the expressions $\big|R_{11} - R_{21}\big|$, $\big|R_{12} - R_{22}\big|$, and $|R_{13}|$ are $o_p(1)$ uniformly over the set $\{\mathcal{G}\big(c_1T^{-1}\psi^{-2}, 0\big)\} \cap \{\tau \geq \tau^0\}$. The same result can be obtained symmetrically on the set $\{\mathcal{G}\big(c_1T^{-1}\psi^{-2}, 0\big)\} \cap \{\tau \leq \tau^0\}$, thereby yielding $o_p(1)$ bounds for these terms uniformly over $\mathcal{G}\big(c_1T^{-1}\psi^{-2}, 0\big)$ Consequently,

$$\begin{aligned} \sup_{\tau\in\mathcal{G}\left((c_1T^{-1}\psi^{-2}),0\right)} Tp^{-1}\big|\hat{\mathcal{U}}(\tau) - \mathcal{U}(\tau)\big| &\leq \sup_{\tau\in\mathcal{G}\left((c_1T^{-1}\psi^{-2}),0\right)} |R_{11} - R_{21}| \\ &\quad + \sup_{\tau\in\mathcal{G}\left((c_1T^{-1}\psi^{-2}),0\right)} 2|R_{12} - R_{22}| + \sup_{\tau\in\mathcal{G}\left((c_1T^{-1}\psi^{-2}),0\right)} 2|R_{13}| = o_p(1) \end{aligned}$$

This completes the proof of **Step 1**.

Step 2: Consider $\tau^* = \tau^0 + rT^{-1}\psi^{-2}$, with $r \in (0, c_1]$. Then using Lemma C.9 we have,

$$p^{-1} \sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^*\rfloor} \sum_{j=1}^{p} \eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T \eta_{(j)}^0 \to_p r\sigma_2^2. \tag{A.7}$$

Next, let $\zeta_t = \sum_{j=1}^p \zeta_{tj} = \sum_{j=1}^p \varepsilon_{tj} z_{t,-j}^T \eta_{(j)}^0$, and note that $\mathrm{var}\big(p^{-1/2}\psi^{-1}\zeta_t\big) = \mathrm{var}\big(\xi_{2,2}^{-1}\zeta_t\big)$. Then using Condition E we obtain,

$$p^{-1} \sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^*\rfloor} \sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j} \eta_{(j)}^0 = p^{-1/2}\psi \sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^*\rfloor} p^{-1/2}\psi^{-1}\zeta_t \Rightarrow \sigma_2^* W_1(r), \tag{A.8}$$

where $W_1(r)$ is a Brownian motion on $[0,\infty)$. Now consider $Tp^{-1}\mathcal{U}(\tau^*)$, and observe that an algebraic simplification yields,

$$\begin{aligned}
Tp^{-1}\mathcal{U}(\tau^*) &= p^{-1}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^*\rfloor}\sum_{j=1}^{p}\big(z_{tj}-z_{t,-j}^T\mu_{(j)}^0\big)^2 - p^{-1}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^*\rfloor}\sum_{j=1}^{p}\big(z_{tj}-z_{t,-j}^T\gamma_{(j)}^0\big)^2\\
&= p^{-1}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^*\rfloor}\sum_{j=1}^{p}\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0 - 2p^{-1}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^*\rfloor}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0\\
&\Rightarrow \big\{\sigma_2^2 r - 2\sigma_2^* W_1(r)\big\},
\end{aligned}$$

where the convergence in distribution follows from (A.7) and (A.8). Similarly for $\tau^*=\tau^0+rT^{-1}\psi^{-2}$, with $r\in[-c_1,0)$, it can be shown that,

$$T\mathcal{U}(\tau^*)\Rightarrow\{\sigma_1^2(-r)-2\sigma_1^*W_2(r)\},$$

where $W_2(r)$ is another Brownian motion on $[0,\infty)$ independent of $W_1(r)$. This completes the proof of **Step 2** and the proof of this theorem. □

B Proofs of results of Section 3

The main result of Section 3 is Theorem 3.1, which forms the basis of the subsequent corollaries. The proof of Theorem 3.1 requires some preliminary work in the form of Theorem B.1, Lemma B.1 and Lemma B.2 below. We begin with Theorem B.1 that provides uniform bounds (over τ) of the ℓ_2 error in the lasso estimates (3.1) obtained from a regression of each column of z on the rest.

Theorem B.1. *Suppose Condition A′ and B holds. Let $u_T\geq 0$ and $\lambda_j=2(\lambda_{1j}+\lambda_{2j})$, where*

$$\lambda_{1j}=c_u\sigma^2\sqrt{(1+\nu^2)}\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}},\;\;\lambda_{2j}=c_u(\sigma^2\vee\phi)\|\eta_{(j)}^0\|_2\max\Big\{\frac{\log(p\vee T)}{Tl_T},\;\frac{u_T}{l_T}\Big\}$$

Then uniformly over all $j=1,...,p$, the following two properties hold with probability at least $1-c_{u2}\exp\big\{-(c_{u3}\log(p\vee T)\big\}$, for some $c_{u2},c_{u3}>0$.
(i) The vectors $\hat\mu_{(j)}(\tau)-\mu_{(j)}^0\in\mathcal{A}_{1j}$, and $\hat\gamma_{(j)}(\tau)-\gamma_{(j)}^0\in\mathcal{A}_{2j}$, where the sets $\mathcal{A}_{ij}$, $i=1,2$, and $j=1,...,p$ are as defined in Condition C.
(ii) For any constant $c_{u1}>0$, we have,

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,0);\\ (\lfloor T\tau\rfloor)\wedge(T-\lfloor T\tau\rfloor)\geq c_{u1}Tl_T}}\|\hat\mu_{(j)}(\tau)-\mu_{(j)}^0\|_2\leq c_u\frac{\sqrt{s}}{\kappa}\lambda_j.$$

The same upper bounds also hold for $\hat\gamma_{(j)}(\tau)-\gamma_{(j)}^0$, uniformly over j and τ.

Proof of Theorem B.1. Consider any $\tau\in\mathcal{G}(u_T,0)$, and w.l.o.g. assume that $\tau\geq\tau^0$. Then for any $j=1,..,p$, by construction of the estimator $\hat\mu_{(j)}(\tau)$, we have the basic inequality,

$$\frac{1}{\lfloor T\tau\rfloor}\sum_{t=1}^{\lfloor T\tau\rfloor}\big(z_{tj}-z_{t,-j}^T\hat\mu_{(j)}(\tau)\big)^2+\lambda_j\|\hat\mu_{(j)}(\tau)\|_1\leq\frac{1}{\lfloor T\tau\rfloor}\sum_{t=1}^{\lfloor T\tau\rfloor}\big(z_{tj}-z_{t,-j}^T\mu_{(j)}^0\big)^2+\lambda_j\|\mu_{(j)}^0\|_1.$$

An algebraic rearrangement of this inequality yields,

$$\frac{1}{\lfloor T\tau \rfloor}\sum_{t=1}^{\lfloor T\tau \rfloor}\big(z_{t,-j}^T(\hat{\mu}_{(j)}-\mu_{(j)}^0)\big)^2+\lambda_j\|\hat{\mu}_{(j)}(\tau)\|_1\leq\lambda_j\|\mu_{(j)}^0\|_1+\frac{2}{\lfloor T\tau \rfloor}\sum_{t=1}^{\lfloor T\tau \rfloor}\tilde{\varepsilon}_{tj}z_{t,-j}^T(\hat{\mu}_{(j)}-\mu_{(j)}^0),$$

where $\tilde{\varepsilon}_{tj}=\varepsilon_{tj}=z_{tj}-z_{t,-j}^T\mu_{(j)}^0$, for $t\leq\lfloor T\tau^0\rfloor$, and $\tilde{\varepsilon}_{tj}=z_{tj}-z_{t,-j}^T\mu_{(j)}^0=\varepsilon_{tj}-z_{t,-j}^T(\mu_{(j)}^0-\gamma_{(j)}^0)$, for $t>\lfloor T\tau^0\rfloor$. A further simplification using these relations yields,

$$\begin{aligned}\frac{1}{\lfloor T\tau \rfloor}\sum_{t=1}^{\lfloor T\tau \rfloor}\big(z_{t,-j}^T(\hat{\mu}_{(j)}-\mu_{(j)}^0)\big)^2+\lambda_j\|\hat{\mu}_{(j)}(\tau)\|_1&\leq\lambda_j\|\mu_{(j)}^0\|_1+\frac{2}{\lfloor T\tau \rfloor}\sum_{t=1}^{\lfloor T\tau \rfloor}\varepsilon_{tj}z_{t,-j}^T(\hat{\mu}_{(j)}-\mu_{(j)}^0)\\&\quad-\frac{2}{\lfloor T\tau \rfloor}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau \rfloor}(\mu_{(j)}^0-\gamma_{(j)}^0)z_{t,-j}z_{t,-j}^T(\hat{\mu}_{(j)}-\mu_{(j)}^0)\\&\leq\lambda\|\mu_{(j)}^0\|_1+\frac{2}{\lfloor T\tau \rfloor}\Big\|\sum_{t=1}^{\lfloor T\tau \rfloor}\varepsilon_{tj}z_{t,-j}^T\Big\|_\infty\|\hat{\mu}_{(j)}-\mu_{(j)}^0\|_1\\&\quad+\frac{2}{\lfloor T\tau \rfloor}\Big\|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau \rfloor}(\mu_{(j)}^0-\gamma_{(j)}^0)z_{t,-j}z_{t,-j}^T\Big\|_\infty\|\hat{\mu}_{(j)}-\mu_{(j)}^0\|_1\end{aligned}\tag{B.1}$$

Now using the bounds of Lemma D.1 we have that,

$$\begin{aligned}&\frac{1}{\lfloor T\tau \rfloor}\Big\|\sum_{t=1}^{\lfloor T\tau \rfloor}\varepsilon_{tj}z_{t,-j}\Big\|_\infty\leq c_u\sigma^2\sqrt{(1+\nu^2)}\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}}=\lambda_{1j}\\&\frac{1}{\lfloor T\tau \rfloor}\Big\|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau \rfloor}\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\Big\|_\infty\leq c_u(\sigma^2\vee\phi)\|\eta_{(j)}^0\|_2\max\Big\{\frac{\log(p\vee T)}{Tl_T},\ \frac{u_T}{l_T}\Big\}=\lambda_{2j},\end{aligned}$$

with probability at least $1-c_{u2}\exp\{-c_{u3}\log(p\vee T)\}$. Applying these bounds in (B.1) yields,

$$\frac{1}{\lfloor T\tau \rfloor}\sum_{t=1}^{\lfloor T\tau \rfloor}\big(z_{t,-j}^T(\hat{\mu}_{(j)}-\mu_{(j)}^0)\big)^2+\lambda_j\|\hat{\mu}_{(j)}(\tau)\|_1\leq\lambda_j\|\mu_{(j)}^0\|_1+(\lambda_{1j}+\lambda_{2j})\|\hat{\mu}_{(j)}(\tau)-\mu_{(j)}^0\|_1,$$

with probability at least $1-c_{u2}\exp\{-c_{u3}\log(p\vee T)\}$. Choosing $\lambda_j\geq 2(\lambda_{1j}+\lambda_{2j})$, leads to $\big\|\big(\hat{\mu}_{(j)}(\tau)\big)_{S_{1j}^c}\big\|_1\leq 3\big\|\big(\hat{\mu}_{(j)}(\tau)-\mu_j^0\big)_{S_{1j}}\big\|_1$, and thus by definition $\hat{\mu}_{(j)}-\mu_{(j)}^0\in\mathcal{A}_{1j}$, with the same probability. This proves the first assertion of this theorem. Next applying the restricted eigenvalue condition of (E.3) to the l.h.s. of the inequality (B.1), we also have that,

$$\kappa\|\hat{\mu}_{(j)}(\tau)-\mu_{(j)}^0\|_2^2\leq 3\lambda\|\hat{\mu}_{(j)}(\tau)-\mu_{(j)}^0\|_1\leq 3\sqrt{s}\lambda_j\|\hat{\mu}_{(j)}(\tau)-\mu_{(j)}^0\|_2.$$

This directly implies that $\|\hat{\mu}_{(j)}(\tau)-\mu_{(j)}^0\|_2\leq 3\sqrt{s}(\lambda_j/\kappa)$, which yields the desired ℓ_2 bound. To finish the proof recall that the stochastic bounds used here hold uniformly over $\mathcal{G}(u_T,0)$, and j, consequently the statements of this theorem also hold uniformly over the same collections. The case of $\tau\leq\tau^0$, and the corresponding results for $\hat{\gamma}_{(j)}(\tau)-\gamma_{(j)}^0$ can be obtained by symmetrical arguments. □

The following lemma obtains ℓ_2 error bounds for the Step 1 edge estimates by utilizing the initializing Condition F and Theorem B.1.

Lemma B.1. *Suppose Condition A$'$, B and F hold. Choose regularizers* λ_j, $j = 1, ..., p$, *as prescribed in Theorem B.1, with* $u_T = \big(c_u l_T \kappa\big)\big/\big(sT^k(\sigma^2 \vee \phi)\big)$. *Then edge estimates* $\check{\mu}_{(j)}$, $j = 1, ..., p$ *of Step 1 of Algorithm 1 satisfy the following bound.*

$$(i)\sqrt{s}\sum_{j=1}^{p}\|\check{\mu}_{(j)} - \mu^0_{(j)}\|_2 \le \frac{c_u\xi_{2,1}}{T^k}, \text{ and } (ii)\Big(s\sum_{j=1}^{p}\|\check{\mu}_{(j)} - \mu^0_{(j)}\|_2^2\Big)^{\frac{1}{2}} \le \frac{c_u\xi_{2,2}}{T^k}$$

with probability $1 - o(1)$. *Corresponding bounds also holds for* $\check{\gamma}_{(j)}$, $j = 1, ..., p$.

Proof of Lemma B.1. We begin by noting that Part (ii) of the initializing Condition F of Algorithm 1 guarantees that $\check{\tau}$ satisfies,

$$|\lfloor T\check{\tau}\rfloor - \lfloor T\tau^0\rfloor| \le \frac{c_u l_T \kappa}{s(\sigma^2 \vee \phi)}T^{(1-k)}$$

In other words, $\check{\tau} \in \mathcal{G}(u_T, 0)$, where $u_T = \big(c_u l_T\kappa\big)\big/\big(sT^k(\sigma^2 \vee \phi)\big)$, where $k < b$. This choice of u_T provides the following relations,

$$\frac{u_T}{l_T} = \frac{c_u\kappa}{(\sigma^2 \vee \phi)T^k s} \ge \frac{\log(p \vee T)}{Tl_T}. \tag{B.2}$$

$$c_u(\sigma^2 \vee \phi)\sqrt{s}\frac{\xi_{2,1}u_T}{\kappa l_T} = \frac{c_u\xi_{2,1}}{T^k\sqrt{s}} \ge c_u\sigma^2\surd(1+\nu^2)\frac{p}{\kappa}\Big\{\frac{s\log(p \vee T)}{Tl_T}\Big\}^{\frac{1}{2}} \tag{B.3}$$

Here the inequality of (B.2) follows from the assumption $c_u\kappa T^{(1-k)}l_T \ge (\sigma^2 \vee \phi)s\log(p \vee T)$ of Condition F. The equality of (B.3) follows directly upon substituting the choice of u_T, and the inequality follows from assumption A$'$(iii) and since w.l.o.g we have $k < b$. Now using this choice of u_T in λ_j of Part (ii) of Theorem B.1 we obtain,

$$\begin{aligned}
\sum_{j=1}^{p}\frac{\sqrt{s}}{\kappa}(\lambda_{1j} + \lambda_{2j}) &\le c_u\sigma^2\surd(1+\nu^2)\frac{p}{\kappa}\Big\{\frac{s\log(p \vee T)}{Tl_T}\Big\}^{\frac{1}{2}} + c_u(\sigma^2 \vee \phi)\xi_{2,1}\frac{\sqrt{s}}{\kappa}\Big\{\frac{\log(p \vee T)}{Tl_T},\ \frac{u_T}{l_T}\Big\}\\
&\le c_u\sigma^2\surd(1+\nu^2)\frac{p}{\kappa}\Big\{\frac{s\log(p \vee T)}{Tl_T}\Big\}^{\frac{1}{2}} + c_u(\sigma^2 \vee \phi)\Big\{\frac{\xi_{2,1}u_T\sqrt{s}}{\kappa l_T}\Big\} \le c_u\frac{\xi_{2,1}}{T^k\sqrt{s}}.
\end{aligned}$$

The second inequality follows from (B.2) and the final inequality follows from (B.3). The bound of Part (i) is now a direct consequence of Theorem B.1. We proceed similarly to prove Part (ii), note that,

$$\begin{aligned}
\sum_{j=1}^{p}\frac{s}{\kappa^2}(\lambda_{1j} + \lambda_{2j})^2 &\le c_u\sigma^4(1+\nu^2)\frac{p}{\kappa^2}\Big\{\frac{s\log(p \vee T)}{Tl_T}\Big\} + c_u(\sigma^4 \vee \phi^2)\xi_{2,2}^2\frac{s}{\kappa^2}\Big\{\frac{\log(p \vee T)}{Tl_T},\ \frac{u_T}{l_T}\Big\}^2\\
&\le c_u\sigma^4(1+\nu^2)\frac{p}{\kappa^2}\Big\{\frac{s\log(p \vee T)}{Tl_T}\Big\} + c_u(\sigma^4 \vee \phi^2)\Big\{\frac{\xi_{2,2}u_T\sqrt{s}}{\kappa l_T}\Big\}^2\\
&\le c_u\sigma^4(1+\nu^2)\frac{p}{\kappa^2}\Big\{\frac{s\log(p \vee T)}{Tl_T}\Big\} + \frac{c_u\xi_{2,2}^2}{sT^{2k}} \le \frac{c_u\xi_{2,2}^2}{sT^{2k}}.
\end{aligned}$$

The final inequality follows from Condition A$'$(iii). Part (ii) is now a direct consequence. $\square$

Lemma B.2. *Suppose Condition A$'$, B and F hold and let* $\check{\mu}_{(j)}$ *and* $\check{\gamma}_{(j)}$, $j = 1, ..., p$ *be edge estimates of Step 1 of Algorithm 1. Additionally, let* $\log(p \vee T) \le Tv_T \le Tu_T$ *be non-negative sequences. Then,*

$$\inf_{\tau \in \mathcal{G}(u_T, v_T)}\mathcal{U}(z, \tau, \hat{\mu}, \hat{\gamma}) \ge \kappa\xi_{2,2}^2\Big[v_T - c_m\max\Big\{\Big(\frac{u_T\log(p \vee T)}{T}\Big)^{\frac{1}{2}},\ \frac{u_T}{T^k}\Big\}\Big] \tag{B.4}$$

with probability at least $1 - o(1)$. *Here* $c_m = \big\{c_u(\sigma^2 \vee \phi)\surd(1+\nu^2)\big\}\big/\big\{\kappa(1 \wedge \phi)\big\}$.

Proof of Lemma B.2. The structure of this proof is similar to that of Lemma 2.1, the distinction being the use of weaker available error bounds of the edge estimates $\check{\mu}_{(j)}$, $\check{\gamma}_{(j)}$, and sharper bounds for other stochastic terms made possible by the additional assumption $\log(p\vee T)\le Tv_T\le Tu_T$. Proceeding as in (A.2) we have that,

$$\inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\mathcal{U}(z,\tau,\check{\mu},\check{\gamma})\quad\ge\quad R1-R2-R3$$

Where $R1,R2$ and R_3 are as defined in (A.2) with $\hat{\mu}_{(j)}$, $\hat{\gamma}_{(j)}$ and $\hat{\eta}_{(j)}$ replaced with $\check{\mu}_{(j)}$, $\check{\gamma}_{(j)}$ and $\check{\eta}_{(j)}=\check{\mu}_{(j)}-\check{\gamma}_{(j)}$, $j=1,...,p$. Now applying the bounds of Lemma D.5 we obtain,

$$\begin{aligned}
R1\quad&\ge\quad\kappa\xi_{2,2}^2\Big[v_T-\frac{c_u\sigma^2}{\kappa}\Big\{\frac{u_T\log(p\vee T)}{T}\Big\}^{\frac12}-c_u(\sigma^2\vee\phi)\frac{u_T}{\kappa\xi_{2,2}}\Big(s\sum_{j=1}^p\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_2^2\Big)^{\frac12}\Big]\\
&\ge\quad\kappa\xi_{2,2}^2\Big[v_T-\frac{c_u\sigma^2}{\kappa}\Big\{\frac{u_T\log(p\vee T)}{T}\Big\}^{\frac12}-c_u(\sigma^2\vee\phi)\frac{u_T}{T^k\kappa}\Big]
\end{aligned}$$

with probability $1-o(1)$. Where the final inequality follows from Lemma B.1. Next we obtain upper bounds for the terms $R2/\kappa\xi_{2,2}^2$ and $R3/\kappa\xi_{2,2}^2$. Consider,

$$\begin{aligned}
\frac{R2}{\kappa\xi_{2,2}^2}\quad&\le\quad c_u\surd(1+\nu^2)\frac{\sigma^2\xi_{2,1}}{\kappa\xi_{2,2}^2}\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12}+c_u\surd(1+\nu^2)\frac{\sigma^2}{\kappa\xi_{2,2}^2}\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12}\sum_{j=1}^p\|\check{\eta}_{(j)}-\eta^0_{(j)}\|_1\\
&\le\quad c_u\surd(1+\nu^2)\frac{\sigma^2}{\kappa\psi}\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12}+c_u\surd(1+\nu^2)\frac{\sigma^2}{\kappa\psi}\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12}\frac{1}{T^k}\\
&\le\quad c_u\surd(1+\nu^2)\frac{\sigma^2}{\kappa\psi}\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12}
\end{aligned}\tag{B.5}$$

with probability $1-o(1)$. Here the first and second inequalities follow from Lemma D.5 and Lemma B.1, respectively. Similarly we can also obtain,

$$\begin{aligned}
\frac{R3}{\kappa\xi_{2,2}^2}\quad&\le\quad c_u(\sigma^2\vee\phi)\frac{u_T}{\kappa\xi_{2,2}}\Big\{s\sum_{j=1}^p\|\check{\gamma}_{(j)}-\gamma^0_{(j)}\|_2^2\Big\}^{\frac12}\Big[1+\frac{1}{\xi_{2,2}}\Big\{s\sum_{j=1}^p\|\check{\eta}_{(j)}-\eta^0_{(j)}\|_2^2\Big\}^{\frac12}\Big]\\
&\le\quad c_{u1}(\sigma^2\vee\phi)\frac{u_T}{\kappa T^k}
\end{aligned}$$

with probability $1-o(1)$. Substituting these bounds in (B.5) and applying a union bound over these three events yields the bound (B.4) uniformly over the set $\{\mathcal{G}(u_T,v_T);\,\tau\ge\tau^0\}$. The mirroring case of $\tau\le\tau^0$ can be obtained by similar arguments. □

Following is the proof of the main result of Section 3.

Proof of Theorem 3.1. This proof relies on the same recursive argument as that of Theorem 2.1, the distinction being that recursions are made on the bound of Lemma B.2 instead of Lemma 2.1. Consider any $Tv_T>\log(p\vee T)$, and apply Lemma B.2 on the set $\mathcal{G}(u_T,v_T)$ to obtain,

$$\begin{aligned}
\inf_{\tau\in\mathcal{G}(1,v_T)}\mathcal{U}(z,\tau,\check{\mu},\check{\gamma})\quad&\ge\quad\kappa\xi_{2,2}^2\Big[v_T-c_m\max\Big\{\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12},\frac{u_T}{T^k}\Big\}\Big]\\
&\ge\quad\kappa\xi_{2,2}^2\Big[v_T-c_m\max\Big\{\Big(\frac{\log(p\vee T)}{T}\Big)^{\frac12},\Big(u_T\frac{\log(p\vee T)}{T}\Big)^k\Big\}\Big]
\end{aligned}\tag{B.6}$$

with probability at least $1-o(1)$. Substituting $u_T = 1$, yields,

$$\inf_{\tau\in\mathcal{G}(1,v_T)} \mathcal{U}(z,\tau,\check{\mu},\check{\gamma}) \geq \kappa\xi_{2,2}^2\Big[v_T - c_m \max\Big\{\Big(\frac{\log(p\vee T)}{T}\Big)^{\frac{1}{2}}, \Big(\frac{\log(p\vee T)}{T}\Big)^{k}\Big\}\Big] \tag{B.7}$$

with probability at least $1-o(1)$. Recall that w.l.og $k < b < (1/2)$, and now choose any $v_T > v_T^* = c_m\big(\log(p\vee T)/T\big)^k$. Then we have $\inf_{\tau\in\mathcal{G}(1,v_T)}\mathcal{U}(z,\tau,\check{\mu},\check{\gamma}) > 0$, thus implying that $\hat{\tau} \notin \mathcal{G}(1,v_T)$, i.e., $\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq Tv_T^*$, with probability at least $1-o(1)$. Now reset $u_T = v_T^*$ and reapply Lemma 2.1 for any $v_T > 0$ to obtain,

$$\inf_{\tau\in\mathcal{G}(u_T,v_T)} \mathcal{U}(z,\tau,\check{\mu},\check{\gamma}) \geq \kappa\xi_{2,2}^2\Big[v_T - c_m \max\Big\{c_m^{1/2}\Big(\frac{\log(p\vee T)}{T}\Big)^{\frac{1}{2}+\frac{k}{2}}, c_m\Big(\frac{\log(p\vee T)}{T}\Big)^{k+k}\Big\}\Big]$$

Again choosing any,

$$v_T > v_T^* = \max\Big\{c_m^{g_2}\Big(\frac{\log(p\vee T)}{T}\Big)^{u_2},\ c_m^2\Big(\frac{\log(p\vee T)}{T}\Big)^{v_2}\Big\}, \tag{B.8}$$

where,

$$g_2 = 1+\frac{1}{2},\ u_2 = \frac{1}{2}+\frac{u_1}{2},\ \text{and}\ v_2 = k+v_1 \geq 2k,\ \text{with}\ u_1 = v_1 = k,$$

we obtain $\inf_{\mathcal{G}(u_T,v_T)}\mathcal{U}(z,\tau,\check{\mu},\check{\gamma}) > 0$, with probability at least $1-o(1)$. Consequently $\hat{\tau}\notin\mathcal{G}(u_T,v_T)$, i.e., $\big|\lfloor T\hat{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq Tv_T^*$. Continuing these recursions by resetting u_T to the bound of the previous recursion, and applying Lemma 2.1, we obtain for the l^{th} recursion,

$$\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq T\max\Big\{c_m^{g_l}\Big(\frac{\log(p\vee T)}{T}\Big)^{u_l},\ c_m^l\Big(\frac{\log(p\vee T)}{T}\Big)^{v_l}\Big\} := T\max\{R_{1l},R_{2l}\},\ \text{where,}$$
$$g_l = \sum_{j=0}^{l-1}\frac{1}{2^j},\ u_l = \frac{1}{2}+\frac{u_{l-1}}{2} = \frac{k}{l}+\sum_{j=1}^{l}\frac{1}{2^j},\ \text{and}\ v_l = k+v_{l-1} \geq lk,\ \text{with}\ u_1 = v_1 = k$$

Next, it is straightforward to observe that for l large enough, $R_{2l} \leq R_{1l}$, for T sufficiently large. Consequently for l large enough we have $\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq TR_{1m}$, with probability at least $1-o(1)$. Finally, we continue these recursions an infinite number of times to obtain, $g_\infty = \sum_{j=0}^{\infty} 1/2^j$, $u_\infty = \sum_{j=1}^{\infty}(1/2^j)$, thus yielding,

$$\big|\lfloor T\tilde{\tau}\rfloor - \lfloor T\tau^0\rfloor\big| \leq T\frac{c_m^2\log(p\vee T)}{T} = c_m^2\log(p\vee T)$$

with probability at least $1-o(1)$. This completes the proof of this result. □

Proof of Corollary 3.1. Under the assumed conditions, we have from Theorem 3.1 that $\hat{\tau}\in\mathcal{G}(u_T,0)$, with probability at least $1-o(1)$, where $u_T = c_m^2T^{-1}\log(p\vee T)$, where c_m is as defined in Lemma B.2. The relation of Part (i) follows directly from Theorem B.1. To obtain Part (ii), substitute this choice of u_T in λ_{2j}, $j = 1,...p$, of Theorem B.1 to obtain,

$$\lambda_{2j} = c_u(\sigma^2\vee\phi)\|\eta_{(j)}^0\|_2\max\Big\{\frac{\log(p\vee T)}{Tl_T},\ c_m^2\frac{\log(p\vee T)}{Tl_T}\Big\} \leq o(1)\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}}$$

Here the final inequality follows since by Condition A$'$(i) we have $\log(p \vee T) = o(Tl_T)$, furthermore from Lemma F.7 we have $\|\eta^0_{(j)}\|_2 \le 2\nu$, $j = 1, ..., p$. Conseuently $\lambda_{2j} \le \lambda_{1j}$, $j = 1, ..., p$, and thus applying Theorem B.1 we obtain,

$$\|\hat{\mu}_{(j)} - \mu^0_{(j)}\|_2 \le c_u \lambda_j \frac{\surd s}{\kappa} \le c_u \surd(1+\nu^2) \frac{\sigma^2}{\kappa} \Big\{ \frac{s \log(p \vee T)}{Tl_T} \Big\}^{\frac{1}{2}}$$

for all $j = 1, ..., p$, with probability at least $1 - o(1)$. Corresponding bound for $\hat{\gamma}_{(j)} - \gamma^0_{(j)}$, $j = 1, ..., p$, can be obtained using symmetrical arguments. This completes the proof of this corollary. □

Proof of Corollary 3.2. Note that Corollary 3.1 has established that the edge estimates $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$, $j = 1, ..., p$, satisfy the requirements of Condition C of Section 2. Thus, this result is now a direct consequence of Theorem 2.1 and Theorem 2.2. □

C Deviation bounds used for proofs of Section 2

Lemma C.1. *Suppose Condition B holds and let ε_{tj} be as defined in (2.4). Then, (i) the r.v. $\varepsilon_{tj} z_{t,-j,k}$ is sub-exponential with parameter $\lambda_1 = 48\sigma^2\surd(1+\nu^2)$, for each $j = 1, ..., p$, $k = 1, ..., p-1$ and $t = 1, ..., T$. (ii) The r.v. $\zeta_t = \sum_{j=1}^p \varepsilon_{tj} z^T_{t,-j}\eta^0_{(j)}$ is sub-exponential with parameter $\lambda_2 = 48\sigma^2\xi_{2,1}\surd(1+\nu^2)$, for each $t = 1, ..., T$. (iii) $E\big[|\zeta_t|^k\big] \le 4\lambda_2^k k^k$, for any $k > 0$.*

Proof of Lemma C.1. Here we only prove Part (ii) of this lemma, Part (i) follows using similar arguments, and Part (iii) follows from properties of sub-exponential random variables, see, Lemma F.2. We begin by noting that the following r.v's are mean zero, $E(\varepsilon_{tj}) = 0$, $E(z^T_{t,-j}\eta^0_{(j)}) = 0$ and $E\big(\varepsilon_{tj} z^T_{t,-j}\eta^0_{(j)}\big) = 0$. Also note that for $t \le \lfloor T\tau^0 \rfloor$, we have,

$$\varepsilon_{tj} = z_{tj} - z^T_{t,-j}\mu^0_{(j)} = \big(1, -\mu^{0T}_{(j)}\big)\big(z_{tj}, z^T_{t,-j}\big)^T$$

Using Lemma F.7 and by properties of sub-gaussian distributions we have ε_{tj}, $1 \le j \le p \sim \text{subG}(\sigma_1)$ with $\sigma_1 = \sigma\surd(1+\nu^2)$. The same also holds for ε_{tj}, for $t > \lfloor T\tau^0 \rfloor$. Similarly, $z_{t,-j}\eta^0_{(j)} \sim \text{subG}(\sigma_2)$ with $\sigma_2 = \sigma\|\eta^0_{(j)}\|_2$. Recall that if $Z \sim \text{subG}(\sigma)$, then the rescaled variable $Z/\sigma \sim \text{subG}(1)$. Next observe that,

$$\frac{\varepsilon_{tj} z^T_{t,-j}\eta^0_{(j)}}{\sigma_1\sigma_2} = \frac{1}{2}\Big\{\Phi\Big(\frac{\varepsilon_{tj}}{\sigma_1} + \frac{z^T_{t,-j}\eta^0_{(j)}}{\sigma_2}\Big) - \Phi\Big(\frac{\varepsilon_{tj}}{\sigma_1}\Big) - \Phi\Big(\frac{z^T_{t,-j}\eta^0_{(j)}}{\sigma_2}\Big)\Big\} = \frac{1}{2}[T1 - T2 - T3]$$

where $\Phi(v) = \|v\|_2^2 - E\big(\|v\|_2^2\big)$. Using Lemma F.3 and Lemma F.5 we have that $T1 \sim \text{subE}(64)$, $T2 \sim \text{subE}(16)$, and $T3 \sim \text{subE}(16)$. Applying Lemma F.4 and rescaling with σ_1, and σ_2 we obtain that $\varepsilon_{tj} z^T_{t,-j}\eta^0_{(j)} \sim \text{subE}(48\sigma_1\sigma_2)$. Another application of Lemma F.4 yields $\zeta_t = \sum_{j=1}^p \varepsilon_{tj} z^T_{t,-j}\eta^0_{(j)} \sim \text{subE}(\lambda_2)$ where

$$\lambda_2 = 48\sigma^2 \sum_{j=1}^p \|\eta^0_{(j)}\|_2 \surd(1+\nu^2) = 48\sigma^2\xi_{2,1}\surd(1+\nu^2)$$

This completes the proof of Part (ii). □

Lemma C.2. *Suppose Condition B holds and let ε_{tj} be as defined in (2.4). Additionally, let u_T, v_T be any non-negative sequences satisfying $0 \le v_T \le u_T$. Then for any $0 < a < 1$, choosing $c_{a1} = 4 \cdot 48 c_{a2}$, with $c_{a2} \ge \surd(1/a)$, we have for $T \ge 2$,*

$$\sup_{\substack{\tau \in \mathcal{G}(u_T, v_T) \\ \tau \ge \tau^0}} \frac{1}{T} \Big| \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j} \eta^0_{(j)} \Big| \le c_{a1} \sigma^2 \xi_{2,1} \surd(1+\nu^2) \Big(\frac{u_T}{T} \Big)^{\frac{1}{2}}, \tag{C.1}$$

with probability at least $1 - a$.

Proof of Lemma C.2. First note that without loss of generality we can assume $u_T \ge (1/T)$. This is because when $u_T < (1/T)$, the set $\mathcal{G}(u_T, 0)$ contains only the singleton τ^0 with a distinct value $\lfloor T\tau^0 \rfloor$. Consequently, the sum of interest in (C.1), is over indices t in an empty set, and is thus trivially zero. Now, let $\zeta_t = \sum_{j=1}^{p} \varepsilon_{tj} z^T_{t,-j} \eta^0_{(j)}$, then using Lemma C.1 we have that $\zeta_t \sim \text{subE}(\lambda)$, where $\lambda = 48 \xi_{2,1} \surd(1+\nu^2)\sigma^2$. Additionally, from part (iii) of Lemma C.1, we have, $\text{var}(\zeta_t) = E(\zeta_t)^2 \le 16\lambda^2$. Consider the set $\mathcal{G}(u_T, v_T) \cap \{\tau \ge \tau^0\}$ and note that in this set, there are at most Tu_T distinct values of $\lfloor T\tau \rfloor$. Applying Kolmogorov's inequality (Theorem F.1) with any $d > 0$ yields,

$$pr\Big(\sup_{\substack{\tau \in \mathcal{G}(u_T, v_T) \\ \tau \ge \tau^0}} \Big| \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \zeta_t \Big| > d \Big) \le \frac{1}{d^2} \sum_{\substack{t \in \mathcal{G}(u_T, v_T) \\ t \ge \tau^0}} \text{var}(\mathrm{z_t}) \le \frac{16 T u_T \lambda^2}{d^2}$$

Choosing $d = 4 c_{a2} \lambda \surd(T u_T)$, with $c_{a2} \ge \surd(1/a)$ yields the statement of the lemma. □

Lemma C.3. *Suppose Condition B holds and let ε_{tj} be as defined in (2.4) and let $0 \le v_T \le u_T$ be any non-negative sequences. Then for any $c_{u2} > 3$ and $c_{u1} \ge 96 c_{u2}$, we have for $T \ge 2$,*

$$(i) \quad \sup_{\substack{\tau \in \mathcal{G}(u_T, v_T) \\ \tau \ge \tau^0}} \frac{1}{T} \Big\| \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \varepsilon_{tj} z^T_{t,-j} \Big\|_\infty \le c_{u1} \sigma^2 \surd(1+\nu^2) \Big(\frac{u_T}{T} \Big)^{\frac{1}{2}} \log(p \vee T),$$

$$(ii) \quad \sup_{\substack{\tau \in \mathcal{G}(u_T, v_T) \\ \tau \ge \tau^0}} \frac{1}{T} \Big| \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p} \varepsilon_{tj} z^T_{t,-j} \big(\hat\eta_{(j)} - \eta^0_{(j)} \big) \Big| \le c_{u1} \sigma^2 \surd(1+\nu^2) \Big(\frac{u_T}{T} \Big)^{\frac{1}{2}} \log(p \vee T) \sum_{j=1}^{p} \| \hat\eta_{(j)} - \eta^0_{(j)} \|_1,$$

each with probability at least $1 - 2\exp\big\{ -(c_{u2} - 3)\log(p \vee T) \big\}$.[i]

Proof of Lemma C.3. Part (ii) is a direct consequence of Part (i), thus we only prove Part (i). Without loss of generality we can assume $v_T \ge (1/T)$. This follows since the only additional distinct element $\lfloor T\tau \rfloor$ in the set $\mathcal{G}(u_T, 0)$ in comparison to $\mathcal{G}(u_T, (1/T))$ is $\lfloor T\tau^0 \rfloor$, and at this value, the sum of interest is over indices t in an empty set and is thus trivially zero.

Let $z_{t,-j} = (z_{t,-j,1}, ..., z_{t,-j,p-1})^T$, then from Lemma C.1 we have $\varepsilon_{tj} z_{t,-j,k} \sim \text{subE}(\lambda_1)$, with $\lambda_1 = 48\surd(1+\nu^2)\sigma^2$. Now applying Bernstein's inequality (Lemma F.6) for any fixed $\tau \in \mathcal{G}(u_T, v_T)$ satisfying $\tau \ge \tau^0$, we have for any $d > 0$,

$$pr\Big(\Big| \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \varepsilon_{tj} z_{t,-j,k} \Big| > d\big(\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor \big) \Big) \le 2 \exp\Big\{ - \frac{\big(\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor \big)}{2} \Big(\frac{d^2}{\lambda_1^2} \wedge \frac{d}{\lambda_1} \Big) \Big\}$$

[i]Here $\big\| \sum \varepsilon_{tj} z^T_{t,-j} \big\|_\infty = \max_{j,k} \big| \sum \varepsilon_{tj} z_{t,-j,k} \big|$.

Choose $d = 2c_{u2}\lambda_1 \log(p \vee T)\big/\sqrt{\big(\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor\big)}$, then,

$$\begin{aligned}\big(\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor\big)\frac{d^2}{2\lambda_1^2} &= 2c_{u2}^2 \log^2(p \vee T), \quad \text{and,}\\ \big(\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor\big)\frac{d}{2\lambda_1} &= c_{u2} \log(p \vee T),\end{aligned}$$

where we have used $\big(\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor\big) \geq Tv_T \geq 1$. A substitution back in the probability bound yields,

$$\Big|\sum_{t=\lfloor T\tau^0 \rfloor+1}^{\lfloor T\tau \rfloor} \varepsilon_{tj} z_{t,-j,k}\Big| \leq 2c_{u2}\lambda_1 \log(p \vee T)\big(\lfloor T\tau \rfloor - \lfloor T\tau^0 \rfloor\big)^{1/2} \leq 2c_{u2}\lambda_1 \log(p \vee T)\big(Tu_T\big)^{\frac{1}{2}},$$

with probability at least $1 - 2\exp\{-c_{u2}\log(p \vee T)\}$. Finally applying a union bound over $j = 1, ..., p$, $k = 1, ..., p-1$ and over the at most T distinct values of $\lfloor T\tau \rfloor$ for $\tau \in \mathcal{G}(u_T, v_T)$, we obtain,

$$\sup_{\substack{\tau \in \mathcal{G}(u_T,v_T)\\ \tau \geq \tau^0}} \Big\|\frac{1}{T}\sum_{t=\lfloor T\tau^0 \rfloor+1}^{\lfloor T\tau \rfloor} \varepsilon_{tj} z_{t,-j,k}\Big\|_\infty \leq 2c_{u2}\lambda_1 \log(p \vee T)\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}},$$

with probability at least $1 - 2\exp\{-(c_{u2} - 3)\log(p \vee T)\}$. This completes the proof of Part (i). $\square$

Lemma C.4. *Suppose Condition B holds and let u_T, v_T be any non-negative sequences satisfying $0 \leq v_T \leq u_T$. Then for any $0 < a < 1$, choosing $c_{a1} = 64c_{a2}$, with $c_{a2} \geq \sqrt{(1/a)}$, we have for $T \geq 2$,*

$$(i) \inf_{\substack{\tau \in \mathcal{G}(u_T,v_T);\\ \tau \geq \tau^0}} \frac{1}{T}\sum_{t=\lfloor T\tau^0 \rfloor+1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p} \eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T \eta_{(j)}^0 \geq v_T \kappa \xi_{2,2}^2 - c_{a1}\sigma^2 \xi_{2,2}^2 \Big(\frac{u_T}{T}\Big)^{\frac{1}{2}},$$

$$(ii) \sup_{\substack{\tau \in \mathcal{G}(u_T,v_T);\\ \tau \geq \tau^0}} \frac{1}{T}\sum_{t=\lfloor T\tau^0 \rfloor+1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p} \eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T \eta_{(j)}^0 \leq u_T \phi \xi_{2,2}^2 + c_{a1}\sigma^2 \xi_{2,2}^2 \Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}$$

with probability at least $1 - a$.

Proof of Lemma C.4. As before in Lemma C.2, w.l.o.g we assume $u_T \geq (1/T)$. Now, we have $\eta_{(j)}^{0T} z_{t,-j} \sim \text{subG}\big(\sigma\|\eta_{(j)}^0\|_2\big)$, consequently, using Lemma F.5 and Lemma F.4 we have

$$\sum_{j=1}^{p}\Big(\|\eta_{(j)}^{0T} z_{t,-j}\|_2^2 - E\|\eta_{(j)}^{0T} z_{t,-j}\|_2^2\Big) \sim \text{subE}\big(\lambda\big), \quad \text{with} \quad \lambda = 16\sigma^2\xi_{2,2}^2.$$

Using moment properties of sub-exponential distributions $\big($Part (iii) of Lemma C.1$\big)$ we also have that

$$\text{var}\Big\{\sum_{j=1}^{p}\Big(\|\eta_{(j)}^{0T} z_{t,-j}\|_2^2 - E\|\eta_{(j)}^{0T} z_{t,-j}\|_2^2\Big)\Big\} \leq 16\lambda^2.$$

Now applying Kolmogorov's inequality (Lemma F.1) we obtain,

$$pr\left\{\sup_{\substack{\tau \in \mathcal{G}(u_T,v_T);\\ \tau \geq \tau^0}} \Big|\sum_{t=\lfloor T\tau^0 \rfloor+1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p}\Big(\|\eta_{(j)}^{0T} z_{t,-j}\|_2^2 - E\|\eta_{(j)}^{0T} z_{t,-j}\|_2^2\Big)\Big| > d\right\} \leq \frac{16\lambda^2 T u_T}{d^2}.$$

Choosing $d = 4c_{a2}\lambda\surd(Tu_T)$, with $c_{a2} \geq \surd(1/a)$ yields,

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\Big(\|\eta_{(j)}^{0T}z_{t,-j}\|_2^2 - E\|\eta_{(j)}^{0T}z_{t,-j}\|_2^2\Big)\Big| \leq 4c_{a2}\lambda\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}$$

with probability at least $1-a$. The statement of this lemma is now a direct consequence. $\square$

We require additional notation for the following results. Consider any sequence of $\alpha_{(j)}, \psi_{(j)} \in \mathbb{R}^{p-1}$, $j = 1, ..., p$, and let α, ψ represent the concatenation of all $\alpha_{(j)}$'s and $\psi_{(j)}$'s. Then define

$$\Phi(\alpha,\psi) = \frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\alpha_{(j)}^T z_{t,-j} z_{t,-j}\psi_{(j)} \tag{C.2}$$

Lemma C.5. *Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.2) and suppose Condition B and C(ii) hold. Let u_T, v_T be any non-negative sequences satisfying $0 \leq v_T \leq u_T$. Then for any $0 < a < 1$, choosing $c_{a1} = 64c_{a2}$, with $c_{a2} \geq \surd(1/a)$, we have for $T \geq 2$,*

$$(i) \quad \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\eta^0,\eta^0) \geq v_T\kappa\xi_{2,2}^2 - c_{a1}\sigma^2\xi_{2,2}^2\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}$$

$$(ii) \quad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\hat\eta-\eta^0,\hat\eta-\eta^0) \leq c_u(\sigma^2\vee\phi)s\log(p\vee T)u_T\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2$$

with probability at least $1-a$, and $1-o(1)$, respectively. Moreover, when $u_T \geq c_{a1}^2\sigma^4/T\phi^2$, we have,

$$(iii) \quad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\eta^0,\eta^0) \leq 2u_T\phi\xi_{2,2}^2,$$

$$(iv) \quad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \big|\Phi(\hat\eta-\eta^0,\eta^0)\big| \leq c_u(\sigma^2\vee\phi)u_T\xi_{2,2}\Big\{s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big\}^{\frac{1}{2}},$$

with probability at least $1-a$, and $1-a-o(1)$, respectively.

Proof of Lemma C.5. Part (i) and Part (iii) of this lemma are a direct consequence of Lemma C.4. To prove Part (ii), first note that,

$$\begin{aligned}\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_1^2 &\leq 2\sum_{j=1}^{p}\Big(\|\hat\mu_{(j)}-\mu_{(j)}^0\|_1^2 + \|\hat\gamma_{(j)}-\gamma_{(j)}^0\|_1^2\Big)\\ &\leq 32s\sum_{j=1}^{p}\Big(\|\hat\mu_{(j)}-\mu_{(j)}^0\|_2^2 + \|\hat\gamma_{(j)}-\gamma_{(j)}^0\|_2^2\Big) \leq 32s\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2, \quad \text{(C.3)}\end{aligned}$$

with probability at least $1-\pi_T = 1-o(1)$. Here the second inequality follows since by Condition C(ii) we have, $\hat\mu_{(j)}-\mu_{(j)}^0 \in \mathcal{A}_{1j}$, and $\hat\gamma_{(j)}-\gamma_{(j)}^0 \in \mathcal{A}_{2j}$, $j = 1, ..., p$. Now applying Lemma E.2, we have,

$$\begin{aligned}\sup_{\tau\in\mathcal{G}(u_T,v_T)} \Phi(\hat\eta-\eta^0,\hat\eta-\eta^0) &\leq c_u(\sigma^2\vee\phi)\log(p\vee T)u_T\Big(\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta^0\|_2^2 + \sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta^0\|_1^2\Big)\\ &\leq c_u(\sigma^2\vee\phi)s\log(p\vee T)u_T\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\end{aligned}$$

with probability at least $1-o(1)$. Here the final inequality follows by using (C.3). The proof of Part (iv) is an application of the Cauchy-Schwarz inequality together with the bounds of Part (ii) and Part (iii),

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\big|\Phi(\hat\eta-\eta^0,\eta^0)\big| \;\le\; \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\big\{\Phi\big(\hat\eta-\eta^0,\hat\eta-\eta^0\big)\big\}^{\frac12}\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\big\{\Phi\big(\eta^0,\eta^0\big)\big\}^{\frac12}.$$

This completes the proof of this lemma. □

Lemma C.6. *Suppose Condition B and C(ii) hold. Let u_T, v_T be any non-negative sequences satisfying $0\le v_T\le u_T$. Then for any $0<a<1$, choosing $c_{a1}=4\cdot 48c_{a2}$, with $c_{a2}\ge\surd(1/a)$, and for $u_T\ge c_{a1}^2\sigma^4\big/(T\phi^2)$, we have for $T\ge 2$,*

$$(i)\quad \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\big(\hat\eta_{(j)}^T z_{t,-j}\big)^2\ge$$

$$\kappa\xi_{2,2}^2\Big[v_T-\frac{c_{a1}\sigma^2}{\kappa}\Big(\frac{u_T}{T}\Big)^{\frac12}-c_u(\sigma^2\vee\phi)\frac{u_T}{\kappa\xi_{2,2}}\Big\{s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big\}^{\frac12}\Big]$$

$$(ii)\quad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(\hat\gamma_{(j)}-\gamma_{(j)}^0)^T z_{t,-j}z_{t,-j}^T\hat\eta_{(j)}\Big|\le$$

$$c_u(\sigma^2\vee\phi)\xi_{2,2}u_T\Big\{s\log(p\vee T)\sum_{j=1}^{p}\|\hat\gamma_{(j)}-\gamma_{(j)}^0\|_2^2\Big\}^{\frac12}\Big[1+\frac{1}{\xi_{2,2}}\Big\{s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big\}^{\frac12}\Big]$$

$$(iii)\quad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\hat\eta_{(j)}\Big|\le$$

$$c_{a1}\surd(1+\nu^2)\sigma^2\xi_{2,1}\Big(\frac{u_T}{T}\Big)^{\frac12}+c_u\surd(1+\nu^2)\sigma^2\Big(\frac{u_T}{T}\Big)^{\frac12}\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_1,$$

each with probability at least $1-a-o(1)$.

Proof of Lemma C.6. Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.2). Then note that $\Phi(\hat\eta,\hat\eta)=\Phi(\eta^0,\eta^0)+2\Phi(\hat\eta-\eta^0,\eta^0)+\Phi(\hat\eta-\eta^0,\hat\eta-\eta^0)$. Using this relation together with the bounds of Part (i) and Part (iv) of Lemma C.5 we obtain,

$$\begin{aligned}\inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\Phi(\hat\eta,\hat\eta) &\ge \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\Phi(\eta^0,\eta^0)-2\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}|\Phi(\hat\eta-\eta^0,\eta^0)|\\ &\ge v_T\kappa\xi_{2,2}^2-c_{a1}\sigma^2\xi_{2,2}^2\Big(\frac{u_T}{T}\Big)^{\frac12}-c_u(\sigma^2\vee\phi)u_T\xi_{2,2}\Big(s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big)^{\frac12}\end{aligned}$$

with probability at least $1-a-o(1)$. To prove Part (ii), note that using identical arguments as in the proof of Lemma C.5 it can be shown that,

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\hat\gamma-\gamma^0,\hat\gamma-\gamma^0)\leq c_u(\sigma^2\vee\phi)s\log(p\vee T)u_T\sum_{j=1}^p\|\hat\gamma_{(j)}-\gamma^0_{(j)}\|_2^2,$$

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \big|\Phi(\hat\gamma-\gamma^0,\eta^0)\big|\leq c_u(\sigma^2\vee\phi)u_T\xi_{2,2}\Big\{s\log(\vee T)\sum_{j=1}^p\|\hat\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}},$$

with probability at least $1-a-o(1)$. The above inequalities and the relation $\Phi\big(\hat\gamma-\gamma^0,\hat\eta\big)\leq\big|\Phi(\hat\gamma-\gamma^0,\hat\eta-\eta^0)\big|+\big|\Phi(\hat\gamma-\gamma^0,\eta^0)\big|$, together with applications of the Cauchy-Schwarz inequality yields,

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \big|\Phi\big(\hat\gamma-\gamma^0,\hat\eta\big)\big| &\leq c_u(\sigma^2\vee\phi)s\log(p\vee T)u_T\Big(\sum_{j=1}^p\|\hat\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big)^{\frac{1}{2}}\Big(\sum_{j=1}^p\|\hat\eta_{(j)}-\eta^0_{(j)}\|_2^2\Big)^{\frac{1}{2}}\\
&\quad+c_u(\sigma^2\vee\phi)u_T\xi_{2,2}\Big\{s\log(p\vee T)\sum_{j=1}^p\|\hat\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}}\\
&\leq c_u(\sigma^2\vee\phi)\xi_{2,2}u_T\Big\{s\log(p\vee T)\sum_{j=1}^p\|\hat\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}}\Big[1+\frac{1}{\xi_{2,2}}\Big\{s\log(p\vee T)\sum_{j=1}^p\|\hat\eta_{(j)}-\eta^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}}\Big]
\end{aligned}$$

with probability at least $1-a-o(1)$. To prove Part (iii), note that,

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\varepsilon_{tj}z_{t,-j}^T\hat\eta_{(j)}\Big| &\leq \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\varepsilon_{tj}z_{t,-j}^T\eta^0_{(j)}\Big|\\
&\quad+\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\varepsilon_{tj}z_{t,-j}^T(\hat\eta_{(j)}-\eta^0_{(j)})\Big|\\
&:= R1+R2.
\end{aligned}$$

Now using Lemma C.2 we have for any $0<a<1$, $R1\leq c_{a1}\sqrt{(1+\nu^2)}\sigma^2\xi_{2,1}\big(u_T/T\big)^{1/2}$, with probability at least $1-a$. Also, using Lemma C.3 we have,

$$R2\leq c_u\sqrt{(1+\nu^2)}\sigma^2\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\log(p\vee T)\sum_{j=1}^p\|\hat\eta_{(j)}-\eta^0_{(j)}\|_1$$

with probability at least $1-o(1)$. Part (iv) now follows by combining bounds for terms $R1$ and $R2$. □

Lemma C.7. *Suppose Condition A and C hold. Then we have,*

$$(i)\quad \sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_2^2 \le c_u(1+\nu^2)\frac{\sigma^4}{\kappa^2}\Big\{\frac{sp\log(p\vee T)}{Tl_T}\Big\},$$

$$(ii)\quad \sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_1 \le c_u\surd(1+\nu^2)\frac{\sigma^2 sp}{\kappa}\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}}$$

$$(iii)\quad \frac{1}{\xi_{2,2}}\Big(s\log(p\vee T)\sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_2^2\Big)^{\frac{1}{2}} \le \frac{c_{u1}}{T^b}=o(1),$$

$$(iv)\quad \frac{1}{\xi_{2,2}^2}\sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_1 \le \frac{c_{u1}}{\psi}\Big\{\frac{1}{\log(p\vee T)}\Big\}^{\frac{1}{2}}$$

with probability at least $1-o(1)$.

Proof of Lemma C.7. Part (i) can be obtained as,

$$\sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_2^2 \le 2\sum_{j=1}^{p}\Big(\|\hat{\mu}_{(j)}-\mu^0_{(j)}\|_2^2+\|\hat{\gamma}_{(j)}-\gamma^0_{(j)}\|_2^2\Big) \le c_u(1+\nu^2)\frac{\sigma^4}{\kappa^2}\Big\{\frac{sp\log(p\vee T)}{Tl_T}\Big\},$$

with probability at least $1-o(1)$. Here the final inequality follows from (1.4). Part (ii) can be obtained quite analogously. To prove Part (iii) note that from Condition A we have $(1/\xi_{2,2})=(1/\psi\surd p)$ and consider,

$$\begin{aligned}\frac{1}{\xi_{2,2}}\Big(s\log(p\vee T)\sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_2^2\Big)^{\frac{1}{2}} &\le \frac{1}{\psi}\Big(sp^{-1}\log(p\vee T)\sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_2^2\Big)^{\frac{1}{2}}\\ &\le c_u\surd(1+\nu^2)\frac{\sigma^2}{\psi\kappa}\Big\{\frac{s\log(p\vee T)}{\surd(Tl_T)}\Big\}\le\frac{c_{u1}}{T^b},\end{aligned}$$

with probability at least $1-o(1)$. Here the second inequality follows by using the bound of Part (i) and the second follows from Condition A. To prove Part (iv) consider,

$$\begin{aligned}\frac{1}{\xi_{2,2}^2}\sum_{j=1}^{p}\|\hat{\eta}_{(j)}-\eta^0_{(j)}\|_1 &\le c_u\surd(1+\nu^2)\frac{\sigma^2 s}{\psi^2\kappa}\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}}\\ &\le \Big\{\frac{1}{\psi\log(p\vee T)}\Big\}c_u\surd(1+\nu^2)\frac{\sigma^2}{\psi\kappa}\Big\{\frac{s\log(p\vee T)}{\surd(Tl_T)}\Big\}\\ &\le \frac{c_{u1}}{\psi}\Big\{\frac{1}{\log(p\vee T)}\Big\}^{\frac{1}{2}}\end{aligned}$$

with probability at least $1-o(1)$. Here the first inequality follows by the assumption $(1/\xi_{2,2})=(1/\psi\surd p)$ together with the bound in Part (ii). The final inequality follows from Condition A. □

Lemma C.8. *Suppose the Conditions of Theorem 2.2 and let* R_{11},R_{12},R_{13}, *and* R_{21},R_{22} *be as defined in its proof. Let* $0<c_1<\infty$ *be any constant, then we have the following bounds.*

$$(i)\quad \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}}|R_{11}-R_{21}|=o(1),\qquad (ii)\quad \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}}|R_{12}-R_{22}|=o(1)$$

$$(iii)\quad \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}}|R_{13}|=o(1)$$

each with probability at least $1-o(1)$.

Proof of Lemma C.8. Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.2) and consider,

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} |R_{11}-R_{21}| &= \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} p^{-1}\Big|\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\Big(\|z_{t,-j}^T\hat\eta_{(j)}\|_2^2-\|z_{t,-j}^T\eta_{(j)}^0\|_2^2\Big)\Big|\\
&= \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} p^{-1}\Big|\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(\hat\eta_{(j)}-\eta_{(j)}^0)^Tz_{t,-j}z_{t,-j}^T(\hat\eta_{(j)}+\eta_{(j)}^0)\Big|\\
&= \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} \Big|Tp^{-1}\Phi(\hat\eta-\eta^0,\hat\eta-\eta^0)+2Tp^{-1}\Phi(\hat\eta-\eta^0,\eta^0)\Big|. \qquad \text{(C.4)}
\end{aligned}$$

Now from Part (ii) of Lemma C.5 we have

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} Tp^{-1}\Phi(\hat\eta-\eta^0,\hat\eta-\eta^0) &\le c_uc_1(\sigma^2\vee\phi)\psi^{-2}p^{-1}s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\\
&= O\Big(\frac{s^2\log^2(p\vee T)}{\psi^{-2}Tl_T}\Big)=o(1), \qquad \text{(C.5)}
\end{aligned}$$

with probability at least $1-o(1)$. Also, from Part (iv) of Lemma C.5, we have for $u_T\ge c_{a1}^2\sigma_x^4/(T\phi^2)$,

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,0);\\ \tau\ge\tau^0}} 2Tp^{-1}\big|\Phi(\hat\eta-\eta^0,\eta^0)\big|\le c_u(\sigma^2\vee\phi)Tu_Tp^{-1}\xi_{2,2}\Big(s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big)^{\frac12} \qquad \text{(C.6)}$$

with probability at least $1-a-o(1)$. Upon choosing $a=\big(64^2\psi^2\sigma^4\big)/(c_1\phi^2)\to0$, we have $c_1T^{-1}\psi^{-2}=c_{a1}^2\sigma_x^4/(T\phi^2)$, consequently from (C.6) we have,

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} 2T\big|\Phi(\hat\eta-\eta^0,\eta^0)\big| &\le c_uc_1(\sigma^2\vee\phi)\frac{\xi_{2,2}}{p\psi^2}\Big(s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big)^{\frac12}\\
&= c_uc_1(\sigma^2\vee\phi)\frac{1}{\xi_{2,2}}\Big(s\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big)^{\frac12}\\
&\le O\Big(\frac{1}{\psi}\frac{s\log(p\vee T)}{\surd(Tl_T)}\Big)=o(1) \qquad \text{(C.7)}
\end{aligned}$$

with probability at least $1-a-o(1)=1-o(1)$. Substituting this uniform bound together with (C.5) back in (C.4) yields Part (i) of this lemma. To prove Part (ii), note that

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} |R_{12}-R_{22}| &= \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} p^{-1}\Big|\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T(\hat\eta_{(j)}-\eta_{(j)}^0)\Big|\\
&= O\Big(p^{-1}\psi^{-1}\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)}-\eta_{(j)}^0\|_1\Big)\\
&\le O\Big(\frac{s\log^{3/2}(p\vee T)}{\psi\surd(Tl_T)}\Big)=o(1),
\end{aligned}$$

with probability at least $1-o(1)$. Here the second equality follows from Part (ii) of Lemma C.3. To prove Part (iii) we first note that the expressions $\Phi\big(\hat\gamma-\gamma^0,\hat\eta-\eta^0\big)$, and $\Phi(\hat\gamma-\gamma^0,\eta^0)$ can be bounded above with probability at least $1-o(1)$, by the same bounds as in (C.5) and (C.7), respectively. Now applications of the Cauchy-Schwarz inequality yields the following bound for the term $|R_{13}|$.

$$\begin{aligned}\sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}}|R_{13}| &= \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}}\Big|\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}(\hat\gamma_{(j)}-\gamma^0_{(j)})^Tz_{t,-j}z_{t,-j}^T\hat\eta_{(j)}\Big|\\ &\le \sup_{\substack{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big);\\ \tau\ge\tau^0}} T\Big\{\big|\Phi(\hat\gamma-\gamma^0,\hat\eta-\eta^0)\big|+\big|\Phi(\hat\gamma-\gamma^0,\eta^0)\big|\Big\}=o(1),\end{aligned}$$

with probability at least $1-o(1)$, thus completing the proof of this lemma. $\square$

Lemma C.9. *Suppose Condition B holds and that $\psi\to0$. Then for any constant any $r>0$, we have,*

$$p^{-1}\Big|\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^0+r\psi^{-2}\rfloor}\sum_{j=1}^{p}\Big(\|z_{t,-j}^T\eta^0_{(j)}\|_2^2-E\|z_{t,-j}^T\eta^0_{(j)}\|_2^2\Big)\Big|=o_p(1)$$

Additionally, if $\xi_{2,2}^{-2}\sum_{j=1}^p E\|z_{t,-j}^T\eta^0_{(j)}\|_2^2\to\sigma^*$, then,

$$p^{-1}\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^0+r\psi^{-2}\rfloor}\sum_{j=1}^{p}\big\|z_{t,-j}^T\eta^0_{(j)}\big\|_2^2\to_p r\sigma^*.$$

Proof of Lemma C.9. We begin with the following observation. For any $\tau\ge\tau^0$, we have the deterministic inequality $T(\tau-\tau^0)-1\le\big(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor\big)\le T(\tau-\tau^0)+1$. It is straightforward to verify that under the assumption $\psi\to0$, this inequality directly yields $c_{u1}r\psi^{-2}\le\big(\lfloor T\tau^0+r\xi^{-2}\rfloor-\lfloor T\tau^0\rfloor\big)\le c_{u2}r\psi^{-2}$. Also, note that from Lemma F.4 and Lemma F.5 we have,

$$p^{-1}\psi^{-2}\sum_{j=1}^{p}\Big(\big\|z_{t,-j}^T\eta^0_{(j)}\big\|_2^2-E\big\|z_{t,-j}^T\eta^0_{(j)}\big\|_2^2\Big)\sim\text{subE}(\lambda),\quad\text{with}\quad\lambda=16\sigma^2. \tag{C.8}$$

Now upon applying Bernstein's inequality (Lemma F.6) together with the above observations, we obtain for any $d>0$,

$$pr\Big\{p^{-1}\Big|\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^0+r\psi^{-2}\rfloor}\sum_{j=1}^{p}\Big(\big\|z_{t,-j}^T\eta^0_{(j)}\big\|_2^2-E\big\|z_{t,-j}^T\eta^0_{(j)}\big\|_2^2\Big)\Big|>c_{u2}dr\Big\}\le2\exp\Big\{-\frac{c_{u1}r\psi^{-2}}{2}\Big(\frac{d^2}{\lambda^2}\wedge\frac{d}{\lambda}\Big)\Big\}.$$

Choosing d as any sequence converging to zero slower than ψ, say $d=\psi^{1-b}$, for any $0<b<1$, and noting that in this case $(d/\lambda)\le1$ for T large, we obtain,

$$p^{-1}\Big|\sum_{\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau^0+r\psi^{-2}\rfloor}\sum_{j=1}^{p}\Big(\big\|z_{t,-j}^T\eta^0_{(j)}\big\|_2^2-E\big\|z_{t,-j}^T\eta^0_{(j)}\big\|_2^2\Big)\Big|=o_p(1),$$

This completes the proof of the first part of this lemma, the second part can be obtained as a direct consequence of Part (i). $\square$

D Deviation bounds used for proofs of Section 3

Lemma D.1. *Suppose Condition A′(i), A′(ii) and B holds, and $c_{u1} > 0$ be any constant. Then uniformly over $j = 1, ..., p$, we have,*

$$\sup_{\substack{\tau\in(0,1);\\ \lfloor T\tau\rfloor\ge c_{u1}Tl_T}} \frac{1}{\lfloor T\tau\rfloor}\Big\|\sum_{t=1}^{\lfloor T\tau\rfloor}\varepsilon_{tj}z_{t,-j}\Big\|_\infty \le 48\sigma^2(c_u/\surd c_{u1})\surd(1+\nu^2)\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac12}$$

with probability at least $1-2\exp\big[-\{(c_u^2/2)-3\}\log(p\vee T)\big]$. *Additionally, let $u_T \ge 0$, be any sequence and $c_u > 0$ any constant, then uniformly over $j = 1, ..., p$, we have,*

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,0);\\ \lfloor T\tau\rfloor\ge c_{u1}Tl_T}} \frac{1}{\lfloor T\tau\rfloor}\Big\|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\Big\|_\infty \quad\le\quad c_{u2}(\sigma^2\vee\phi)\|\eta_{(j)}^0\|_2\max\Big\{\frac{\log(p\vee T)}{Tl_T},\ \frac{u_T}{l_T}\Big\},$$

with probability $1-2\exp\big\{-c_{u3}\log(p\vee T)\big\}$, *with* $c_{u2} = (1+48c_u)/c_{u1}$, $c_{u3} = \{(c_u\wedge c_u^2)/2\}-3$.

Proof of Lemma D.1. We begin with proving Part (i). Using Lemma C.1 we have that $\varepsilon_{tj}z_{t,-j,k} \sim$ subE(λ_1), with $\lambda_1 = 48\sigma^2\surd(1+\nu^2)$. For any $\tau\in(0,1)$ satisfying $\lfloor T\tau\rfloor \ge c_{u1}Tl_T$, applying Lemma F.6 we have for any $d > 0$,

$$pr\Big(\Big|\sum_{t=1}^{\lfloor T\tau\rfloor}\varepsilon_{tj}z_{t,-j,k}\Big| > d\lfloor T\tau\rfloor\Big) \le 2\exp\Big\{-\frac{\lfloor T\tau\rfloor}{2}\Big(\frac{d^2}{\lambda_1^2}\wedge\frac{d}{\lambda_1}\Big)\Big\}.$$

Choose $d = c_u\lambda_1\surd\{\log(p\vee T)\big/\lfloor T\tau\rfloor\}$, and recall that by choice we have $\lfloor T\tau\rfloor \ge c_{u1}Tl_T$, and from Condition A′(i) we have $\log(p\vee T) \le c_{u1}Tl_T$. Thus, $d/\lambda_1 \le 1$, and consequently $(d^2/\lambda_1^2) \le (d/\lambda_1)$. Using these relations the above probability bound yields,

$$\frac{1}{\lfloor T\tau\rfloor}\Big|\sum_{t=1}^{\lfloor T\tau\rfloor}\varepsilon_{tj}z_{t,-j,k}\Big| \le (c_u/\surd c_{u1})\lambda_1\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac12}$$

with probability at least $1-2\exp\big\{-(c_u^2/2)\log(p\vee T)\big\}$. Part (i) now follows by applying a union bound over $k = 1, ..., (p-1)$, $j = 1, ..., p$ and over the at most T distinct values of $\lfloor T\tau\rfloor$.

To prove Part (ii), first note that using similar arguments as in Lemma C.1 we have that $\eta_{(j)}^{0T}z_{t,-j}z_{t,-j,k} - E\big(\eta_{(j)}^{0T}z_{t,-j}z_{t,-j,k}\big) \sim$ subE(λ_1), with $\lambda_1 = 48\sigma^2\|\eta_{(j)}^0\|_2$. For any $\tau\in\mathcal{G}(u_T,0)$, satisfying $\lfloor T\tau\rfloor \ge c_{u1}Tl_T$, applying a union bound over $k = 1, ..., p-1$, on the Bernstein's inequality (Lemma C.1) yields the following probability bound,

$$pr\Big\{\Big\|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\big(\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}-\eta_{(j)}^{0T}\Delta_{-j,-j}\big)\Big\|_\infty > d(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)\Big\} \le$$

$$2p\exp\Big\{-\frac{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}{2}\Big(\frac{d^2}{\lambda_1^2}\wedge\frac{d}{\lambda_1}\Big)\Big\} \qquad \text{(D.1)}$$

Now upon choosing,

$$d = c_u\lambda_1\max\Big[\Big\{\frac{\log(p\vee T)}{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}\Big\}^{\frac12},\ \frac{\log(p\vee T)}{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}\Big],$$

it can be verified that [j],

$$d\frac{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}{\lfloor T\tau\rfloor}\le\frac{c_u}{c_{u1}}\lambda_1\max\Big\{\frac{\log(p\vee T)}{Tl_T},\ \frac{u_T}{l_T}\Big\},\quad\text{and,}$$
$$\frac{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}{2}\Big(\frac{d^2}{\lambda_1^2}\wedge\frac{d}{\lambda_1}\Big)=\frac{(c_u\wedge c_u^2)}{2}\log(p\vee T)\tag{D.2}$$

Substituting the relations of (D.2) in the probability bound (D.1) we obtain,

$$\frac{1}{\lfloor T\tau\rfloor}\Big\|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\big(\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}-\eta_{(j)}^{0T}\Delta_{-j,-j}\big)\Big\|_\infty\le\frac{c_u}{c_{u1}}\lambda_1\max\Big\{\frac{\log(p\vee T)}{Tl_T},\ \frac{u_T}{l_T}\Big\}$$

with probability at least $1-2p\exp\big[\{(c_u\wedge c_u^2)/2\}\log(p\vee T)\big]$. Next, using the bounded eigenvalue assumption of Condition B we have that,

$$\frac{1}{\lfloor T\tau\rfloor}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\eta_{(j)}^{0T}\Delta_{-j,-j}\le\|\eta_{(j)}^0\|_2\phi\frac{u_T}{c_{u1}l_T}$$

Using this relation in the probability bound now yields,

$$\frac{1}{\lfloor T\tau\rfloor}\Big\|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}\Big\|_\infty\le c_{u2}\phi\|\eta_{(j)}^0\|_2\frac{u_T}{l_T}+c_{u3}\sigma^2\|\eta_{(j)}^0\|_2\max\Big\{\frac{\log(p\vee T)}{Tl_T},\ \frac{u_T}{l_T}\Big\},$$

with probability at least $1-2p\exp\big[\{(c_u\wedge c_u^2)/2\}\log(p\vee T)\big]$, where $c_{u2}=1/c_{u1}$, and $c_{u3}=48c_u/c_{u1}$. Uniformity over τ can be obtained by using a union bound over the atmost T distinct values of $\lfloor T\tau\rfloor$, and similarly over $j=1,...,p$, by using another union bound. This completes the proof of the lemma. □

Remark D.1. *Consider,*

$$d=c_u\lambda_1\max\Big[\Big\{\frac{\log(p\vee T)}{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}\Big\}^{\frac{1}{2}},\ \frac{\log(p\vee T)}{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}\Big],\tag{D.3}$$

observe that when $\log(p\vee T)/(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)\ge1$, *then the maximum of the two terms in the expression (D.3) is* $\log(p\vee T)/(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)$. *In this case,*

$$\Big(\frac{d^2}{\lambda_1^2}\wedge\frac{d}{\lambda_1}\Big)=(c_u^2\wedge c_u)\frac{\log(p\vee T)}{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}.\tag{D.4}$$

In the case where $\log(p\vee T)/(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)<1$, *the maximum in the expression* (D.3) *becomes* $\sqrt{\{\log(p\vee T)/(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)\}}$, *however the minimum in the expression (D.4) remains the same.*

Lemma D.2. *Suppose Condition B holds and let* ε_{tj} *be as defined in (2.4). Let* $T\ge\log(p\vee T)$ *and* $\log(p\vee T)\le Tv_T\le Tu_T$ *be any non-negative sequences. Then for any* $c_u>0$, *we have,*

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T)\\ \tau\ge\tau^0}}\frac{1}{T}\Big\|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\varepsilon_{tj}z_{t,-j}^T\Big\|_\infty\le48\surd(2c_u)\sigma^2\surd(1+\nu^2)\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac{1}{2}},$$

with probability at least $1-2\exp\big\{-(c_{u1}-3)\log(p\vee T)\big\}$, *with* $c_{u1}=c_u\wedge\surd(c_u/2)$.

[j]See, Remark D.1

Proof of Lemma D.2. The proof of this result is very similar to that of Lemma C.3, the difference being utilization of the additional assumption $Tv_T \geq \log(p \vee T)$, in order to obtain this sharper bound. Proceeding as in (C.2) we have,

$$pr\Big(\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor} \varepsilon_{tj} z_{t,-j,k}\Big| > d\big(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor\big)\Big) \leq 2\exp\Big\{-\frac{\big(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor\big)}{2}\Big(\frac{d^2}{\lambda_1^2} \wedge \frac{d}{\lambda_1}\Big)\Big\},$$

where $\lambda_1 = 48\sigma^2\surd(1+\nu^2)$. Choose $d = \lambda_1\{2c_u \log(p \vee T)/(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor)\}^{1/2}$, then,

$$\begin{aligned}
\big(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor\big)\frac{d^2}{2\lambda_1^2} &= c_u \log(p \vee T), \quad \text{and,} \\
\big(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor\big)\frac{d}{2\lambda_1} &= \surd(c_u/2)\{\log(p \vee T)(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor)\}^{1/2} \geq \surd(c_u/2)\log(p \vee T),
\end{aligned}$$

where we used $\big(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor\big) \geq Tv_T \geq \log(p \vee T)$. Substituting back in the probability bound yields,

$$\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor} \varepsilon_{tj} z_{t,-j,k}\Big| \leq \lambda_1\Big\{\frac{2c_u u_T \log(p \vee T)}{T}\Big\}^{1/2},$$

with probability at least $1 - 2\exp\{-c_{u1}\log(p \vee T)\}$, with $c_{u1} = c_u \wedge \surd(c_u/2)$. Finally applying a union bound over $j = 1, ..., p$, $k = 1, ..., p-1$ and over the at most T distinct values of $\lfloor T\tau\rfloor$ for $\tau \in \mathcal{G}(u_T, v_T)$, yields the statement of this lemma. □

Lemma D.3. *Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.2) and suppose Condition B holds and $T \geq \log(p \vee T)$. Additionally, let u_T, v_T be non-negative sequences satisfying $\log(p \vee T) \leq Tv_T \leq Tu_T$. Then for any constant $c_u > 0$, we have,*

$$(i) \qquad \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\eta^0, \eta^0) \geq v_T\kappa\xi_{2,2}^2 - 16\surd(2c_u)\sigma^2\xi_{2,2}^2\Big(\frac{u_T\log(p \vee T)}{T}\Big)^{\frac{1}{2}},$$

$$(ii) \qquad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\eta^0, \eta^0) \leq u_T\phi\xi_{2,2}^2 + 16\surd(2c_u)\sigma^2\xi_{2,2}^2\Big(\frac{u_T\log(p \vee T)}{T}\Big)^{\frac{1}{2}}$$

with probability at least $1 - 2\exp\{-(c_{u1}-1)\log(p \vee T)\}$, where $c_{u1} = c_u \wedge \surd(c_u/2)$.

Proof of Lemma D.3. Note that $\sum_{j=1}^{p}\big(\|z_{t,-j}^T\eta_{(j)}^0\|_2^2 - E\|z_{t,-j}^T\eta_{(j)}^0\|_2^2\big) \sim \text{subE}(\lambda)$, where $\lambda = 16\sigma^2\xi_{2,2}^2$. For any fixed $\tau \in \mathcal{G}(u_T, v_T)$, applying the Bernstein's inequality (Lemma F.6) we obtain,

$$\begin{aligned}
pr\Big\{\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\big(\|z_{t,-j}^T\eta_{(j)}^0\|_2^2 - E\|z_{t,-j}^T\eta_{(j)}^0\|_2^2\big)\Big| \geq d(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor)\Big\}& \\
\leq 2\exp\Big\{-\frac{(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor)}{2}\Big(\frac{d^2}{\lambda^2} \wedge \frac{d}{\lambda}\Big)\Big\}&
\end{aligned}$$

Choose $d = \lambda\{2c_u \log(p \vee T)/(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor)\}^{1/2}$ and observe that,

$$\begin{aligned}
(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor)\frac{d^2}{2\lambda^2} &= c_u \log(p \vee T) \\
(\lfloor T\tau\rfloor - \lfloor T\tau^0\rfloor)\frac{d}{2\lambda} &= \surd(c_u/2)\{Tv_T\log(p \vee T)\}^{1/2} \geq \surd(c_u/2)\log(p \vee T)
\end{aligned}$$

where the inequality follows from the assumption $Tv_T \ge \log(p\vee T)$. A substitution back in the above probability bound yields,

$$\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\big(\|z_{t,-j}^T\eta_{(j)}^0\|_2^2 - E\|z_{t,-j}^T\eta_{(j)}^0\|_2^2\big)\Big| \le \surd(2c_u)\lambda\Big\{\frac{u_T\log(p\vee T)}{T}\Big\}^{\frac{1}{2}} \tag{D.5}$$

with probability at least $1-2\exp\big(-c_{u1}\log(p\vee T)\big)$, $c_{u1}=c_u\wedge\surd(c_u/2)$. Applying a union bound over at most T distinct values of $\lfloor T\tau\rfloor$, yields the bound (D.5) uniformly over τ. The statements of this lemma are now a direct consequence. $\square$

Lemma D.4. *Let $\Phi(\cdot,\cdot)$ be as defined in (C.2) and suppose Condition B holds and $T\ge\log(p\vee T)$. Let $\check\mu_{(j)}$ and $\check\gamma_{(j)}$, $j=1,...,p$ be Step 1 edge estimates of Algorithm 1, and u_T,v_T be any non-negative sequences satisfying $\log(p\vee T)\le Tv_T\le Tu_T$. Then,*

$$(i)\qquad \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\Phi(\eta^0,\eta^0)\ge v_T\kappa\xi_{2,2}^2 - c_u\sigma^2\xi_{2,2}^2\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac{1}{2}},$$

$$(ii)\qquad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\Phi(\check\eta-\eta^0,\check\eta-\eta^0)\le c_u(\sigma^2\vee\phi)u_T\Big(s\sum_{j=1}^p\|\check\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big)$$

with probability $1-o(1)$. Furthermore, when $u_T\ge c_u\sigma^4\log(p\vee T)\big/T\phi^2$, we have,

$$(iii)\qquad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\Phi(\eta^0,\eta^0)\le 2u_T\phi\xi_{2,2}^2,$$

$$(iv)\qquad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\big|\Phi(\check\eta-\eta^0,\eta^0)\big|\le c_u(\sigma^2\vee\phi)u_T\xi_{2,2}\Big\{s\sum_{j=1}^p\|\check\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big\}^{\frac{1}{2}},$$

with probability at least $1-o(1)$.

Proof of Lemma D.4. Part (i) and Part (iii) are a direct consequence of Lemma D.3. To prove Part (ii), first note from Theorem B.1 we have that $\check\mu_{(j)}-\mu_{(j)}^0\in\mathcal{A}_{1j}$, and $\check\gamma_{(j)}-\gamma_{(j)}^0\in\mathcal{A}_{2j}$, $j=1,...,p$, with probability at least $1-o(1)$. It can be verified that this property yields $\|\check\eta_{(j)}-\eta_{(j)}^0\|_1\le c_u\surd s\|\check\eta_{(j)}-\eta_{(j)}^0\|_2$. (see, e.g. (C.3)). Now applying Part (ii) of E.2 yields,

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\Phi(\check\eta-\eta^0,\check\eta-\eta^0)\le c_u(\sigma^2\vee\phi)u_T\Big(s\sum_{j=1}^p\|\check\eta_{(j)}-\eta_{(j)}^0\|_2^2\Big)$$

with probability at least $1-o(1)$. Part (iv) follows by an application of the Cauchy-Schwarz inequality together with the bounds of Part (ii) and Part (iii) (see, (C.4)). This completes the proof of this lemma. $\square$

Lemma D.5. *Suppose Condition B holds and* $T \geq \log(p \vee T)$. *Let* $\check{\mu}_{(j)}$, $\check{\gamma}_{(j)}$, $j = 1, ..., p$ *be Step 1 estimates of Algorithm 1, and assume* u_T, v_T *satisfy* $\log(p \vee T) \leq Tv_T \leq Tu_T$. *Then,*

$$(i) \inf_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} \frac{1}{T} \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p} \|\check{\eta}_{(j)}^T z_{t,-j}\|_2^2 \geq$$

$$\kappa \xi_{2,2}^2 \Big[v_T - \frac{c_u \sigma^2}{\kappa} \Big\{ \frac{u_T \log(p \vee T)}{T} \Big\}^{\frac{1}{2}} - c_u(\sigma^2 \vee \phi) \frac{u_T}{\kappa \xi_{2,2}} \Big(s \sum_{j=1}^{p} \|\hat{\eta}_{(j)} - \eta_{(j)}^0\|_2^2 \Big)^{\frac{1}{2}} \Big]$$

$$(ii) \sup_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} \frac{1}{T} \Big| \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p} (\check{\gamma}_{(j)} - \gamma_{(j)}^0)^T z_{t,-j} z_{t,-j}^T \check{\eta}_{(j)} \Big| \leq$$

$$c_u(\sigma^2 \vee \phi) \xi_{2,2} u_T \Big\{ s \sum_{j=1}^{p} \|\check{\gamma}_{(j)} - \gamma_{(j)}^0\|_2^2 \Big\}^{\frac{1}{2}} \Big[1 + \frac{1}{\xi_{2,2}} \Big\{ s \sum_{j=1}^{p} \|\check{\eta}_{(j)} - \eta_{(j)}^0\|_2^2 \Big\}^{\frac{1}{2}} \Big]$$

$$(iii) \sup_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} \frac{1}{T} \Big| \sum_{t=\lfloor T\tau^0 \rfloor + 1}^{\lfloor T\tau \rfloor} \sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j}^T \check{\eta}_{(j)} \Big| \leq$$

$$c_u \surd(1 + \nu^2) \sigma^2 \xi_{2,1} \Big(\frac{u_T \log(p \vee T)}{T} \Big)^{\frac{1}{2}} + c_u \surd(1 + \nu^2) \sigma^2 \Big(\frac{u_T \log(p \vee T)}{T} \Big)^{\frac{1}{2}} \sum_{j=1}^{p} \|\check{\eta}_{(j)} - \eta_{(j)}^0\|_1,$$

each with probability at least $1 - o(1)$.

Proof of Lemma D.5. Let $\Phi(\cdot\,, \cdot)$ be as defined in (C.2). Then note that $\Phi(\check{\eta}, \check{\eta}) = \Phi(\eta^0, \eta^0) + 2\Phi(\check{\eta} - \eta^0, \eta^0) + \Phi(\check{\eta} - \eta^0, \check{\eta} - \eta^0)$. Using this relation together with the bounds of Part (i) and Part (iv) of Lemma D.4 we obtain,

$$\begin{aligned} \inf_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} \Phi(\check{\eta}, \check{\eta}) &\geq \inf_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} \Phi(\eta^0, \eta^0) - 2 \sup_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} |\Phi(\check{\eta} - \eta^0, \eta^0)| \\ &\geq v_T \kappa \xi_{2,2}^2 - c_u \sigma^2 \xi_{2,2}^2 \Big\{ \frac{u_T \log(p \vee T)}{T} \Big\}^{\frac{1}{2}} - c_u(\sigma^2 \vee \phi) u_T \xi_{2,2} \Big(s \sum_{j=1}^{p} \|\check{\eta}_{(j)} - \eta_{(j)}^0\|_2^2 \Big)^{\frac{1}{2}} \end{aligned}$$

with probability at least $1 - o(1)$. To prove Part (ii), note that using identical arguments as in the proof of Lemma D.4 it can be shown that,

$$\sup_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} \Phi(\check{\gamma} - \gamma^0, \check{\gamma} - \gamma^0) \leq c_u(\sigma^2 \vee \phi) u_T s \sum_{j=1}^{p} \|\check{\gamma}_{(j)} - \gamma_{(j)}^0\|_2^2,$$

$$\sup_{\substack{\tau \in \mathcal{G}(u_T, v_T); \\ \tau \geq \tau^0}} \big|\Phi(\check{\gamma} - \gamma^0, \eta^0)\big| \leq c_u(\sigma^2 \vee \phi) u_T \xi_{2,2} \Big\{ s \sum_{j=1}^{p} \|\check{\gamma}_{(j)} - \gamma_{(j)}^0\|_2^2 \Big\}^{\frac{1}{2}},$$

with probability at least $1 - o(1)$. The above inequalities and the relation $\Phi\big(\check{\gamma} - \gamma^0, \check{\eta}\big) \leq \big|\Phi(\check{\gamma} -$

$\gamma^0, \check\eta - \eta^0)\big| + \big|\Phi(\check\gamma - \gamma^0, \eta^0)\big|$, together with applications of the Cauchy-Schwarz inequality yields,

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \big|\Phi\big(\check\gamma-\gamma^0,\check\eta\big)\big| &\leq c_u(\sigma^2\vee\phi)u_T\Big(s\sum_{j=1}^p\|\check\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big)^{\frac12}\Big(s\sum_{j=1}^p\|\check\eta_{(j)}-\eta^0_{(j)}\|_2^2\Big)^{\frac12}\\
&\quad + c_u(\sigma^2\vee\phi)u_T\xi_{2,2}\Big\{s\sum_{j=1}^p\|\check\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big\}^{\frac12}\\
&\leq c_u(\sigma^2\vee\phi)\xi_{2,2}u_T\Big\{s\sum_{j=1}^p\|\check\gamma_{(j)}-\gamma^0_{(j)}\|_2^2\Big\}^{\frac12}\Big[1+\frac{1}{\xi_{2,2}}\Big\{s\sum_{j=1}^p\|\check\eta_{(j)}-\eta^0_{(j)}\|_2^2\Big\}^{\frac12}\Big]
\end{aligned}$$

with probability at least $1-o(1)$. To prove Part (iii), note that,

$$\begin{aligned}
\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac1T\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\varepsilon_{tj}z_{t,-j}^T\check\eta_{(j)}\Big| &\leq \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac1T\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\varepsilon_{tj}z_{t,-j}^T\eta^0_{(j)}\Big|\\
&\quad + \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac1T\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\varepsilon_{tj}z_{t,-j}^T(\check\eta_{(j)}-\eta^0_{(j)})\Big|\\
&:= R1+R2.
\end{aligned}$$

Now using Lemma D.2, we have

$$\begin{aligned}
R1 &\leq c_u\sqrt{(1+\nu^2)}\sigma^2\xi_{2,1}\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12},\ \text{and}\\
R2 &\leq c_u\sqrt{(1+\nu^2)}\sigma^2\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac12}\sum_{j=1}^p\|\check\eta_{(j)}-\eta^0_{(j)}\|_1 \qquad \text{(D.6)}
\end{aligned}$$

with probability at least $1-o(1)$. Part (iv) now follows by combining bounds for terms $R1$ and $R2$. □

E Uniform (over τ) Restricted Eigenvalue Condition

Lemma E.1. *Let $z_t\in\mathbb{R}^p$, $t=1,...,n$ be independent* subG(σ) *r.v's and $\lambda=16\sigma^2$. Additionally, for any $s\geq1$, let $\mathcal{K}_p(s)=\{\delta\in\mathbb{R}^p;\ \|\delta\|_1\leq1,\ \|\delta\|_0\leq s\}$. Then for non-negative $0\leq v_T\leq u_T$, and any $d_1>0$, we have $T\geq2$,*

$$pr\Bigg[\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\sup_{\delta\in\mathcal{K}_p(2s)}\frac1T\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\big\{\|z_t^T\delta\|_2^2-E\|z_t^T\delta\|_2^2\big\}\Big|\geq d_1u_T\Bigg]\leq 2\exp\Big\{-\frac{Tv_T}{2}\Big(\frac{d_1^2}{\lambda^2}\wedge\frac{d_1}{\lambda}\Big)+3s\log(p\vee T)\Big\}$$

Proof of Lemma E.1. Consider any fixed $\delta\in\mathbb{R}^p$, with $\|\delta\|_2\leq1$, then from Lemma F.5 we have $\|z_t^T\delta\|_2^2-E\|z_t^T\delta\|_2^2\sim$ subE(λ), with $\lambda=16\sigma^2$. Now, for any fixed $\tau\in\mathcal{G}(u_T,v_T)$, $\tau\geq\tau^0$ applying

Lemma F.6 (Bernstein's inequality) we have,

$$pr\Big(\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\|z_t^T\delta\|_2^2-E\|z_t^T\delta\|_2^2\Big|>d(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)\Big)\le 2\exp\Big\{-\frac{(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)}{2}\Big(\frac{d^2}{\lambda^2}\wedge\frac{d}{\lambda}\Big)\Big\}$$

Choose $d=d_1Tu_T/(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)$ and observe that by definition of the set $\mathcal{G}(u_T,v_T)$, we have $Tv_T\le(\lfloor T\tau\rfloor-\lfloor T\tau^0\rfloor)\le Tu_T$, this in turn yields $d_1\le d$, and consequently,

$$pr\Big(\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\|z_t^T\delta\|_2^2-E\|z_t^T\delta\|_2^2\Big|\ge d_1u_T\Big)\le 2\exp\Big\{-\frac{Tv_T}{2}\Big(\frac{d_1^2}{\lambda^2}\wedge\frac{d_1}{\lambda}\Big)\Big\} \tag{E.1}$$

Using the inequality (E.1) and a covering number argument, it can be shown that (see, Lemma 15 of the supplementary materials of Loh and Wainwright [2012]) for any $s\ge1$,

$$pr\Big(\sup_{\delta\in\mathcal{K}_p(2s)}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\|z_t^T\delta\|_2^2-E\|z_t^T\delta\|_2^2\Big|\ge d_1u_T\Big)\le 2\exp\Big\{-\frac{Tv_T}{2}\Big(\frac{d_1^2}{\lambda^2}\wedge\frac{d_1}{\lambda}\Big)+2s\log(p\vee T)\Big\}.$$

Finally, uniformity over the set $\mathcal{G}(u_T,v_T)$ can be obtained by applying a union bound over the at most T distinct values of $\lfloor T\tau\rfloor$ for $\tau\in\mathcal{G}(u_T,v_T)$, thus yielding the statement of this lemma. □

Lemma E.2. *Suppose Condition B holds and let $0\le v_T\le u_T$ be any non-negative sequences. Then for all $\delta_{(j)}\in\mathbb{R}^{p-1}$, $j=1,...,p$, and $T\ge2$, we have,*

$$(i)\quad\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\delta_{(j)}^Tz_{t,-j}z_{t,-j}^T\delta_{(j)}\le c_u(\sigma^2\vee\phi)u_T\log(p\vee T)\Big(\sum_{j=1}^p\|\delta_{(j)}\|_2^2+\sum_{j=1}^p\|\delta_{(j)}\|_1^2\Big)$$

with probability at least $1-2\exp\big\{-\log(p\vee T)\big\}$. Additionally assuming that $T\ge\log(p\vee T)$ and v_T satisfies $Tv_T\ge\log(p\vee T)$, then for all $\delta_{(j)}\in\mathbb{R}^{p-1}$, $j=1,...,p$,

$$(ii)\quad\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\frac{1}{T}\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\delta_{(j)}^Tz_{t,-j}z_{t,-j}^T\delta_{(j)}\le c_u(\sigma^2\vee\phi)u_T\Big(\sum_{j=1}^p\|\delta_{(j)}\|_2^2+\sum_{j=1}^p\|\delta_{(j)}\|_1^2\Big)$$

with probability at least $1-2\exp\big\{-\log(p\vee T)\big\}$.

Proof. w.l.o.g. assume $v_T\ge(1/T)$ (see, Lemma C.3). Now for any $s\ge1$, consider any non-negative u_T, any $\delta_{(j)}\in\mathcal{K}_{p-1}(2s)$, $j=1,...,p$. Then for any $d_1>0$, applying a union bound to the result of Lemma E.1 over the components $j=1,...,p$ we obtain,

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\ \sup_{\substack{\delta_{(j)}\in\mathcal{K}(2s);\\ j=1,...,p}}\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\|z_{t,-j}^T\delta_{(j)}\|_2^2-E\|z_{t,-j}^T\delta_{(j)}\|_2^2\Big|\le d_1u_T \tag{E.2}$$

with probability at least $1-2\exp\Big\{-\frac{Tv_T}{2}\Big(\frac{d_1^2}{\lambda^2}\wedge\frac{d_1}{\lambda}\Big)+4s\log(p\vee T)\Big\}$. It can be shown that the bound (E.2) in turn implies that (see, Lemma 12 of supplement of Loh and Wainwright [2012]), for all $\tau\in\mathcal{G}(u_T,v_T)$, and for all $\delta_{(j)}\in\mathbb{R}^{p-1}$, $j=1,...,p$,

$$\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^p\|z_{t,-j}^T\delta_{(j)}\|_2^2-E\|z_{t,-j}^T\delta_{(j)}\|_2^2\Big|\le 27d_1u_T(\sum_{j=1}^p\|\delta_{(j)}\|_2^2+(1/s)\sum_{j=1}^p\|\delta_{(j)}\|_1^2)$$

with probability at least $1-2\exp\Big\{-\frac{Tv_T}{2}\Big(\frac{d_1^2}{\lambda^2}\wedge\frac{d_1}{\lambda}\Big)+4s\log(p\vee T)\Big\}$. Now choose $d_1=10\lambda\log(p\vee T)$, and note that $\frac{Tv_T}{2}\Big(\frac{d_1^2}{\lambda^2}\wedge\frac{d_1}{\lambda}\Big)\geq 5\log(p\vee T)$. This follows since $Tv_T\geq 1$, and that $d_1/\lambda\geq 1$. A substitution back in the probability bound yields,

$$\frac{1}{T}\Big|\sum_{t=\lfloor T\tau^0\rfloor+1}^{\lfloor T\tau\rfloor}\sum_{j=1}^{p}\|z_{t,-j}^T\delta_{(j)}\|_2^2-E\|z_{t,-j}^T\delta_{(j)}\|_2^2\Big|\leq 270\lambda u_T\log(p\vee T)\Big\{\sum_{j=1}^{p}\|\delta_{(j)}\|_2^2+\frac{1}{s}\sum_{j=1}^{p}\|\delta_{(j)}\|_1^2\Big\},$$

with probability at least $1-2\exp\big\{-5\log(p\vee T)+4s\log(p\vee T)\big\}$. The statement of Part (i) follows upon setting $s=1$. The proof of Part (ii) is quite analogous. This can be obtained by proceeding as earlier with (E.2) above, and additionally utilizing $Tv_T\geq\log(p\vee T)$, and setting $d_1=10\lambda$, instead of the choice made for Part (i). This completes the proof of this result. $\square$

Lemma E.3. *Suppose Condition A' and B hold, then for $i=1,2$, we have,*

$$\min_{j=1,\dots,p;}\inf_{\substack{\tau\in(0,1);\\ \tau\geq c_{u1}l_T}}\inf_{\substack{\delta\in\mathcal{A}_{ij};\\ \|\delta\|_2=1}}\frac{1}{\lfloor T\tau\rfloor}\sum_{t=1}^{\lfloor T\tau\rfloor}\delta^T z_{t,-j}z_{t,-j}^T\delta\geq\frac{\kappa}{2}.$$

with probability at least $1-2\exp\{-c_u\log(p\vee T)\}$, for some $c_u>0$ and for T sufficiently large.

Lemma E.3 is a nearly direct extension of the usual restricted eigenvalue condition. Its proof is analogous to those available in the literature, for e.g., Corollary 1 of Loh and Wainwright [2012]. In comparison to the typical restricted eigenvalue condition, Lemma E.3 has additional uniformity over τ, i and j, which can be obtained by simply using additional union bounds.

F Auxiliary results

In the following Definition's **??**, **??**, and Lemma's F.1-F.6, we provide basic properties of subgaussian and subexponential distributions. These are largely reproduced from Vershynin [2019] and Rigollet [2015]. Theorem F.1 and F.2 below reproduce the Kolmogorov's inequality and the argmax theorem. Lemma F.7 provides an upper bound for the ℓ_2 norm of the parameter vectors defined in Section 1.

Sub-gaussian r.v.: A random variable $X\in\mathbb{R}$ is said to be sub-gaussian with parameter $\sigma>0$ $\big($denoted by $X\sim\text{subG}(\sigma)\big)$ if $E(X)=0$ and its moment generating function

$$E(\mathrm{e}^{tX})\leq\mathrm{e}^{t^2\sigma^2/2},\qquad\forall\ t\in\mathbb{R}$$

Furthermore, a random vector $X\in\mathbb{R}^p$ is said to be sub-gaussian with parameter σ, if the inner products $\langle X,v\rangle\sim\text{subG}(\sigma)$ for any $v\in\mathbb{R}^p$ with $\|v\|_2=1$.

Sub-exponential r.v.: A random variable $X\in\mathbb{R}$ is said to be sub-exponential with parameter $\sigma>0$ $\big($denoted by $X\sim\text{subE}(\sigma)\big)$ if $E(X)=0$ and its moment generating function

$$E(\mathrm{e}^{tX})\leq\mathrm{e}^{t^2\sigma^2/2},\qquad\forall\ |t|\leq\frac{1}{\sigma}$$

Lemma F.1. *[Tail bounds] (i) If $X \sim \mathrm{subG}(\sigma)$, then,*

$$pr(|X| \geq \lambda) \leq 2\exp(-\lambda^2/2\sigma^2).$$

(ii) If $X \sim \mathrm{subE}(\sigma)$, then

$$pr(|X| \geq \lambda) \leq 2\exp\Big\{-\frac{1}{2}\Big(\frac{\lambda^2}{\sigma^2} \wedge \frac{\lambda}{\sigma}\Big)\Big\}.$$

Proof of Lemma F.1 . This proof is a simple application of the Markov inequality. For any $t > 0$,

$$pr(X \geq \lambda) = pr(tX \geq t\lambda) \leq \frac{E\mathrm{e}^{tX}}{\mathrm{e}^{t\lambda}} = \mathrm{e}^{-t\lambda + t^2\sigma^2/2}.$$

Minimizing over $t > 0$, yields the choice $t^* = \lambda/\sigma^2$, and substituting in the above bound ,we obtain,

$$pr(X \geq \lambda) \leq \inf_{t>0} \mathrm{e}^{-t\lambda + t^2\sigma^2/2} = e^{-\lambda^2/2\sigma^2}.$$

Repeating the same for $P(X \leq -\lambda)$ yields part (i) of the lemma. To prove Part (ii), repeat the above argument with $t \in (0, 1/\sigma]$, to obtain,

$$pr(X \geq \lambda) = pr(tX \geq t\lambda) \leq \mathrm{e}^{-t\lambda + t^2\sigma^2/2}. \tag{F.1}$$

As in the subgaussian case, to obtain the tightest bound one needs to find t^* that minimizes $-t\lambda + t^2\sigma^2/2$, with the additional constraint for this subexponential case that $t \in (0, 1/\sigma]$. We know that the unconstrained minimum occurs at $t^* = \lambda/\sigma^2 > 0$. Now consider two cases:

1. If $t^* < (0, 1/\sigma] \Leftrightarrow \lambda \leq \sigma$ then the unconstrained minimum is same as the constrained minimum, and substituting this value yields the same tail behavior as the subgaussian case.

2. If $t^* > (1/\sigma) \Leftrightarrow \lambda > \sigma$, then note that $-t\lambda + t^2\sigma^2/2$ is decreasing in t, in the interval $(0, (1/\sigma)]$, thus the minimum occurs at the boundary $t = 1/\sigma$. Substituting in the tail bound we obtain for this case,

   $$pr(X \geq \lambda) \leq \mathrm{e}^{-t\lambda + t^2\sigma^2/2} = \exp\{-(\lambda/\sigma) + (1/2)\} \leq \exp(-\lambda/2\sigma),$$

   where the final inequality follows since $\lambda > \sigma$.

Part (ii) of the lemma is obtained by combining the results of the above two cases. □

Lemma F.2 (Moment bounds). *(i) If $X \sim \mathrm{subG}(\sigma)$, then*

$$E|X|^k \leq 3k\sigma^k k^{k/2}, \qquad k \geq 1.$$

(ii) If $X \sim \mathrm{subE}(\sigma)$, then

$$E|X|^k \leq 4\sigma^k k^k, \qquad k > 0.$$

Proof of Lemma F.2. Consider $X \sim \mathrm{subG}(\sigma)$, and w.l.o.g assume that $\sigma = 1$ (else define $X^* = X/si$). Using the integrated tail probability expectation formula, we have for any $k > 0$,

$$\begin{aligned} E|X|^k &= \int_0^\infty pr(|X|^k > t)dt = \int_0^\infty pr(|X| > t^{1/k})dt \\ &\leq 2\int_0^\infty \exp\Big(-\frac{t^{2/k}}{2}\Big)dt \\ &= 2^{k/2}k\int_0^\infty \mathrm{e}^{-u}u^{k/2-1}du, \qquad u = \frac{t^{2/k}}{2} \\ &= 2^{k/2}k\Gamma(k/2) \end{aligned}$$

Here the first inequality follows from the tail bound Lemma F.1. Now, for $x \geq 1/2$, we have the inequality $\Gamma(x) \leq 3x^x$, thus for $k \geq 1$ we have, $\Gamma(k/2) \leq 3(k/2)^{(k/2)}$. A substitution back in the moment bound yields desired bound of Part (i).

To prove the moment bound of Part (ii). As before, w.l.o.g. assume $\sigma = 1$. Consider the inequality,

$$|x|^k \leq k^k(e^x + e^{-x})$$

which is valid for all $x \in \mathbb{R}$ and $k > 0$. Substitute x=X and take expectation to get,

$$E|X|^k \leq k^k(E\mathrm{e}^X + E\mathrm{e}^{-X}).$$

Since in this case $\sigma = 1$, from the mgf condition, at $t = \pm 1$ we have, $E\mathrm{e}^X \leq \mathrm{e}^{1/2} \leq 2$, and $E\mathrm{e}^{-X} \leq 2$. Thus for any $k > 0$,

$$E|X|^k \leq 4k^k$$

This yields the desired moment bound of Part (ii). □

Lemma F.3. *Assume that* $X \sim \mathrm{subG}(\sigma)$, *and that* $\alpha \in \mathbb{R}$, *then* $\alpha X \sim \mathrm{subG}(|\alpha|\sigma)$. *Moreover if* $X_1 \sim \mathrm{subG}(\sigma_1)$ *and* $X_2 \sim \mathrm{subG}(\sigma_2)$, *then* $X_1 + X_2 \sim \mathrm{subG}(\sigma_1 + \sigma_2)$.

Proof of Lemma F.3. The first part follows directly from the inequality $E(\mathrm{e}^{t\alpha X}) \leq \exp(t^2\alpha^2\sigma^2/2)$. To prove Part (ii) use the Hölder's inequality to obtain,

$$\begin{aligned} E(\mathrm{e}^{t(X_1+X_2)}) &= E(\mathrm{e}^{tX_1}\mathrm{e}^{tX_2}) \leq \{E(\mathrm{e}^{tX_1p})\}^{\frac{1}{p}}\{E(\mathrm{e}^{tX_2q})\}^{\frac{1}{q}} \\ &\leq \mathrm{e}^{\frac{t^2}{2}\sigma_1^2p^2}\mathrm{e}^{\frac{t^2}{2}\sigma_2^2q^2} = \mathrm{e}^{\frac{t^2}{2}(p\sigma_1^2+q\sigma_2^2)} \end{aligned}$$

where $p, q \in [1, \infty]$, with $1/p + 1/q = 1$. Choose $p^* = (\sigma_2/\sigma_1) + 1$, $q^* = (\sigma_1/\sigma_2) + 1$ to obtain $E(\mathrm{e}^{t(X_1+X_2)}) \leq \exp\big\{\frac{t^2}{2}(\sigma_1 + \sigma_2)^2\big\}$. This completes the proof of this lemma. □

Lemma F.4. *Assume that* $X \sim \mathrm{subE}(\sigma)$, *and that* $\alpha \in \mathbb{R}$, *then* $\alpha X \sim \mathrm{subE}(|\alpha|\sigma)$. *Moreover, assume that* $X_1 \sim \mathrm{subE}(\sigma_1)$ *and* $X_2 \sim \mathrm{subE}(\sigma_2)$, *then* $X_1 + X_2 \sim \mathrm{subE}(\sigma_1 + \sigma_2)$.

The proof of Lemma F.4 is analogous to that of Lemma F.3 and is thus omitted.

Lemma F.5 (Lemma 1.12 of Rigollet [2015]). *Let* $X \sim \text{subG}(\sigma)$ *then the random variable* $Z = X^2 - E[X^2]$ *is sub-exponential:* $Z \sim \text{subE}(16\sigma^2)$.

The next result is the Bernstein's inequality, reproduced from Lemma 1.13 of Rigollet [2015].

Lemma F.6 (Bernstein's inequality). *Let* $X_1, X_2, ..., X_T$ *be independent random variables such that* $X_t \sim \text{subE}(\sigma)$. *Then for any* $d > 0$ *we have,*

$$pr(|\bar{X}| > d) \leq 2\exp\Big\{-\frac{T}{2}\Big(\frac{d^2}{\sigma^2} \wedge \frac{d}{\sigma}\Big)\Big\}$$

The next result is the Kolmogorov's inequality reproduced from Hájek and Rényi [1955]

Theorem F.1 (Kolmogorov's inequality). *If* $\xi_1, \xi_2, ...$ *is a sequence of mutually independent random variables with mean values* $E(\xi_k) = 0$ *and finite variance* $\text{var}(\xi_k) = D_k^2$ $(k = 1, 2, ...)$, *we have, for any* $\varepsilon > 0$,

$$pr\Big(\max_{1\leq k\leq m}\big|\xi_1 + \xi_2 + ... + \xi_k\big| > \varepsilon\Big) \leq \frac{1}{\varepsilon^2}\sum_{k=1}^{m} D_k^2$$

The following theorem is the well known 'Argmax' theorem reproduced from Theorem 3.2.2 of Vaart and Wellner [1996]

Theorem F.2 (Argmax Theorem). *Let* $\mathcal{M}_n, \mathcal{M}$ *be stochastic processes indexed by a metric space* H *such that* $\mathcal{M}_n \Rightarrow \mathcal{M}$ *in* $\ell^\infty(K)$ *for every compact set* $K \subseteq H$[k]. *Suppose that almost all sample paths* $h \to \mathcal{M}(h)$ *are upper semicontinuous and posses a unique maximum at a (random) point* $\hat{h}$, *which as a random map in* H *is tight. If the sequence* $\hat{h}_n$ *is uniformly tight and satisfies* $\mathcal{M}_n(\hat{h}_n) \geq \sup_h \mathcal{M}_n(h) - o_p(1)$, *then* $\hat{h}_n \Rightarrow \hat{h}$ *in* H.

Lemma F.7. *Suppose condition B holds, and let* $\mu^0_{(j)}$ *and* $\gamma^0_{(j)}$, *be as defined in (1.2). Then we have,*

$$\max_{1\leq j\leq p}\Big(\|\mu^0_{(j)}\|_2 \vee \|\gamma^0_{(j)}\|_2\Big) \leq \nu,$$

Proof of Lemma F.7. Let $\Omega = \Sigma^{-1}$ be the precision matrix corresponding to Σ. Then we can write $\Omega_{jj} = -(\Sigma_{jj} - \Sigma_{j,-j}\mu^0_{(j)})^{-1}$, and $\Omega_{-j,j} = -\Omega_{jj}\mu^0_{(j)}$, for each $j = 1, ..., p$, (see, e.g., Yuan [2010]). We also have that $1/\phi \leq \max_j |\Omega_{jj}| \leq 1/\kappa$. Now note that the ℓ_2 norm of the rows (or columns) of Ω are bounded above, i.e., $\|\Omega_{j\cdot}\|_2 = \|\Omega e_j\|_2 \leq 1/\kappa$. This finally implies that

$$\|\mu^0_{(j)}\|_2 = \|-\Omega_{-j,j}/\Omega_{jj}\|_2 \leq \|\Omega_{j\cdot}\|_2/|\Omega_{jj}| \leq \frac{\phi}{\kappa} = \nu \tag{F.2}$$

Since the r.h.s. in (F.2) is free of j, this implies that $\max_j \|\mu^0_{(j)}\| \leq \nu$. Identical arguments can be used to show that $\max_j \|\gamma^0_{(j)}\| \leq \nu$. These two statements together imply the statement of the lemma. □

[k] i.e., $\sup_{h\in K}\big|\mathcal{M}_n(h) - \mathcal{M}(h)\big| \to^p 0$.

G Additional numerical results

T	p	bias ($\hat{\tau}$) ($\times 10^2$)	bias ($\tilde{\tau}$) ($\times 10^2$)	rmse ($\hat{\tau}$) ($\times 10^2$)	rmse ($\tilde{\tau}$) ($\times 10^2$)	SE ($\lfloor T\tilde{\tau} \rfloor$)	c_α
200	100	5.28	3.06	9.56	7.67	2.27	10.22
200	200	11.77	7.52	15.53	12.53	2.27	10.22
200	300	13.07	8.21	16.70	13.51	2.26	10.22
200	400	13.70	8.73	17.56	14.22	2.27	10.22
275	100	1.18	0.20	2.93	0.53	2.27	10.22
275	200	1.05	0.30	2.63	0.74	2.27	10.22
275	300	3.34	1.09	7.03	3.62	2.26	10.22
275	400	2.81	0.85	6.45	3.02	2.27	10.22
350	100	0.16	0.06	0.61	0.28	2.27	10.22
350	200	0.17	0.04	0.74	0.29	2.27	10.22
350	300	0.28	0.02	0.92	0.41	2.26	10.22
350	400	0.27	0.05	1.01	0.35	2.27	10.22
425	100	0.30	0.10	0.67	0.17	2.27	10.22
425	200	0.20	0.11	0.37	0.20	2.27	10.22
425	300	0.19	0.11	0.37	0.21	2.26	10.22
425	400	0.18	0.12	0.32	0.20	2.27	10.22

Table 3: Summary of numerical results at $\tau^0 = 0.69$. The estimates $\hat{\tau}$, and $\tilde{\tau}$ are those of Step 1 and Step 2 of Algorithm 1, respectively.

T	p	bias ($\hat{\tau}$) ($\times 10^2$)	bias ($\tilde{\tau}$) ($\times 10^2$)	rmse ($\hat{\tau}$) ($\times 10^2$)	rmse ($\tilde{\tau}$) ($\times 10^2$)	SE ($\lfloor T\tilde{\tau} \rfloor$)	c_α
200	100	0.36	0.18	1.27	1.04	2.27	10.22
200	200	0.26	0.15	0.70	0.46	2.27	10.22
200	300	0.12	0.07	0.34	0.22	2.26	10.22
200	400	0.10	0.07	0.33	0.26	2.27	10.22
275	100	0.00	0.05	0.27	0.15	2.27	10.22
275	200	0.02	0.06	0.20	0.14	2.27	10.22
275	300	0.03	0.07	0.22	0.11	2.26	10.22
275	400	0.06	0.07	0.14	0.11	2.27	10.22
350	100	0.20	0.17	0.26	0.21	2.27	10.22
350	200	0.15	0.15	0.17	0.15	2.27	10.22
350	300	0.15	0.15	0.16	0.15	2.26	10.22
350	400	0.16	0.15	0.17	0.16	2.27	10.22
425	100	0.09	0.07	0.15	0.11	2.27	10.22
425	200	0.08	0.07	0.12	0.10	2.27	10.22
425	300	0.07	0.06	0.09	0.06	2.26	10.22
425	400	0.06	0.06	0.08	0.07	2.27	10.22

Table 4: Summary of numerical results at $\tau^0 = 0.77$. The estimates $\hat{\tau}$, and $\tilde{\tau}$ are those of Step 1 and Step 2 of Algorithm 1, respectively.