Journal of Machine Learning Research 24 (2023) 1-68 Submitted 10/22; Revised 5/23; Published 5/23

Inference on the Change Point under a High Dimensional Covariance Shift

Abhishek Kaul ABHISHEK.KAUL@WSU.EDU
Department of Mathematics and Statistics
Washington State University
Pullman, WA 99164, USA.

Hongjin Zhang HONGJIN.ZHANG@WSU.EDU
Department of Mathematics and Statistics
Washington State University
Pullman, WA 99164, USA.

Konstantinos Tsampourakis K.TSAMPOURAKIS@WSU.EDU
School of Mathematics
University of Edinburgh
Edinburgh, Scotland, EH9 3FD.

George Michailidis GMICHAIL@UFL.EDU
Department of Statistics and the Informatics Institute
University of Florida
Gainsville, FL 32611-8545.

Editor: Zaid Harchaoui

Abstract

We consider the problem of constructing asymptotically valid confidence intervals for the change point in a high-dimensional covariance shift setting. A novel estimator for the change point parameter is developed, and its *asymptotic distribution* under *high dimensional scaling* obtained. We establish that the proposed estimator exhibits a sharp $O_p(\psi^{-2})$ rate of convergence, wherein ψ represents the jump size between model parameters before and after the change point. Further, the form of the asymptotic distributions under both a vanishing and a non-vanishing regime of the jump size are characterized. In the former case, it corresponds to the *argmax* of an asymmetric Brownian motion, while in the latter case to the *argmax* of an asymmetric random walk. We then obtain the relationship between these distributions, which allows construction of regime (vanishing vs non-vanishing) adaptive confidence intervals. Easy to implement algorithms for the proposed methodology are developed and their performance illustrated on synthetic and real data sets.

Keywords: High dimensional scaling, covariance/precision shift, dynamic graphical models, change point, inference, limiting distribution

1. Introduction

The study of change points in statistical models has a long history in the literature, including both first order problems, such as mean shifts (Bai, 1994; Fryzlewicz, 2014; Wang and Samworth, 2018) and regression models (Bai, 1997; Kaul et al., 2019; Wang et al., 2019a), and second order problems such as covariance shifts Wang et al. (2021); Bai et al. (2020)

and more recently network and graphical models (Keshavarz et al., 2020; Keshavarz and Michailidis, 2020). This paper focuses on shifts in second order structure in an *offline* setting, where the entire data set is available prior to change point estimation and inference.

The main statistical tasks in change point analysis are the following: (i) *whether change point(s) exist* in the data, (ii) assuming their existence, *estimation* of their location and (iii) *post-estimation inference*. A brief literature review follows in context of these tasks. Aue et al. (2009) considers the detection of change points problem in a fixed dimension p time series model subject to temporal m-dependence. Johnstone (2001) and Birke and Dette (2005) provide further results for the detection problem in a diverging p $(p/T \to 0)$ setting in the same covariance shift model. A high dimensional covariance setting is considered in Avanesov et al. (2018). The online version of the detection problem for a graphical model in a diverging p setting is considered in Keshavarz and Michailidis (2020), wherein a regularized likelihood ratio type statistic is employed. The estimation of change point(s) problem has also received a lot of attention. In a fixed p setting for the Gaussian graphical model, Kolar et al. (2010) and Kolar and Xing (2012) consider a fused lasso regularization approach with a squared loss and a likelihood based loss, respectively. The diverging p framework under multiple change points is considered in Wang et al. (2021) and Gibberd and Roy (2017). The former proposes a modified CUSUM estimator in conjunction with binary segmentation and the latter a likelihood based estimator together with fused lasso regularization. Estimation of change points under high dimensional scaling for a Markov random field has been considered in Bybee and Atchadé (2018) and Roy et al. (2017). Finally, Barigozzi et al. (2018) consider change point estimation in a factor model in the diverging p setting, which is inherently related to the underlying second order structure.

The third problem of post-estimation inference on the change point in context of a second order shift, is currently unexplored in the literature and constitutes the primary objective of this work. To describe our objectives precisely, we first introduce the statistical model of interest. Consider multivariate data collected for T time periods and at a certain point during that time period, their covariance matrix exhibits a change. Specifically, let,

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

with $z_t \in \mathbb{R}^p$, $t = 1, ..., T$. The variables $w_t, x_t \in \mathbb{R}^p$ are independent and piecewise identically distributed zero mean sub-Gaussian random variables (r.v.'s), with unknown $p \times p$ covariances Σ and Δ, respectively, i.e., $w_t \sim^{i.i.d}$ subgaussian$(0, \Sigma)$, $t = 1, .., \tau^0$, and $w_t \sim^{i.i.d}$ subgaussian$(0, \Delta)$, $t = \tau^0 + 1, .., T$. The change point $\tau^0 \in \{1, ..., (T-1)\}$ together with the underlying covariance matrices are to be estimated from the available data. We allow p to diverge potentially at an exponentially rate, i.e., $\log p = o(T^\delta)$, for some $0 < \delta < 1$, while imposing a sparsity assumption on the model parameters specified in Section 2.

Main contributions and related literature.

1. This work aims to develop methods and results that enable inference for the change point τ^0 of the second order shift model (1.1). Specifically, introduce an estimator $\tilde{\tau}$ that possesses well defined limiting distributions, thereby allowing construction of asymptotically valid confidence intervals. The methodology should work under high dimensional scaling

for the covariance matrices Σ and Δ, and also allow for diminishing jump size across the pre- and post-change point covariances, measured by a suitable metric. To our knowledge, there is no existing method/result in the literature that allows one to perform inference on the change point estimator under a *second order shift* in a multivariate framework in a diverging or a high dimensional p setting.

It is worth noting that there exists limited literature for post-estimation inference on change point(s) involving high-dimensional models even for first order (mean) shifts. In a fixed p setting, Bai (1994, 1997) provide limiting distributions for mean shift parameters and regression coefficients, respectively. The case of diverging p is considered in Bhattacharjee et al. (2017, 2020) and Wang et al. (2019b). The high dimensional case is considered in Kaul et al. (2021) and Kaul and Michailidis (2021). All listed articles also focus on a single change point, with the last reference being a notable exception.

2. Existing work in first order settings establishes that the distributional behavior of change point estimators is split into two distinct regimes, characterized by a vanishing or a non-vanishing jump size. We obtain these distributions for the posited model (1.1) with an analogous regime distinction. However, determining *a priori* the jump size regime is not possible in practice. Hence, a new problem emerges, namely *which of the two limiting distributions* should be used to construct confidence intervals for the estimated change point. The traditional answer to this problem is that of implementing a regime adaptive bootstrap procedure, proposed in Antoch et al. (1995) in a $p = 1$ setting and also considered in Cho and Kirch (2021) for mean shift models and in Bhattacharjee et al. (2020) for stochastic block network models. We address this question in a different manner. We establish a novel result that illustrates the inherent asymptotic adaptivity of the limiting distribution under the non-vanishing jump size regime to that under the vanishing jump regime. Effectively, this result shows that if one always employs the former distribution to obtain confidence intervals, then they remain asymptotically valid, even if the regime was *mis-specified.* Hence, it eliminates the need for implementing a computationally expensive bootstrap procedure.

The closest comparable articles are those of (Kaul et al., 2021; Kaul and Michailidis, 2021) who consider the first order problem of high dimensional inference on the change point for a mean shift. In addition to novel methodology developed in this article for the considered second order setting of a covariance shift, there are also several striking theoretical distinctions with the mean shift setting described throughout the article. Specifically, a fundamental distinction of a variance inflation between the vanishing and non-vanishing regimes exists under the considered framework, a phenomenon that does not occur in the mean shift setting (Remark 6). Further, a new result on regime adaptation is presented, which has no existing counterpart in the literature. Finally, insightful distinctions in the sufficient conditions required to obtain the established inferential properties are mentioned.

Proposed methodology and other preliminaries.

We shall estimate the change point in model (1.1) by instead approaching the problem via the inverse (precision matrices) of the corresponding covariances Σ and Δ. In the static setting without change points, this approach has been shown to be especially useful under high dimensionality being analytically tractable and efficient alternative, where sample covariances are known to be inconsistent, see, for e.g. the methods of *neighborhood selection* in Meinshausen et al. (2006) and *graphical Lasso* in Friedman et al. (2008); Yuan (2010).

The following notation is needed in the sequel. For any matrix $W_{p\times p}$, define $W_{-i,j}$ as the j^{th} column of W with the i^{th} entry removed, and similarly define $W_{i,-j}$. Also define the submatrix $W_{-i,-j}$ with the i^{th} row and the j^{th} column of W removed. Next, define,

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

The vectors $\mu^0_{(j)}$, $\gamma^0_{(j)}$, $j = 1,...,p$ are motivated by their direct relation to the underlying precision matrices; specifically, letting $\Omega = \Sigma^{-1}$, it is well known that (see, e.g., Yuan (2010)) $\mu^0_{(j)} = \big(\Omega_{-j,j}/\Omega_{j,j}\big)$. This relationship between the coefficients in (1.2) and underlying precision matrices constitute the main building blocks of graphical lasso.

The adjacency matrix ($|\text{sign}(\Omega)|$) of a precision matrix Ω is also referred to as a *graphical model*. These provide a visual interpretable network wherein edges represent conditional dependencies amongst components (nodes) of the underlying random variables. A considerable body of literature on such models has been developed owing to a wide variety of applications in both static and dynamic frameworks. They have been extensively utilized in genetics and genomics (Sinoquet, 2014), metabolomics (Basu et al., 2017) and neuroimaging studies (Cribben et al., 2012). 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.

Let $\eta^0_{(j)} = \big(\mu^0_{(j)} - \gamma^0_{(j)}\big)$, $j = 1,...,p$, and define the jump size for model (1.1),

$$\xi_{2,2} = \Big(\sum_{j=1}^{p}\|\eta^0_{(j)}\|_2^2\Big)^{\frac{1}{2}}, \quad \text{and} \quad \psi = \xi_{2,2}/\sqrt{p}.^{1} \tag{1.3}$$

The quantities $\xi_{2,2}$, and ψ reflect the magnitude of the difference between the pre- and post-change precision matrices, the latter being a normalized version that plays a central role in subsequent analysis. Henceforth, we refer to ψ as the *jump size*. Note that $\xi_{2,2}$ or ψ are non-zero, either if the conditional dependence structure (edges) exhibits changes, or the magnitude of the model parameters changes. This definition of the jump size is somewhat similar to that in Kolar and Xing (2012), who define it as $\min_j \|\eta^0_{(j)}\|_2$. The advantage of using 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.

Next, define the following criterion function for estimating τ^0. Let $z_t \in \mathbb{R}^p, t = 1,..,T$ be the observed realizations, and μ, and γ be the concatenation of $\mu_{(j)}$, $\gamma_{(j)}$s. Then, consider the aggregate squared loss,

$$Q(\tau,\mu,\gamma) \;=\; \frac{1}{T}\Big[\sum_{t=1}^{\tau}\sum_{j=1}^{p}(z_{tj} - z^T_{t,-j}\mu_{(j)})^2 + \sum_{t=(\tau+1)}^{T}\sum_{j=1}^{p}(z_{tj} - z^T_{t,-j}\gamma_{(j)})^2\Big], \tag{1.4}$$

1. The jump sizes $\xi_{2,2}$, and ψ depend on the sampling period T via the dimension p, as in this high-dimensional setting p, is allowed to be sequence in T.

with $\tau \in \{1, ..., (T-1)\}$. Suppose for the time being that estimates $\hat\mu$ and $\hat\gamma$, are available. As a first task, we examine the estimation behavior of the following plug-in estimator $\tilde\tau$ of the change point, with respect to the properties of these preliminary coefficient estimates,

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

The plug-in estimator (1.5) shall yield a sharp rate of estimation $O_p(\psi^{-2})$, which in turn provides *sufficient regularity for limiting distributions to exist*, despite the presence of potential high dimensionality. The preliminary estimates $\hat\mu$, $\hat\gamma$ are required to satisfy the following ℓ_2 bound,

$$\max_{1\le j\le p}\Big(\|\hat\mu_{(j)} - \mu^0_{(j)}\|_2 \vee \|\hat\gamma_{(j)} - \gamma^0_{(j)}\|_2\Big) \le c_u\surd(1+\nu^2)\frac{\sigma^2}{\kappa}\Big\{\frac{s\log(p\vee T)}{T\ell_T}\Big\}^{\frac{1}{2}} \tag{1.6}$$

with probability at least $1-o(1)$. In Section 2, we establish the main results, under this general condition. Section 3 discusses in detail how to construct such preliminary estimates under the intuitive rate condition,

$$\Big(\frac{1}{\psi}\Big)\Big\{\frac{s\log^{3/2}(p\vee T)}{\surd(T\ell_T)}\Big\} = O(1), \tag{1.7}$$

with s being the sparsity of the precision matrices and $T\ell_T$ being the separation of the change point from the parametric boundary, i.e.. $T\ell_T = \tau^0 \wedge (T-\tau^0)$.

Notation: $\mathbb{R}$ denotes 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 $a, b \in \mathbb{R}$. 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 are with respect to the sample size $T \to \infty$, unless mentioned otherwise. We use $\Rightarrow$ to represent convergence in distribution. For a positive sequence a_n, we say a sequence of r.v.' $X_n = O_p(a_n)$ if for any $0 < \delta < 1$, we have $|X_n| \le C_\delta a_n$ with probaility at least $1-\delta$, where the constant C_δ may depend on the δ.

2. Theoretical properties of the change point estimator

Next, we state sufficient conditions required for the main theoretical results regarding $\tilde\tau$ in (1.5). These results include a sharp $O_p(\psi^{-2})$ rate of estimation, its limiting distributions in the two *regimes* together with the regime adaptation result previously mentioned.

2.1 Main assumptions

Condition A (assumption on the model parameters): *Let* $S_{1j} = \{k;\, \mu^0_{(j)k} \ne 0\}$, *and* $S_{2j} = \{k;\, \gamma^0_{(j)k} \ne 0\}$, $1 \le j \le p$ *be sets of non-zero indices.*
(i) Assume that $\max_{1\le j\le p}|S_{1j}| \vee |S_{2j}| = s \ge 1$.
(ii) Assume a change point exists and is sufficiently separated from the boundaries $\{0, T\}$,

i.e., for some positive sequence $\ell_T \to 0$, *we have* $\big(\tau^0 \wedge (T-\tau^0)\big) \geq T\ell_T \to \infty$
(iii) Let ψ *be as defined in (1.3). Then, for an appropriately chosen small enough constant* $c_{u1} > 0$, *the following relations hold,*

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

$$(b) \quad c_u\surd(1+\nu^2)\frac{\sigma^2}{\psi\kappa}\left\{\frac{s\log(p\vee T)}{T^{\left(\frac{1}{2}-b\right)}\surd\ell_T}\right\} \leq c_{u1},$$

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

Condition A controls rate at which the sparsity level s of the precision matrices and its dimension p diverge as a function of T. Further, it controls the jump size ψ and the distance ℓ_T of the change point from the parametric boundary. Condition A(iii) encompasses the two regimes of interest on the asymptotic behavior of the jump size. Specifically, it allows for a potentially vanishing jump size, $\psi \to 0$, when $s\log^{3/2}(p \vee T) = o\big(\surd(T\ell_T)\big)$. Alternatively, s, p can diverge at an arbitrary rate provided the jump size is large enough to compensate for the increasing dimensions s, p, so that Condition A(iii) holds.

To our knowledge, this is the weakest condition assumed on the jump size in the second order shift literature. The constant $b > 0$ in A(iii)(b) is any fixed number between $(0, 1/2)$. The rate conditions (a) and (b) of A(iii) are stated in the given form to provide generality and neither (a) or (b) implies the other without additional restrictions; for e.g., (b) implies (a) if $\log p \leq c_u T^{2b}$, while (a) implies (b) if $\log p \geq c_u T^{2b}$. For fixed s, p, l_T, the rate required for the minimum jump size ψ in Part (iii) can be replaced with $T^{\left(\frac{1}{2}-b\right)}\psi \to \infty$. This condition is identical to Assumption A7 in Bai (1997) and serves an analogous role.

Sparsity of $\mu^0_{(j)}$ and $\gamma^0_{(j)}$ is equivalent to sparsity of precision matrices Σ^{-1} and Δ^{-1}. In context of the associated graphical models, it assumes each node has at most s connecting edges out of a total $(p-1)$ possible edges. This is a direct extension of the same assumption in (Meinshausen et al., 2006; Yuan, 2010).

Condition B (assumption on the underlying distributions):
(i) The vectors $w_t = (w_{t1}, ..., w_{tp})^T$, $t = 1, .., \tau^0$, *and* $x_t = (x_{t1}, ..., x_{tp})^T$, $t = \tau^0 + 1, ...T$, *are independent sub-Gaussian r.v's with variance proxy* $\sigma^2 \leq c_u$. *(see Definition 38)*
(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* $\nu = \phi/\kappa$.

The sub-Gaussian assumption is a significant relaxation to the more typical Gaussian assumption made in the literature for graphical models. It allows asymmetric distributions, including centered mixtures of Gaussian distributions. Conditions B can be stated more succinctly as boundedness of the *Orlicz* norm (see e.g., Wang et al. (2021)). We choose this representation in favor of explicit clarity on underlying parametric assumptions, see, Vershynin (2019) for several other equivalent characterizations of sub-Gaussian distributions.

For the remainder of this section, we are *agnostic* on the choice of the estimator for nuisance parameters $\mu^0_{(j)}$ and $\gamma^0_{(j)}$, $j = 1, ..., p$, and instead require the following condition.

Condition C (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 (1.6).*
(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 and is known to hold in the static setting by common precision matrix estimation methods, including neighborhood selection and Graphical Lasso. Condition C(ii) provides a restriction on the sparsity level of the estimated edge parameters and is common in the ℓ_1 regularization literature. In Section 3, the estimates of the nuisance parameters developed satisfy this condition. Further, other common regularization mechanisms, such as SCAD or the Dantzig selector are also applicable.

Feasibility of this condition is illustrated in detail in Section 3, where we show that it can be eliminated and replaced by the intuitive rate restriction,

$$\Big(\frac{1}{\psi}\Big)\Big\{\frac{s\log^{3/2}(p \vee T)}{\surd(T\ell_T)}\Big\} = O(1),$$

Further, a new estimator is proposed that satisfies the bound in (1.6). Finally, it is shown how existing change point estimators in the literature -e.g., Lee et al. (2016), Bybee and Atchadé (2018); Wang et al. (2021)- can serve towards the development of these preliminary estimates, even though they do not possess the properties needed for inference; see, Corollary 11 and Algorithm 1 of Section 3.

2.2 Rate of Estimation of $\tilde{\tau}$

We begin with our first result establishing the rate of convergence of the proposed estimator.

Theorem 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 15. Then, for T sufficiently large, we have*

$$(1 + \nu^2)^{-1}(\sigma^2 \vee \phi)^{-2}\kappa^2\psi^2|\tilde{\tau} - \tau^0| \le c_u^2 c_{a1}^2, \tag{2.1}$$

with probability at least $1 - 3a - o(1)$. *Equivalently, we have,* $\big(\tilde{\tau} - \tau^0\big) = O_p(\psi^{-2})$.

Theorem 1 provides an $O_p(\psi^{-2})$ rate of estimation, wherein the bounding constant c_a depends on the probability of the bound. This is sharper and in contrast to existing localizing bounds in the literature, for example an $O\big(\log(p \vee T)\big)$ bound in Roy et al. (2017) that holds with probability $1 - o(1)$; to observe this improvement note that the result in Theorem 1 implies the latter, but the converse is not true. Moreover, its importance in context of inference can be noted by observing that the bound of Theorem 1 traps the sequence in $\psi^{-2}(\tilde{\tau} - \tau^0)$ in a finite interval with a specific probability, thus providing the feasibility of constructing a confidence interval. Also note that existing bounds as described earlier do not yield this property, as their approach does not attempt to obtain a sharp bound and instead conservatively bounds the sequence $\psi^{-2}(\tilde{\tau} - \tau^0)$ under consideration, by dominating the bounding constant by a diverging sequence ($\log p$ or similar). As a consequence, these bounds do not provide a precise measure of probability and only guarantee that the sequence converges to one. As such, it is the sharpest rate of estimation for a change point estimator available in the literature that allows diverging or high dimensional

p. For example, the corresponding estimation rate in Kolar and Xing (2012) and Gibberd and Roy (2017) is $O\big(\psi^{-2}\log(p\vee T)\big)$, in Li et al. (2019) is $O(\psi^{-2}\log^4 T)$, each holding with probability $1-o(1)$. Note that the jump size ψ appearing in the respective rates for (Gibberd and Roy, 2017; Li et al., 2019; Roy et al., 2017) corresponds to the normalized measure $\psi=\|\Sigma-\Delta\|_F/\sqrt{p}$.

This result does not come at a price of stronger parametric assumptions. Specifically, the jump size can potentially diminish to zero with the assumption (1.7) only requiring ψ to be at least of order $O\big(s\log^{3/2}(p\vee T)/\sqrt{(T\ell_T)}\big)$. This is the weakest condition available on the jump size in the literature for a graphical model; for example, Kolar and Xing (2012) require it to be $O(p\log T/T)^{1/2}$, Gibberd and Roy (2017) require $O\big(p\sqrt{(\log p^{\beta/2}/T)}\big)$, while several other articles also assume this jump size to be bounded below.

2.3 Asymptotic distributions of $\tilde\tau$ and their relationship

Towards obtaining asymptotic distributions consider the following process, let $W_1(r)$, and $W_2(r)$ be two independent Brownian motions defined on $[0,\infty)$ and define,

$$Z(r):=Z\big(r,\sigma^2,\sigma^{*2}\big)=\begin{cases}2W_1(r)-|r| & \text{if } r<0,\\ 0 & \text{if } r=0,\\ \frac{2\sigma_2^*}{\sigma_1^*}W_2(r)-\frac{\sigma_2^2}{\sigma_1^2}|r| & \text{if } r>0,\end{cases} \tag{2.2}$$

where $0<\sigma_1,\sigma_2,\sigma_1^*,\sigma_2^*<\infty$, are estimable parameters that control both the variance and the negative drift of the process $Z(r)$. Next, assume the following mild technical conditions that control the processes (2.2). For this purpose, define

$$\varepsilon_{tj}=\begin{cases}z_{tj}-z_{t,-j}^T\mu_{(j)}^0, & t=1,...,\tau^0\\ z_{tj}-z_{t,-j}^T\gamma_{(j)}^0, & t=\tau^0+1,...,T.\end{cases} \tag{2.3}$$

Condition D: *(i) Assume that the following limits exist,*

$$\xi_{2,2}^{-2}\sum_{j=1}^p\eta_{(j)}^{0T}\Sigma_{-j,-j}\eta_{(j)}^0\to\sigma_1^2,\quad \text{and}\quad \xi_{2,2}^{-2}\sum_{j=1}^p\eta_{(j)}^{0T}\Delta_{-j,-j}\eta_{(j)}^0\to\sigma_2^2,\quad 0<\sigma_1^2,\sigma_2^2<\infty.$$

(ii) For ε_{tj}, for $t=1,...,T$, and $j=1,...,p$, as defined in (2.3), assume that,

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

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

where $0<\sigma_1^{*2},\sigma_2^{*2}<\infty$.

All limits are with respect to T. The limits of Condition D are acting in T via the dimension p and the jump size $\xi_{2,2}$. The limits σ_1^2 and σ_2^2 control the negative drift and σ_1^{*2} and σ_2^{*2} control the variance of the process (2.2).

Note that finiteness of the limits appearing in Condition D is already guaranteed by prior assumptions, and this condition only assumes their stability, i.e., we only assume here that the sequence under consideration is not a bounded oscillating sequence such as $\{-1^n\}_{n=1}^{\infty}$. Assumptions of this form are very commonly utilized in the literature to establish limiting distributions. The most common one is the stability of the Gram matrix (X'X/n), assumed in linear regression settings for inference on the regression coefficients.

To see the described boundedness, first consider Condition D(i) and note that the assumed convergence is on a sequence that is guaranteed to be bounded, i.e.,

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

wherein the inequalities follow from the bounded eigenvalues assumption on the covariance matrix Σ $\big($Condition B(ii)$\big)$, and analogous for the post-change covariance matrix Δ. An easier to interpret, but stronger sufficient condition for the finiteness for the limits in Condition D(i) is as follows. Let $\Sigma = \big[\sigma_{ij}\big]_{i,j=1,\ldots,p}$, and Δ be symmetric matrices such that,

$$\|\Sigma\|_1 = \max_{1\le j\le p} \sum_{i=1}^{p} |\sigma_{ij}| < \infty,$$

and analogous for the matrix Δ. Then, we get

$$\xi_{2,2}^{-2}\sum_{j=1}^{p} \eta_{(j)}^{0T}\Sigma_{-j,-j}\eta_{(j)}^0 \le \|\Sigma\|_\infty\|\Sigma\|_1\xi_{2,2}^{-2}\sum_{j=1}^{p}\|\eta_{(j)}\|_2^2 = \|\Sigma\|_\infty\|\Sigma\|_1 < \infty,$$

wherein the inequality follows from the relation $\|\Sigma\|_2^2 \le \|\Sigma\|_\infty\|\Sigma\|_1$, with $\|\Sigma\|_2$ denoting the operator norm. In other words, finiteness of the limits of D(i) are guaranteed by absolute summability of components of each row (or column) of the underlying covariances, which are in turn satisfied by large classes of such matrices, including Toeplitz and banded ones.

Next, finiteness of the assumed limits in Condition D(ii) can be illustrated by using properties of sub-Gaussian distributions assumed earlier in Condition B. Specifically, let $\zeta_{tj} = \varepsilon_{tj}z_{t,-j}^T\eta^0$ and $\zeta_t = \sum_{j=1}^{p}\zeta_{tj}$, and note that $E(\zeta_t) = 0$. Further, using part (ii) of Lemma 19 we get that $\zeta_t \sim \text{subE}(\lambda)$, $\lambda = O(\xi_{2,1})$, with $\xi_{2,1} = \sum_{j=1}^{p}\|\eta_{(j)}^0\|_2$. Hence, $\xi_{2,2}^{-1}p^{-1}\text{var}(\zeta_t) = O\big(\xi_{2,1}^2/p\xi_{2,2}^2\big) = O(1) < \infty$, which follows by utilizing the elementary relation $\xi_{2,1} \le \sqrt{p}\xi_{2,2}$ between the $1-$norm and $2-$norm.

Next, we state the result for the asymptotic distribution of the change point estimator for the *vanishing jump size regime* $\psi \to 0$.

Theorem 2 *(Vanishing jump size regime) Suppose Conditions A, B, C, and D hold. Further, assume that* $\psi \to 0$, *while satisfying,*

$$\frac{1}{\psi}\Big\{\frac{s\log^{3/2}(p\vee T)}{\sqrt{(T\ell_T)}}\Big\} = o(1). \tag{2.4}$$

Then, the estimator $\tilde\tau$ *of (1.5) has the following limiting distribution.*

$$(\sigma_1^*)^{-2}\sigma_1^4\psi^2(\tilde\tau - \tau^0) \Rightarrow \arg\max_{r\in\mathbb{R}} Z(r). \tag{2.5}$$

where $Z(r)$ *is defined in (2.2).*

The cumulative distribution function of this limiting distribution is readily available in Bai (1997), thereby allowing straightforward computation of its quantiles.

Remark 3 *(On adaptation to nuisance estimates)* The posited limit distribution is the same as one would obtain when the nuisance parameters μ^0, γ^0 were *known.* This is despite $\tilde{\tau}$ utilizing $2p$ estimated vectors $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$, $j = 1, ..., p$, each of dimension $(p-1)$. This is effectively the adaptation property as described in Bickel (1982), but in a high dimensional setting and within a change point parameter context.

Remark 4 *(On the sparsity requirement and interplay with jump size)* A note of interest concerns the jump size scaling in Condition A and its relation to the inference properties. We note that the scaling $\psi \geq \big(c_u s \log^{3/2}(p \vee T)/\sqrt{T\ell_T}\big)$ viewed from a sparsity (s) perspective assumes a more sparse regime than the scaling $\psi \geq c_u\big\{s\log(p\vee T)/T\ell_T\big\}^{1/2}$ for which near optimal estimation results have been established in context of other dynamic models such as that of linear regression, see, e.g. Rinaldo et al. (2020). Assuming an increased sparsity level is a key distinction that makes the inference results feasible. Some evidence pointing to the sharpness of this assumption follows. In a linear regression framework, Lemma 4 in Rinaldo et al. (2020) shows that the minimax optimal rate of estimation under a scaling $\psi \geq c_u\big\{s/T\big\}^{1/2}$ is $s\psi^{-2}$, i.e., slower than $O_p(\psi^{-2})$ obtained above and in turn disallowing inference. Thus, at the very least, one may conclude that the sparsity level necessary for being able to carry out inference should be diverging at a slower rate, as that assumed in Condition A. Additional indirect evidence for the sharpness of this super-sparse scaling arises from recent results on inference for a regression coefficient in the presence of high dimensionality. The debiased lasso (Van de Geer et al. (2014)) and orthogonalized moment estimators (Belloni et al., 2014) and Ning et al. (2017) developed for this purpose, require a similar super-sparsity assumption $s\log p/\sqrt{T} = o(1)$ for validity of inference results, over an ordinary sparsity assumption.

Next, we obtain the limiting distribution in the non-vanishing jump size regime $\psi \to \psi_\infty$, $0 < \psi_\infty < \infty$. Let $\mathcal{L}$ represent the form of the distribution of the limiting random variable of the sequence $p^{-1}\sum_{j=1}^{p}\Big\{2\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 - \eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0\Big\}$. Then, define the following two sided random walk with negative drift, initialized at the origin

$$\mathcal{C}_\infty(r) := \mathcal{C}_\infty\big(r, \psi_\infty, \sigma^2, \bar{\sigma}^2\big) = \begin{cases} \sum_{t=1}^{-r} z_t, & r \in \mathbb{N}^- = \{-1,-2,-3,...\} \\ 0, & r = 0 \\ \sum_{t=1}^{r} z_t^*, & r \in \mathbb{N}^+ = \{1,2,3,...\}. \end{cases} \tag{2.6}$$

Further, $z_t \sim^{i.i.d} \mathcal{L}\big(-\psi_\infty^2\sigma_1^2, \bar{\sigma}_1^2\big)$ and $z_t^* \sim^{i.i.d} \mathcal{L}(-\psi_\infty^2\sigma_2^2, \bar{\sigma}_2^2)$, and z_t and z_t^* are also independent of each other over all t. The notation in the arguments of $\mathcal{L}(\cdot\,,\cdot)$ correspond to the mean and variance of this distribution. The quantities $0 < \sigma_1, \sigma_2 < \infty$ are the same as in the construction of the process $Z(r)$. The parameters $0 < \bar{\sigma}_1^2, \bar{\sigma}_2^2 < \infty$ are estimable variance parameters of this limiting process which are related, but not identical to those under the vanishing regime. To ensure regularity of this limiting process we require an additional distributional assumption.

Condition B′ (further distributional assumption): *Suppose Conditions B and D hold. Let* σ_1^2, σ_2^2 *be as defined in Condition D and let* $\bar\sigma_1^2 = \lim_T \text{var}\Big[p^{-1}\sum_{j=1}^p\Big\{2\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 - \eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0\Big\}\Big]$, $t \le \tau^0$, *and similarly define* $\bar\sigma_2^2$ *for* $t > \tau^0$, *such that* $0 < \bar\sigma_1^2, \bar\sigma_2^2 < \infty$. *Then, assume*

$$p^{-1}\sum_{j=1}^p\Big\{2\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 - \eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0\Big\} \;\Rightarrow\; \mathcal{L}\big(-\psi_\infty^2\sigma_1^2,\, \bar\sigma_1^2\big), \quad \text{for } t \le \tau^0$$
$$p^{-1}\sum_{j=1}^p\Big\{2\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 - \eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0\Big\} \;\Rightarrow\; \mathcal{L}\big(-\psi_\infty^2\sigma_2^2,\, \bar\sigma_2^2\big), \quad \text{for } t > \tau^0$$

for some distribution law $\mathcal{L}$ *which is continuous and supported in* $\mathbb{R}$.

The only additional requirement imposed by Condition B′, in comparison to Conditions B and D, is that the r.v.'s under consideration are continuously distributed; this is trivially true in the typical Gaussian graphical model framework. The arguments in notation $\mathcal{L}(\mu,\sigma^2)$ are used to represent mean and variance of the distribution, i.e, $E\mathcal{L}(\mu,\sigma^2) = \mu$, and $\text{var}\big(\mathcal{L}(\mu,\sigma^2)\big) = \sigma^2$. The notation $\mathcal{L}(\mu,\sigma^2)$ is only for ease of presentation and does not imply that $\mathcal{L}$ is characterized by only its mean and variance.

To illustrate the mildness of Condition B′, consider the mean of the sequence of r.v.'s under consideration for $t \le \tau^0$,

$$Ep^{-1}\sum_{j=1}^p\Big\{\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 - \eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0\Big\} = -p^{-1}\sum_{j=1}^p\eta_{(j)}^{0T}\Sigma\eta_{(j)}^0 \to -\psi_\infty^2\sigma_1^2,$$

and analogously for $t > \tau^0$. The equality follows since $E\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0 = \eta_{(j)}^{0T}\Sigma\eta_{(j)}^0$, and $E\varepsilon_{tj} = Ez_{t,-j}^T\eta_{(j)}^0 = 0$, and moreover, ε_{tj} and $z_{t,-j}$ are uncorrelated by construction in (2.3). Then, convergence in expected value follows from Condition D(i) provided that $\psi \to \psi_\infty$.

Next, consider the variance of these r.v.'s. From the properties of sub-Gaussian and sub-exponential distributions (also see, discussion after Condition D), we have,

$$\text{var}\Big[p^{-1}\sum_{j=1}^p 2\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0\Big] = O\big(\xi_{2,2}^2/p\big) = O(\psi_\infty^2) < \infty, \quad \text{and}$$
$$\text{var}\Big[p^{-1}\sum_{j=1}^p \eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0\Big] = O\big(\xi_{2,2}^4/p^2\big) = O(\psi_\infty^4) < \infty, \quad \text{thus,}$$
$$\text{var}\Big[p^{-1}\sum_{j=1}^p\Big\{2\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 - \eta_{(j)}^{0T}z_{t,-j}z_{t,-j}^T\eta_{(j)}^0\Big\}\Big] = O(\psi_\infty^2) + O(\psi_\infty^4) < \infty. \quad (2.7)$$

Relation (2.7) implies the sequence of r.v.'s in Condition B′ have bounded variances, thereby the distribution of the limiting random variable is well defined ($\mathcal{L} < \infty$, a.s.). Condition B′ simply provides a notation $\mathcal{L}$ to whatever distribution this may be, with appropriate variance notation $\bar\sigma_1^2$ or $\bar\sigma_2^2$, respectively. One may observe that thus far in our discussion no additional assumption has been made in Condition B′ in addition to Condition B and Condition D and the change of regime to the non-vanishing jump size.

The only additional assumption of Condition B$'$, of continuity of the distribution $\mathcal{L}$ is assumed for the regularity of the *argmax* of this two sided random walk (see, Lemma 16).

Theorem 5 *(Non-vanishing jump regime) Suppose Conditions A, B$'$, C, and D hold. Further, assume that $\psi \to \psi_\infty$, $0 < \psi_\infty < \infty$, and that $\big\{s\log^{3/2}(p \vee T)\big/\sqrt{(Tl_T)}\big\} = o(1)$. Then, the estimator $\tilde{\tau}$ in (1.5) has the following limiting distribution:*

$$(\tilde{\tau} - \tau^0) \Rightarrow \arg\max_{r \in \mathbb{Z}} \mathcal{C}_\infty(r). \tag{2.8}$$

where $\mathcal{C}_\infty(r)$ is as defined in (2.6).

The process $\mathcal{C}_\infty(r)$ is a two sided random walk with negative drift and continuously distributed increments. Although the limit distribution is well defined, its distribution function does not have any known characterization. Its quantiles can be approximated numerically by first drawing sample path realizations of the given stochastic process, in turn providing realizations of its *argmax*.

Remark 6 *(Comparison of limiting distribution results obtained to those established for mean shift models)* An important observation distinguishing these processes is that the limiting process in the vanishing regime is characterized by the sequence $\zeta_t = p^{-1}\sum_{j=1}^p \varepsilon_{tj} z_{t,-j}^T \eta^0_{(j)}$, $t = 1, ..., T$, whereas in the non-vanishing regime by the sequence $\zeta_t = p^{-1}\sum_{j=1}^p \Big\{\varepsilon_{tj} z_{t,-j}^T \eta^0_{(j)} - \eta^{0T}_{(j)} z_{t,-j} z_{t,-j}^T \eta^0_{(j)}\Big\}$, $t = 1, ..., T$. A somewhat unusual consequence is that the increments of the limiting process change from symmetric to asymmetric in the vanishing and non-vanishing regimes, respectively. This observation further distinguishes the above result from that for mean shift models, such as that obtained in Kaul et al. (2021), wherein the same sequence of r.v.'s characterizes limiting processes for both vanishing and non-vanishing regimes. Another consequence is that the presence of an additional quadratic form in the sequence of interest leads to an inflation in the variance of the process in the non-vanishing regime. The reason for this can be observed from (2.7), where the variance of the quadratic form is $O(\psi^4)$, whereas the variance of the remainder is $O(\psi^2)$. Thus, in the vanishing regime the first part of the r.v. under consideration dominates the quadratic form, which is no longer true in the non-vanishing jump regime.

The above results provide the capability of performing inference on τ^0 given the regime of the jump size ψ. However, determination of whether one operates in a vanishing vs. a non-vanishing regime is unverifiable in practice. Consequently, it is unclear which of the two confidence intervals constructed using Theorems 2 or 5 would provide a more precise representation in a real data setting. In order to address this problem, we examine the relationship between the two limiting distributions obtained in (2.5) and (2.8) and show that if the underlying regime is a vanishing one, then the distribution of the non-vanishing regime undergoes an asymptotic adaptation to it. This result has significant implications which are discussed after the statement of the result below.

The following result establishes an asymptotic relationship between two distributions. It is independent of the considered model framework. However, it is stated in coherent notations to allow its direct applicability. For explicit clarity on the acting limits in this result, we introduce subscripts on all involved sequences which have thus far been implicit.

Theorem 7 *Suppose $\mathcal{L}(\mu,\sigma^2)$ is any distribution law, with $E\mathcal{L}(\mu,\sigma^2)=\mu$ and finite variance, $\mathrm{var}\mathcal{L}(\mu,\sigma^2)=\sigma^2<\infty$. Let $\mathcal{L}$ be continuously distributed and invariant under scalar addition and multiplication*[2]*. Furthermore, let ψ_T, $\sigma^2_{1T},\sigma^2_{2T}$ and $\bar\sigma^2_{1T},\bar\sigma^2_{2T}$ be positive sequences in T, such that, as $T\to\infty$, the following limits hold. (i) $\psi_T\to 0$, (ii) $\sigma^2_{1T}\to\sigma^2_1$, and $\sigma^2_{2T}\to\sigma^2_2$, and finally, (iii) $\psi_T^{-2}\bar\sigma^2_{1T}\to 4\sigma_1^{*2}$ and $\psi_T^{-2}\bar\sigma^2_{2T}\to 4\sigma_2^{*2}$, where $0<\sigma_1,\sigma_2,\sigma_1^*,\sigma_2^*<\infty$. Then, we have the following,*

$$(\sigma_1^*)^{-2}\sigma_1^4\psi_T^2\arg\max_{r\in\mathbb{Z}}\mathcal{C}_\infty\big(r,\psi_T,\sigma_T^2,\bar\sigma_T^2\big)\Rightarrow\arg\max_{r\in\mathbb{R}}Z\big(r,\sigma^2,\sigma^{*2}\big),\quad \text{when } T\to\infty, \tag{2.9}$$

where $\mathcal{C}_\infty\big(r,\psi_T,\sigma_T^2,\bar\sigma_T^2\big)$ and $Z\big(r,\sigma^2,\sigma^{*2}\big)$ are as defined in (2.6) and (2.2), respectively.

Theorem 7 shows that the distributions in the non-vanishing and vanishing jump size regimes correspond to the *discrete* and *continuous* versions of the *same underlying* stochastic process (when viewed in a limiting sense). This is conceptually akin to the elementary limiting relationship between geometric and exponential distributions. Technically, Theorem 7 is perhaps not particularly surprising when viewed from a probabilistic perspective. All it says is that a suitably normalized random walk converges to a Brownian motion (functional central limit theorem (Donsker's Theorem), see, e.g., Theorem 4.3.2 of Whitt (2002)). The convergence of the *argmax* following an inclusion of a negative drift can then be viewed roughly as an application of a version of the continuous mapping theorem.

Remark 8 In order to view Theorem 7 in context of the model (1.1) and the limiting results (2) and (5) one can set distribution $\mathcal{L}$ as that in Condition B' and the underlying sequences as follows. 1. The sequence ψ_T is the jump size as defined in (1.3). 2. The sequences σ^2_{1T} and σ^2_{2T} represent finite sample negative drifts as defined in Condition D, i.e.,

$$\sigma^2_{1T}=\xi_{2,2}^{-2}\sum_{j=1}^{p}\eta^{0T}_{(j)}\Sigma_{-j,-j}\eta^0_{(j)},$$

and symmetrically define σ^2_{2T}, w.r.t covariance matrix Δ. 3. The sequences $\bar\sigma^2_{1T}$ and $\bar\sigma^2_{2T}$ represent finite sample variances as defined in Condition B', i.e.,

$$\bar\sigma^2_{1T}=\mathrm{var}\Big[p^{-1}\sum_{j=1}^{p}\Big\{2\varepsilon_{tj}z^T_{t,-j}\eta^0_{(j)}-\eta^{0T}_{(j)}z_{t,-j}z^T_{t,-j}\eta^0_{(j)}\Big\}\Big],\quad \text{for any } t\le\tau^0,$$

and symmetrically define $\bar\sigma^2_{2T}$, for $t>\tau^0$. Then, all assumptions of Theorem 7 hold. Specifically, $\psi_T\to 0$ from regime mis-specification. $\sigma^2_{1T}\to\sigma^2_1$, and $\sigma^2_{2T}\to\sigma^2_2$, from Condition D. Finally, $\psi_T^{-2}\bar\sigma^2_{1T}\to 4\sigma_1^{*2}$ and $\psi_T^{-2}\bar\sigma^2_{2T}\to 4\sigma_2^{*2}$, where σ_1^{*2} and σ_2^{*2} are also defined in Condition D. The latter convergence follows from (2.7) and the discussion in Remark 6.

To discuss implications of Theorem 7, note the normalization on the lhs of (2.5) is the same as on the lhs of (2.9). On the other hand, the lhs of (2.8) is un-normalized, thus Theorem 7 provides a direct connection between confidence intervals for τ^0 in either of the vanishing and non-vanishing regimes. More precisely, one can construct intervals as,

$$\mathrm{CI}(\tilde\tau):=\big[(\tilde\tau-q^{nv}_\alpha),(\tilde\tau+q^{nv}_\alpha)\big], \tag{2.10}$$

2. $a\mathcal{L}(\mu,\sigma^2)+b\sim\mathcal{L}(a\mu+b,\,a^2\sigma^2)$, for any constants $a,b<\infty$.

where q_α^{nv} is the symmetric quantile so that $(-q_\alpha^{nv}, q_\alpha^{nv})$ contains $(1-\alpha)$ of the probability mass of (2.8) of the non-vanishing regime. The interval $\mathrm{CI}(\tilde{\tau})$ is clearly valid w.r.t the non-vanishing regime. On the other hand, an interval constructed directly using Theorem 2 for the vanishing case would be $\big[\tilde{\tau} \pm \sigma_1^{*2}\sigma_1^{-4}\psi^{-2}q_\alpha^v\big]$, where q_α^v is the corresponding quantile of the distribution in this case. Instead of using this direct formulation, obtaining the required quantile using the finite sample approximation of Theorem 7, one obtains the asymptotically equivalent $(\sigma_1^*)^2\sigma_1^{-4}\psi^{-2}q_\alpha^v \asymp q_\alpha^{nv}$. Substituting back in the interval yields exactly the construction of $\mathrm{CI}(\tilde{\tau})$. Thus, the results of Theorem 2, Theorem 5 and their relationship in Theorem 7 together imply that doing so shall yield the desired $(1-\alpha)$ asymptotic coverage, $pr\Big((\tilde{\tau} - q_\alpha^{nv}) \le \tau^0 \le (\tilde{\tau} + q_\alpha^{nv})\Big) \to (1-\alpha)$, irrespective of whether the underlying regime is vanishing or non-vanishing. In other words, the interval $\mathrm{CI}(\tilde{\tau})$ is indeed regime adaptive. A summary of the implementation details for computation of $CI(\tilde{\tau})$ is provided as Algorithm 4 in Section 4.

The above analysis can be appreciated by a comparison to the only other method that provides this feature, which is regime adaptive bootstrap, for e.g., Antoch et al. (1995); Cho and Kirch (2021). It proceeds by obtaining an empirical distribution of the change point estimator. Our approach on the other hand completely bypasses the need to perform such computationally intensive examination of the empirical distribution, which is especially useful under the considered high dimensional setting. This comparison is also only heuristic since no such bootstrap method is available in the considered second order setting.

Finally, we briefly discuss the converse case, i.e., what would happen if we use the confidence intervals of the vanishing jump size regime in the non-vanishing regime. The reason one may be tempted to perform this converse routine may be due to the ease of computing of quantiles. However, an immediate concern is regarding the variance inflation of the non-vanishing regime discussed in Remark 6, i.e., the asymptotic variance σ^{*2} implied by the vanishing regime will underestimate the true variance under a finite jump size. Thus, employing confidence intervals obtained for the vanishing regime, under regime mis-specification may yield inaccurate coverage. A potential numerical correction for this problem is to utilize a corrected version $(\psi^{-2}\bar{\sigma}^2/4)$ in place of σ^{*2} as suggested by the relation between these two variances (see, (3) of Remark 8) while utilizing the result of the vanishing regime. A detailed examination of this problem and theoretical validation of this correction remains an important question, however, it is outside the scope of this article. We also mention that an alternative recent approach has also been proposed in Theorem 28 of Bhattacharjee et al. (2020) in the direction of regime adaptation, which relies on simulated sampling of the data generating process and generating an empirical distribution of change point estimates which in turn can be utilized to obtain regime free confidence intervals. This approach is conceptually closer to the traditional bootstrap method and is thus numerically significantly more intensive. Further, it relies on several additional assumptions on the signal to noise ratio for the setting under consideration. We conclude the Section with a remark on implementation of the interval $\mathrm{CI}(\tilde{\tau})$.

Remark 9 *(Numerical approximations of distribution law $\mathcal{L}$)* One requires quantiles from the limiting distributions in order to construct $\mathrm{CI}(\tilde{\tau})$. Unlike the distribution in Theorem 2, this cdf is not available analytically. To construct confidence intervals, one needs to simulate sample paths of the random walk $\mathcal{C}_\infty(r)$, and then obtain realizations of its *argmax* to calibrate the quantiles of $\mathcal{L}$. Specifically, realizations from the incremental distributions $\mathcal{L}\big(-\psi^2\sigma_1^2,\,\bar\sigma_1^2\big)$ and $\mathcal{L}\big(-\psi^2\sigma_2^2,\,\bar\sigma_2^2\big)$ of Condition B$'$ need to be simulated. Note that the means can be computed as plug in estimates from the estimated jump size and the given form in Remark 8. The variances $\bar\sigma_1^2$ and σ_2^2 can also be estimated as piecewise sample variances from the observed data by noting that one has available T predicted realizations, $\hat\zeta_t = p^{-1}\sum_{j=1}^{p}\Big\{2\hat\varepsilon_{tj}z_{t,-j}^T\hat\eta_{(j)} - \hat\eta_{(j)}z_{t,-j}z_{t,-j}^T\hat\eta_{(j)}\Big\}$, $t=1,...,T$. Thus, the only missing link that remains is the form of the distribution $\mathcal{L}$. Since no explicit assumptions on the form of the underlying data generating distribution have been made in the article, thus identifying the distribution $\mathcal{L}$ is not analytically feasible. For the Gaussian case, the distribution $\mathcal{L}$ becomes an average of inter-dependent Variance-Gamma distributed random variables, which to the best of our knowledge has no known analytical form. We overcome this hurdle of choosing the form of $\mathcal{L}$ by performing an empirical fit to the predicted realizations $\hat\zeta_t$, $t=1,...,T$, by means of the Kolmogorov-Smirnov goodness of fit test. Details of this process are described in Section 4 and Algorithm 3 therein.

3. Constructing a feasible $O_p(\psi^{-2})$ estimator for τ^0

The results of Section 2 rely on estimates of the nuisance parameters satisfying Condition C. Procedures to obtain such estimates and the required theoretical guarantees are discussed next. We start by introducing additional notation. For any $\tau\in\{(1,...,(T-1))\}$, consider the ℓ_1 regularized (Lasso) estimates of the regression of each column of the observed variables $z\in\mathbb{R}^{T\times p}$ on the remaining columns, for each of the two binary partitions induced by τ. Specifically, for each $j=1,...,p$, define,

$$\hat\mu_{(j)}(\tau) = \arg\min_{\mu_{(j)}\in\mathbb{R}^{p-1}}\Big\{\frac{1}{\tau}\sum_{t=1}^{\tau}\big(z_{tj}-z_{t,-j}^T\mu_{(j)}\big)^2 + \lambda_j\|\mu_{(j)}\|_1\Big\}, \tag{3.1}$$
$$\hat\gamma_{(j)}(\tau) = \arg\min_{\gamma_{(j)}\in\mathbb{R}^{p-1}}\Big\{\frac{1}{(T-\tau)}\sum_{t=\tau+1}^{T}\big(z_{tj}-z_{t,-j}^T\gamma_{(j)}\big)^2 + \lambda_j\|\gamma_{(j)}\|_1\Big\},$$

where $\lambda_j>0$. Towards obtaining these estimates, we begin with a modification of Condition A that is sufficient for this Section and is weaker than the original one.

Condition A$'$ (assumption on model parameters): *Assume Condition A(i) and A(ii) and in place of A(iii) assume the following weaker requirement.*

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

for some $0<b<(1/2)$, *where* $c_{u1}>0$ *is a suitably chosen small constant.*

The following result studies the behavior of the estimates (3.1) uniformly over the collection $\mathcal{G}(u_T,v_T)$ of τ as defined in (A.2).

Theorem 10 *Suppose Condition A$'$ and B hold. Let* $0\le u_T\le 1$ *be any sequence and* $\lambda_j = 2(\lambda_{1j}+\lambda_{2j})$, *where*

$$\lambda_{1j} = c_u\sigma^2\surd(1+\nu^2)\Big\{\frac{\log(p\vee T)}{T\ell_T}\Big\}^{\frac{1}{2}},\quad \lambda_{2j} = c_u(\sigma^2\vee\phi)\|\eta^0_{(j)}\|_2\max\Big\{\frac{\log(p\vee T)}{T\ell_T},\ \frac{u_T}{\ell_T}\Big\}$$

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 $\hat{\mu}_{(j)}(\tau) - \mu^0_{(j)} \in \mathcal{A}_{1j}$, and $\hat{\gamma}_{(j)}(\tau) - \gamma^0_{(j)} \in \mathcal{A}_{2j}$, 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,

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

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

From a practical perspective, Theorem 10 aids in showing that the only requirement for the main results of Section 2 to hold is solely the availability of any preliminary near optimal estimate $\hat{\tau}$ lying in a wider neighborhood of τ^0 in comparison to $\tilde{\tau}$. Specifically, consider any estimate satisfying

$$(\hat{\tau} - \tau^0) = O\big(\psi^{-2}\log(p \vee T)\big), \quad \text{w.p. } 1 - o(1). \tag{3.2}$$

Recall from the construction (A.2) of set $\mathcal{G}(u_T, 0)$, that u_T is the sequence that measures the outer bound of the distance with the underlying change point τ^0, i.e., in the case of $\hat{\tau}$ in (3.2), we have $u_T = c\psi^{-2}T^{-1}\log(p \vee T)$. Substituting this choice of u_T in Theorem 10 and simplifying resulting expressions, it may be observed that the resulting nuisance estimates $\hat{\mu}_{(j)}(\hat{\tau})$, $\hat{\gamma}_{(j)}(\hat{\tau})$, $j = 1, ..., p$ satisfy all requirements of Condition C. Note that the weaker bound (3.2) is the resulting localization error of the existing methods in the literature for a second order shift. This allows us to integrate these existing methods and develop a method that allows one to attain a sharp rate of estimation together with the inference properties described in Section 2. This is described below in Algorithm 1.

Algorithm 1: $O_p(\psi^{-2})$ estimation of τ^0 :

Step 1: Implement any estimator $\hat{\tau}$ 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,

$$\tilde{\tau} = \arg\min_{\tau\in\{1,...,(T-1)\}} Q(\tau, \hat{\mu}, \hat{\gamma})$$

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

The most direct estimator that can be utilized in Step 1 is that of a full grid search, i.e., searching over all combinations of $(\tau, \mu(\tau), \gamma(\tau))$,

$$\hat{\tau} = \arg\min_{\tau\in\{1,...,(T-1)\}} Q(\tau, \mu(\tau), \gamma(\tau))$$

Such full grid search estimators have been studied in several contexts and it can be shown via standard arguments, such as those in Lee et al. (2016) that they satisfy the weaker rate of estimation (3.2). These arguments are not repeated here to avoid redundancy. Other estimators from the literature that can be used in Step 1 of Algorithm 1 include the ones introduced in Bybee and Atchadé (2018); Wang et al. (2021) that aim to avoid a brute force search and save on computational time. Note that all of these estimators only come with the weaker bound (3.2) and do not provide the sharp rate of Theorem 1. Further, Algorithm 1 can be viewed as performing a refitting of the change point estimate in Step 2, which leads to improving the $O_p\big(\psi^{-2}\log(p\vee T)\big)$ rate to the attainable sharp rate $O_p(\psi^{-2})$ and thereby enabling inference on the change point as per the discussion of Section 2. This is summarized in the following Corollary.

Corollary 11 *Suppose Conditions A' and B hold and that $\big(\max_j \|\eta^0_{(j)}\|_2/(\psi\sqrt{s})\big)=O(1)$. Then, the nuisance estimates $\hat\mu_{(j)}$, and $\hat\gamma_{(j)}$, $j=1,...,p$, from Step 2 of Algorithm 1 satisfy all requirements of Condition C. Upon assuming Condition A, $\tilde\tau$ of Algorithm 1 satisfies the $O_p(\psi^{-2})$ rate of Theorem 1. Further assuming Condition D and (2.4), $\tilde\tau$ possesses the limiting distribution of Theorem 2 under a vanishing jump size regime. Finally, assuming Condition B', additionally yields the limiting distribution of Theorem 5 of the non-vanishing regime, together with the regime adaptation of Theorem 7.*

Algorithm 1 presents all necessary steps for a feasible implementation of the proposed methodology, while Corollary 11 summarizes estimation and inferential properties of the resulting change point estimate.

Next, we introduce a new computationally efficient estimator that can be used in Step 1 of Algorithm 1. The key idea is that a *single refitting* step leads to an improvement in the rate of estimation from (3.2) to $O_p(\psi^{-2})$. Similarly, upon backtracking one more step, one can further relax the required localization (3.2) to a nearly arbitrarily large interval around the change point parameter.

The twice iterative approach of the estimator to be considered is as follows. 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 arbitrary $\check\tau\in(0,1)$ (see, Condition E below) possess sufficient information, so that a single step update $\hat\tau=\tilde\tau(\check\mu,\check\gamma)$, moves into a near optimal neighborhood (nbd.) $O_p\big(\psi^{-2}\log(p\vee T)\big)$. This provides the intermediate relation (3.2). Next, proceeding as in Algorithm 1, this near optimal estimate $\hat\tau$, upon another refitting yields an improvement to optimality, specifically and $O_p(\psi^{-2})$ rate and thus ensuring that the results of Section 2 hold. More specifically, we shall establish that Algorithm 2 moves any starting value in this $o(T^{1-k})$ neighborhood into an optimal neighborhood in two iterations, i.e., $o(T^{1-k})$-nbd.$\xrightarrow{\text{Step1}}$ near optimal-nbd., $O_p(\psi^{-2}\log p)\xrightarrow{\text{Step2}}$ optimal-nbd., $O_p(\psi^{-2})$. Note the sequential improvement in the rate of convergence from initialization to Step 2. Moreover, the improvement to optimality occurs in exactly two iterations. The procedure is stated as Algorithm 2 below and is described visually in Figure 1.

The only additional requirement of Algorithm 2 in comparison to Algorithm 1 is the following mild condition imposed on the initializer $\check\tau$ of Algorithm 2.

Condition E (initializer): *Assume initializer $\check\tau$ of Algorithm 2 satisfies,*

$$(i)\ \check\tau\wedge(T-\check\tau)\geq c_uT\ell_T,\qquad (ii)\ |\check\tau-\tau^0|\leq\frac{c_u\kappa\ell_T}{s(\sigma^2\vee\phi)}T^{(1-k)},$$

Algorithm 2: $O_p(\psi^{-2})$ estimation of τ^0 :

(Initialize): Initialize $\check{\tau} \in \{1, ..., (T-1)\}$.

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,

$$\hat{\tau} = \arg\min_{\tau \in \{1,...,(T-1)\}} Q(\tau, \check{\mu}, \check{\gamma})$$

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,

$$\tilde{\tau} = \arg\min_{\tau \in \{1,...,(T-1)\}} Q(\tau, \hat{\mu}, \hat{\gamma})$$

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

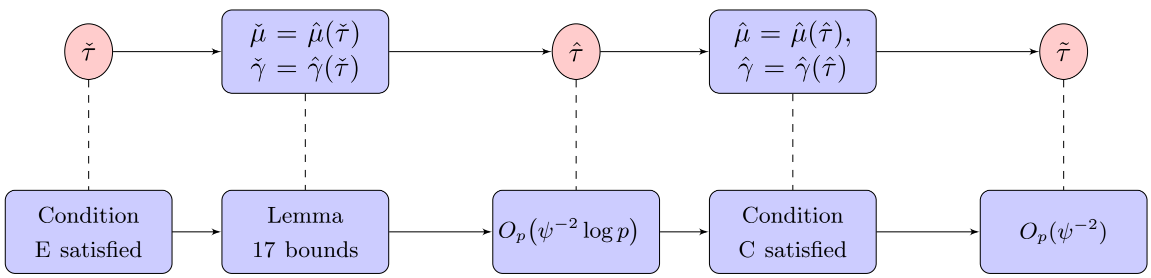

Figure 1: A schematic of the proposed Algorithm 2 and its underlying working mechanism.

for any constant $k > 0$.[3]

The first requirement of Condition E is clearly innocuous and simply requires a separation of the chosen initializer from the boundaries of the parametric space of the change point which is satisfied with any $(\check{\tau}/T) \in [c_{u1}, c_{u2}] \subset (0, 1)$. Regarding the second requirement, for simplicity consider the case when $\ell_T \geq c_u$, i.e., the true change (τ^0/T) lies in some bounded subset of $(0, 1)$, and the sparsity parameter is bounded above by a constant. Then, the requirement reduces to $|\check{\tau} - \tau^0| = o(T^{1-k})$, where the constant k is any arbitrarily small, but fixed value; in other words, the initializer may be in any arbitrary polynomial neighborhood $o(T^{(1-k)})$ of τ^0.

Remark 12 *(Computational Complexity)* The complexity of Algorithm 1 depends on the method utilized in Step 1, i.e., the implementation time scales as $O(Step1) + O(p^4 + T)$. For Algorithm 2, the corresponding time scaling is $O(T + p^4)$. Here $O(T)$ corresponds to the one-dimensional minimizations required to update the change point parameters, the remaining comes from the p Lasso optimizations carried out to update the coefficients μ, γ, with each Lasso optimization scaling as $O(p^3)$ via the algorithm LARS (see, page 443, Efron et al. (2004)). For comparison purposes, the complexity of a full grid search via the loss (1.4) would be $O(Tp^4)$, as these methods in principle attempt to simultaneously optimize the change point and coefficient parameters. Similar grid search methods are common in the change point literature, for e.g., Lee et al. (2016). We note that the added cost towards

3. Without loss of generality we assume $k < b$, where b is as defined in Condition A$'$.

computation of the confidence intervals (see, Algorithm 4 in Section 4) is observed to be negligible relative to the estimation process as it only requires Monte Carlo sampling.

A theoretically valid initializer $\check{\tau}$ in an very wide $o(T^{1-k})$ neighborhood of τ^0 can be obtained by a preliminary coarse grid search as follows: consider T^k equally separated values in $\mathcal{P} \subset \{1, ..., T\}$ forming a coarse grid of possible initializers, for any arbitrarily chosen, but fixed value k. Then, select the best fitting value $\check{\tau}$ for Algorithm 2, i.e., $\check{\tau} = \arg\min_{\tau \in \mathcal{P}} Q\big(\tau, \mu(\tau), \gamma(\tau)\big)$. Then, by leveraging arguments analogous to those developed for the proof of Theorem 1, or more generally similar to those available in the literature pertaining to grid search approaches for e.g., Bhattacharjee et al. (2017) and Lee et al. (2016),[4] it can be shown that this best fitting value is closest to the true change point τ^0, amongst the choices available in the coarse grid. As a consequence, by the pigeonhole principle it must be in an $o(T^{1-k})$ neighborhood of τ^0, and hence a valid initializer. A similar preliminary coarse grid search has also been heuristically utilized in Roy et al. (2017) in a different high dimensional model setting, in Kaul et al. (2019, 2021) for mean shifts, and most recently in McGonigle and Peng (2021). All simulation experiments in Section 4 consider a preliminary search grid of $\check{\tau} \in \{\lfloor 0.25 \cdot T \rfloor, \lfloor 0.5 \cdot T \rfloor, \lfloor 0.75 \cdot T \rfloor\}$ to select the initializer for Algorithm 2. The following Theorem establishes that $\hat{\tau}$ of Step 1 of Algorithm 2 lies in an $O\big(\psi^{-2} \log(p \vee T)\big)$ neighborhood of τ^0, i.e., satisfies (3.2).

Theorem 13 *Suppose Conditions A', B and E hold. Let $\hat{\tau}$ be the change point estimate in Step 1 of Algorithm 2. Then, for sufficiently large T, we have,*

$$\psi^2 (1 + \nu^2)^{-1} (\sigma^2 \vee \phi)^{-2} \kappa^2 \big|\hat{\tau} - \tau^0\big| \le c_u \log(p \vee T) \tag{3.3}$$

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

Theorem 13 establishes that the behavior of Step 2 in Algorithm 2 is now identical to that Algorithm 1. Specifically, the resulting estimate $\tilde{\tau}$ satisfies the desirable estimation and inference properties developed in Section 2. This is summarized in the following Corollary.

Corollary 14 *Suppose the conditions of Corollary 11 hold and additionally assume that the initializer $\check{\tau}$ of Algorithm 2 satisfied Condition E. Then, $\tilde{\tau}$ of Algorithm 2 satisfies the sharp rate of estimation of Theorem 1 and the possesses the limiting distributions of Theorem 2 and Theorem 5 together with their inter-relationship of Theorem 7.*

4. Implementation details and performance evaluation

Construction of confidence intervals requires evaluation of the quantile q_α^{nv} of the argmax of the two sided random walk. This is computed by simulating 3000 realizations of this distribution and its Monte Carlo approximation. This approximation requires estimation of the drift and variance parameters σ_1^2, σ_2^2 and $\sigma_1^{*2}, \sigma_2^{*2}, \bar{\sigma}_1^2, \bar{\sigma}_2^2$, as defined in Remark 8, as well as identification of $\mathcal{L}$ of Condition B'. The following discussion provides the methods employed for these calculations.

4. Note that a full grid search is simply the case where $\mathcal{P} = \{1, .., (T-1)\}$.

4.1 Implementation details

Computation of asymptotic variances and negative drifts. Step 2 of Algorithm 1 or 2 yields ℓ_1 regularized estimates $\hat{\mu}_{(j)}$ and $\hat{\gamma}_{(j)}$. In order to alleviate finite sample regularization biases we undertake a supplemental step of refitting these coefficient estimates as ordinary least squares on estimated non-zero indices, i.e., let $\hat{S}_{1j} = \{k\,;\,\hat{\mu}_{(j)k} \neq 0\}$ and $\hat{S}_{2j} = \{k\,;\,\hat{\gamma}_{(j)k} \neq 0\}$, $j = 1, ..., p$. Then define,

$$\tilde{\mu}_{(j)} = \arg\min_{\substack{\mu_{(j)}\in\mathbb{R}^p;\\ \mu_{(j)k}=0;\\ \forall k\in\hat{S}_{1j}^c}} \frac{1}{\tilde{\tau}} \sum_{t=1}^{\tilde{\tau}} \big(z_{tj} - z_{t,-j}^T \mu_{(j)}\big)^2, \quad j = 1, ..., p$$

and symmetrically define $\tilde{\gamma}_{(j)}$, $j = 1, ..., p$. It is known that such refitted estimates preserve the rate of convergence of the regularized version while reducing finite sample biases, see, e.g. Belloni et al. (2011). The jump sizes $\xi_{2,2}$ and ψ are then estimated using these refitted coefficient vectors, i.e., let $\tilde{\eta}_{(j)} = \tilde{\mu}_{(j)} - \tilde{\gamma}_{(j)}$, then $\tilde{\xi}_{2,2} = \big(\sum_{j=1}^p \|\tilde{\eta}_{(j)}\|_2^2\big)^{1/2}$ and $\tilde{\psi} = \tilde{\xi}_{2,2}/\sqrt{p}$.

Recall the definition of drift parameter $\sigma_1^2 = \xi_{2,2}^{-2} \sum_{j=1}^p \eta_{(j)}^{0T} \Sigma_{-j,-j} \eta_{(j)}^0$, and similar for σ_2^2 from Remark 8. Plug in versions are computed by utilizing the above described $\tilde{\xi}_{2,2}$ and $\tilde{\eta}_{(j)}$, $j = 1, ..., p$. The covariances in the above calculation are chosen as the sample covariances $\tilde{\Sigma}$ and $\tilde{\Delta}$ computed on the binary partition induced by $\tilde{\tau}$. Since we are not interested in the estimation of the covariances themselves but instead the quardratic form described above, thus utilizing the sample covariances is effectively identical to utilizing refitted covariances on the adjacency matrix defined by the jump parameters $\tilde{\eta}_{(j)}$, in turn making this shortcut valid despite potential high dimensionality.

Finally consider asymptotic variances $\bar{\sigma}_1^2, \bar{\sigma}_2^2$. Plug in estimates are infeasible here since no closed form expressions are available for these variances. Instead, define

$$\tilde{\varepsilon}_{tj} = \begin{cases} z_{tj} - z_{t,-j}^T \tilde{\mu}_{(j)}, & t = 1, ..., \tilde{\tau} \\ z_{tj} - z_{t,-j}^T \tilde{\gamma}_{(j)}, & t = \tilde{\tau} + 1, ..., T, \end{cases}$$

Next, recall from Remark 8, $\bar{\sigma}_1^2$ is defined as $\mathrm{var}\big[p^{-1/2} \sum_{j=1}^p \varepsilon_{tj} z_{t,-j}^T \eta_{(j)}^0 - \eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T \eta_{(j)}^0\big]$. Thus, one can obtain $\tilde{\tau}$ predicted realizations as $\tilde{\zeta}_t^* = p^{-1/2} \sum_{j=1}^p \tilde{\varepsilon}_{tj} z_{t,-j}^T \tilde{\eta}_{(j)} - \tilde{\eta}_{(j)}^T z_{t,-j} z_{t,-j}^T \tilde{\eta}_{(j)}$, $t = 1, ..., \tilde{\tau}$. The parameter $\bar{\sigma}_1^2$ is then estimated as the sample variance of these realizations,

$$\tilde{\bar{\sigma}}_1^2 = \frac{1}{\tilde{\tau}} \sum_{j=1}^{\tilde{\tau}} \Big(\tilde{\zeta}_t^* - \bar{\tilde{\zeta}}_t^*\Big)^2.$$

The parameter $\bar{\sigma}_2^2$ is approximated analogously as the sample variance of predicted realizations $\tilde{\zeta}_t^*$ from the post binary partition $\tilde{\tau} + 1, ..., T$.

Empirically fitting the distribution law $\mathcal{L}$: Here we illustrate the process employed to empirically fit a distribution $\mathcal{L}$ of Condition B'. The distribution under question is that of the sequence $p^{-1} \sum_{j=1}^p \big\{2\varepsilon_{tj} z_{t,-j}^T \eta_{(j)}^0 - \eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T \eta_{(j)}^0\big\}$ in the limit. As described above, the available data z and plug in estimates of underlying parameters allow one to obtain predicted realizations from this distribution. Figure 2 provides an example of the centered and scaled distribution of these realizations when the data generating process is Gaussian.

Algorithm 3: Empirically fitting a centered and scaled χ^2_k to the distribution law $\mathcal{L}$.

Step 1: Obtain refitted estimates $\tilde{\mu}_{(j)}$, $\tilde{\gamma}_{(j)}$ and $\tilde{\eta}_{(j)}$, $j = 1, ..., p$ and obtain predicted realizations,

$$\tilde{\zeta}_t = p^{-1} \sum_{j=1}^{p} \big\{ 2\tilde{\varepsilon}_{tj} z_{t,-j}^T \tilde{\eta}_{(j)} - \tilde{\eta}_{(j)}^T z_{t,-j} z_{t,-j}^T \tilde{\eta}_{(j)} \big\}, \quad t = 1, ..., T. \tag{4.1}$$

Step 2: Piecewise center and scale the predicted realizations $\tilde{\zeta}_t$, i.e.

$$\tilde{\zeta}_t^* = \begin{cases} (\tilde{\zeta}_t - \bar{\zeta}_{t1})/sd_1, & t = 1, ..., \tilde{\tau} \\ (\tilde{\zeta}_t - \bar{\zeta}_{t2})/sd_2, & t = \tilde{\tau} + 1, ..., T. \end{cases}$$

Here $\bar{\zeta}_{t1}, \bar{\zeta}_{t2}$ and sd_1, sd_2 are the piecewise sample means and standard deviations, respectively.

Step 3: Consider a negative centered and scaled χ^2_k distribution with k degrees of freedom, i.e., $X = -(\chi^2_k - k)/\sqrt{(}2k)$ and utilize the Kolmogorv-Smirnov (K-S) goodness of fit test to check for the empirical fit between X and the realizations $\tilde{\zeta}_t^*$, $t = 1, ..., T$.

Step 4: Repeat Step 3 on a grid of values for the degrees of freedom $k \in \{1, 2, 3....\}$ and choose k as the maximizing value of the p-value of the K-S test.

The following two key observations are in order: First, given the sub-Gaussian assumption of Condition B, the distribution under investigation must be sub-exponential. This observation allows considerable reduction of the potential distributions to be tested to a sub-class of well known sub-expoential distributions. Next, note that the second part of the sequence under consideration is a quadratic form, thus it induces a skewness in the distribution with the underlying skewness diminishing with a decreasing jump size[5]. Since this quadratic form appears with a negative sign in the distribution of interest, thus the skewness appears through a larger left tail.

In view of the above, we consider a negative centered and scaled chi-square distribution as an empirical fit. The negative sign switches the right skew of a chi-square to a left skew. Further, an increasing degrees of freedom parameter of this chi-square allows one to fit a distribution from complete left skew to perfect symmetry. Specifically, we utilize Algorithm 3, where the degrees of freedom of this chi-square distribution are selected so as to maximize the p-value of the Kolmogorov-Smirnov goodness of fit test, i.e., so as to provide the best fitting chi-square approximation to the underlying distribution. An illustration of the fitted distribution using Algorithm 3 is provided in Figure 2.

4.2 Numerical results

Evaluation of Regime Adaptivity: Next, we construct confidence intervals as described in (2.10). In view of Theorem 7 and the ensuing discussion, these confidence intervals are

5. Recall from (2.7), variance of the quadratic form is $O(\psi^4)$, whereas the first symmetric part is $O(\psi^2)$.

Algorithm 4: Overall Inference Procedure for τ^0

Step 1: Implement Algorithm 1 or Algorithm 2 to obtain estimate $\tilde{\tau}$.

Step 2: Obtain quantile q_α^{nv} at any given coverage α of limiting distribution of the non-vanishing regime as follows:

a: Obtain estimates for jump size $(\tilde{\psi})$, drift $(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2)$ and variance $(\tilde{\tilde{\sigma}}_1^2, \tilde{\tilde{\sigma}}_2^2)$ as described in Section 4.1.

b: Implement Algorithm 3 to identify incremental distribution $\mathcal{L}$.

c: Repeatedly sample the distribution $\arg\max_r \mathcal{C}_\infty(r)$ with parameters and distribution identified in Step 2a and Step 2b, respectively. Obtain the Monte Carlo approximation of q_α^{nv} as the symmetric sample quantile that cuts-off central $(1-\alpha)$ proportion of the obtained realizations.

(Output): Confidence interval: $\left[\tilde{\tau} \pm q_\alpha^{nv}\right]$

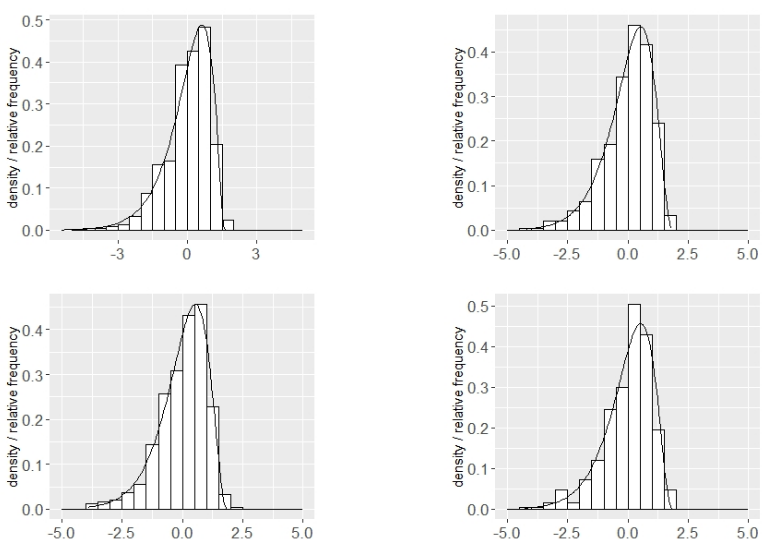

Figure 2: Histograms of $\tilde{\zeta}_t^*$, along with superimposed densities of negative centered and scaled χ_k^2 distributions with df identified via K-S goodness of fit test. *Top panels:* $p = 25$, and $p = 50$, (fitted distributions with df=$5, 7$, respectively. p-values of K-S test: 0.90, 0.95, respectively.) *Bottom panels:* $p = 150$ and $p = 250$. (fitted distributions with df=$7, 8$, respectively. p-values of K-S test: 0.99, 0.82, respectively.)

Jump Size ψ	Ratio of quantiles $\left(q_\alpha^{nv}/(\sigma_1^{*2}\sigma_1^{-4}\psi^{-2}q_\alpha^v)\right)$			
	$\alpha=0.2$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
2.00	0	0	0	0
1.74	0	0	0	0
1.48	0	0	0	0
1.22	0	0.789	0.543	0.304
0.97	0.811	0.493	0.679	0.571
0.71	0.877	0.800	0.735	0.721
0.45	0.898	0.873	0.752	0.717
0.20	1.032	1.025	0.951	0.945

Jump Size ψ	Ratio of quantiles $\left(q_\alpha^{nv}/(\sigma_1^{*2}\sigma_1^{-4}\psi^{-2}q_\alpha^v)\right)$			
	$\alpha=0.2$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
2.00	0	0	0	0.807
1.74	0	0	1.094	0.613
1.48	0	1.153	0.795	0.891
1.22	1.298	0.789	1.087	0.914
0.97	0.811	0.986	1.019	1.143
0.71	0.877	1.066	1.102	0.927
0.45	0.898	0.873	0.903	0.886
0.20	0.963	0.961	1.009	0.945

Table 1: Comparison of quantiles of distributions in the non-vanishing vs the normalized version of those of the vanishing regime. Expected behavior: $\left(q_\alpha^{nv}/(\sigma_1^{*2}\sigma_1^{-4}\psi^{-2}q_\alpha^v)\right) \to 1$ as $\psi \to 0$,. *Left panel:* Incremental distribution $\mathcal{L}$: centered and scaled χ_k^2, $k=5$, *Right panel*: Incremental distribution $\mathcal{L}:\mathcal{N}(0,1)$

regime adaptive, a property illustrated next. Specifically, the observable consequence of Theorem 7 that allows regime adaptation is that for any significance level α, we have the asymptotic equivalence $q_\alpha^{nv} \asymp \sigma_1^{*2}\sigma_1^{-4}\psi^{-2}q_\alpha^v$, as $\psi \to 0$, where $q_\alpha^{nv}, q_\alpha^v$ are the quantiles of the distributions in the non-vanishing and vanishing regime, respectively. In other words, one should observe the ratio $q_\alpha^{nv}/\left(\sigma_1^{*2}\sigma_1^{-4}\psi^{-2}q_\alpha^v\right) \to 1$ as $\psi \to 0$.

We consider eight equally separated and decreasing values of $\psi \in \{2,..,0.2\}$. Two cases of the form of the incremental distribution $\mathcal{L}$ are considered: (1) a centered and scaled χ_k^2 distribution in keeping with earlier discussion, and (2) a standard normal distribution. For a clear illustration, we set the drift and variance parameters as $\sigma_1^2=\sigma_2^2=\sigma_1^{*2}=\sigma_2^{*2}=1$. The variance parameter of the random walk process is accordingly set to $\bar{\sigma}_1^{*2}=\sigma_2^{*2}=1/4$ (as dictated by Theorem 7). Quantiles for the non-vanishing case are computed as described via a Monte Carlo simulation of 3000 sample paths of the underlying random walk process, whereas quantiles of the vanishing case are computed via its cumulative distribution function. Results are shown in Table 1, which confirm the theoretical assertion. Specifically, the ratio under consideration approaches 1 for all considered values of α as the jump size vanishes. This is also illustrated visually in Figure 3.

Evaluation of Estimation and Inference Properties: Next, an evaluation of the proposed estimation and inference methodology is provided on synthetic data. We note that while there is no existing inference methodology, however in order to provide a benchmark from an estimation perspective, we also compare the performance of Algorithm 2 to the method of Bybee and Atchadé (2018). The method being based on the likelihood criterion, and optimized via the Simulated Annealing algorithm. Its implementation is carried out via the authors developed `R`-package `ChangepontsHD` with recommended tuning settings and is referred to as *BA 2018* in the following.

In all to follow, w_t, x_t of model (1.1) are independent, p-dimensional, Gaussian r.v.'s with distinct covariance structures. Specifically, we set $w_t \sim \mathcal{N}(0,\Sigma)$, $t=1,...,\tau^0$ and $x_t \sim \mathcal{N}(0,\Delta)$, $t=\tau^0+1,...,T$. The observation period T is set to $\{300,400,500\}$, the dimension p to $\{25,50,150,250\}$ and the relative location of the change point $(\tau^0/T) \in \{0.2,0.4,0.6,0.8\}$. All computations are carried out in `R`, and all Lasso optimizations of 3.1 are carried out using the `glmnet` package. In all cases, the initializer for Algorithm 2 is selected via a preliminary search grid of $\check{\tau} \in \{0.25,0.5,0.75\}$ as described in the discussion ensuing Condition E. The significance level is set to $\alpha \in \{0.05,0.01\}$ in all cases.

Structure of the covariance matrices: To construct Σ, we consider a Toeplitz type matrix Γ with the $(l,m)^{th}$ component set as $\Gamma_{(l,m)} = \rho^{|l-m|^a}$, $l,m = 1,...,p$. We set $\rho = 0.4$ and $a = 1/\log s$, where s specified below.[6]. Then, set $\Sigma =\cdot A\cdot\Gamma$, where $\cdot$ denotes a componentwise product. The matrix A is constructed as a symmetric block diagonal matrix with alternating signs $\{-1,1\}$ within each block of size $s\times s$. This allows both positive and negative correlations in Σ and also induces a sparsity structure with each row and column having s non-zero components. The post-change point covariance Δ is chosen as a banded matrix. The non-zero correlations for each row and column of Δ are chosen as a sequence of s equally spaced values between $\{\rho_2 = 0.5,...,0\}_{s\times 1}$. The sparsity of both Σ and Δ are set at 15% of the dimension size.

Selection of tuning parameters: The tuning parameters λ_j, $j = 1,...,p$ used to obtain ℓ_1 regularized mean estimates are selected based on a BIC type criterion. Specifically, we set $\lambda_j = \lambda$, $j = 1,...,p$, and evaluate $\hat\mu_{(j)}(\lambda)$, and $\hat\gamma_{(j)}(\lambda)$ over an equally spaced grid of seventy five values in the interval $(0,1)$. Upon letting $\hat S = \cup_{j=1}^p\big[\{k;\ \hat\mu_{(j)}k \neq 0\}\cup\{k;\ \hat\gamma_{(j)}k\neq 0\}\big]$ we evaluate the criteria,

$$BIC(\lambda,\tau) = \sum_{t=1}^{\tau}\sum_{j=1}^{p}\big(z_{tj} - z_{t,-j}^T\hat\mu_{(j)}(\lambda)\big)^2 + \sum_{t=\tau+1}^{T}\sum_{j=1}^{p}\big(z_{tj} - z_{t,-j}^T\hat\gamma_{(j)}(\lambda)\big)^2 + |\hat S|\log T. \quad (4.2)$$

Set λ as the minimizer of $BIC(\lambda,\check\tau)$, and $BIC(\lambda,\hat\tau)$. for Step 1 and Step 2 of Algorithm 2.

We note some additional pertinent aspects. The problem under consideration relies on recovery of two $p\times p$ matrices, the number of free parameters in each being $p(p-1)/2$. Second, high dimensionality in the considered framework is characterized as $\log p = o(T\ell_T)$, where ℓ_T is the separation of τ^0 from the parametric boundary. The largest sample size is $T = 500$ and closest to boundary change point is at $\lfloor 0.2\cdot T\rfloor$, consequently the effective sample size here is 100. The appropriate comparison of dimensionality p is with these effective sample sizes. Finally, the considered problem is of estimation and inference on τ^0, which is discrete, and so are the estimates $\tilde\tau$ as well as the associated confidence intervals This is distinct from a conventional inference problem under a continuous parametric space. Specifically, in the construction (2.10) the quantiles of the limiting distribution are discrete values and can very well be identically zero. Effectively, a confidence interval can also be identical to a point estimate, while additionally providing an uncertainty measurement. Technically, there is also the trivial asymptotic regime, where the signal is so large that the asymptotic distribution becomes degenerate, see, e.g. Page 6 of Bhattacharjee et al. (2019).

The following metrics are used for assessing the performance of the methodology: bias $(\big|E(\hat\tau-\tau^0)\big|)$, root mean squared error (RMSE, $E^{1/2}(\hat\tau-\tau^0)^2$), coverage (relative frequency of the number of times τ^0 lies in the confidence interval) and the average margin of error (average over replicates of the margin of error of each confidence interval). Estimation performance in Table 2 are based on 100 replications. All other inference results reported metrics are based on 500 replications. Results are provided in Tables 2, 3 and 4.

Discussion of the Results: We note that estimates obtained via the proposed method exhibit very little bias and are tightly distributed, as seen from the RMSE metric. Further, an expected deterioration is present as p increases, which is compensated by an increasing

6. We choose the $\log s$ root of $|l-m|$ so as to somewhat preserve the magnitude of correlations and in turn condition dependencies

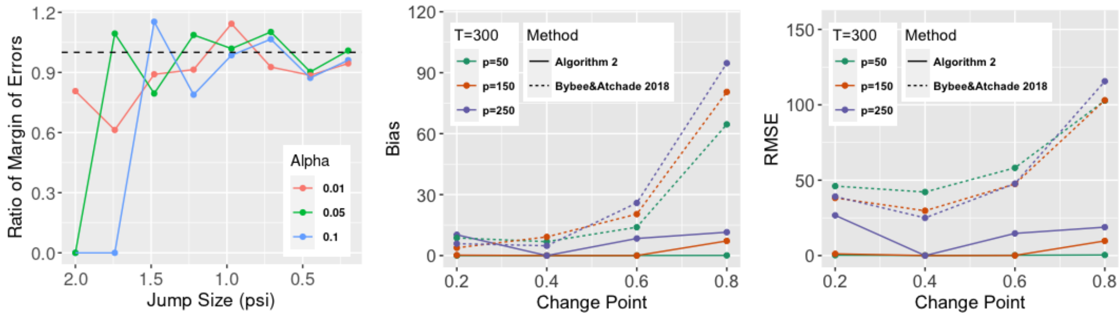

Figure 3: *Left Panel*: Illustration of regime adaptivity. At all considered values of $\alpha = 0.1, 0.05, 0.01$, the ratio of margin of errors $\left(q_{\alpha}^{nv}/(\sigma_1^{*2}\sigma_1^{-4}\psi^{-2}q_{\alpha}^{v})\right) \to 1$ under the non-vanishing vs vanishing regime, as the jump size $\to 0$. *Center and Right Panel*: Comparison of estimation results (bias, RMSE) with BA 2018 at $T = 300$. x-axis: change point in a fraction scale (τ^0/T).

p	$\frac{\tau^0}{T}$	$T=300$		$T=400$		$T=500$	
		Algorithm 2	BA 2018	Algorithm 2	BA 2018	Algorithm 2	BA 2018
		Bias (RMSE)	Bias (RMSE)	Bias (RMSE)	Bias (RMSE)	Bias (RMSE)	Bias (RMSE)
50	0.2	0.02 (0.24)	8.68 (46.08)	0.04 (0.20)	13.28 (74.07)	0.01 (0.10)	0.22 (63.35)
	0.4	0.01 (0.10)	6.85 (42.16)	0.01 (0.17)	8.55 (75.91)	0.00 (0.00)	7.41 (80.87)
	0.6	0.04 (0.20)	13.96 (58.19)	0.03 (0.17)	21.85 (74.64)	0.00 (0.14)	5.51 (60.06)
	0.8	0.08 (0.40)	64.58 (102.80)	0.00 (0.00)	64.70 (126.40)	0.05 (0.22)	61.98 (148.65)
150	0.2	0.21 (1.16)	3.84 (38.23)	0.02 (0.14)	6.16 (53.50)	0.00 (0.00)	1.20 (57.25)
	0.4	0.00 (0.00)	9.18 (29.74)	0.02 (0.14)	10.30 (32.85)	0.00 (0.00)	11.88 (38.58)
	0.6	0.01 (0.10)	20.41 (47.47)	0.01 (0.10)	26.86 (72.91)	0.00 (0.00)	26.30 (80.22)
	0.8	7.21 (9.73)	80.51 (103.00)	3.34 (7.40)	95.76 (130.58)	0.08 (0.28)	71.99 (120.55)
250	0.2	10.23 (26.70)	5.87 (39.17)	1.58 (10.74)	10.82 (62.55)	0.52 (5.00)	7.95 (59.01)
	0.4	0.00 (0.00)	4.84 (24.99)	0.00 (0.00)	3.97 (31.54)	0.00 (0.00)	10.51 (36.36)
	0.6	8.41 (14.72)	25.91 (47.70)	0.03 (0.17)	16.39 (44.49)	0.00 (0.00)	33.36 (81.01)
	0.8	11.50 (18.85)	94.69 (115.61)	12.97 (15.69)	110.63 (142.27)	8.50 (13.10)	92.32 (132.21)

Table 2: Estimation performance of proposed Algorithm 2 and BA 2018. Reported results based on 100 Monte Carlo replications.

T. Results are observed to be nearly uniformly better than the benchmark method of Bybee and Atchadé (2018), as also visualized in Figure 3. This is agreement with the estimation result of Theorem 1 yielding a sharper rate of estimation.

From an inference perspective, the results are in agreement with our theoretical developments. The coverage is observed to deviate below the nominal level only for the case of large p and small T (for e.g., cases $p = 250, T = 300, \tau^0 = 0.8$). An increase in the sampling period compensates for this and improves coverage to or above the nominal level. The deviation of coverage above the nominal level (conservative coverage) is an expected feature under change point inference particularly due to the inherent discreteness of the problem. This feature is observed for large values of p, T (driven by p). The reason for this is as follows. Recall the distribution of the change point estimator is driven by a random walk with increments as $\mathcal{L}(-\psi^2\sigma^2, \bar{\sigma}^2)$, with the negative mean causing the negative drift of the process. By the design of our simulation, the sparsity s of the underlying precision matrices is growing with p, consequently the underlying parameters of these increments are also varying. In particular, the drift relative to the standard deviation of these increments (signal/noise) is observed to be growing, i.e., $(\psi^2\sigma^2/\bar{\sigma})$ is increasing with p. In effect,

T	p	$\tau^0 = \lfloor T\cdot 0.6 \rfloor$			$\tau^0 = \lfloor T\cdot 0.8 \rfloor$		
		bias (RMSE)	$\alpha = 0.05$	$\alpha = 0.01$	bias (RMSE)	$\alpha = 0.05$	$\alpha = 0.01$
			Coverage (av. ME)			Coverage (av. ME)	
300	25	0.018 (0.279)	0.95 (0.190)	0.99 (0.910)	0.088 (0.400)	0.93 (0.198)	0.98 (0.882)
300	50	0.022 (0.265)	0.97 (0.008)	0.99 (0.456)	0.098 (0.417)	0.93 (0.020)	0.96 (0.430)
300	150	0.050 (0.688)	0.98 (0.002)	0.98 (0.024)	7.342 (9.891)	0.35 (0.314)	0.35 (0.708)
300	250	6.768 (13.32)	0.65 (0.190)	0.66 (0.362)	11.82 (14.71)	0.16 (0.282)	0.16 (0.830)
400	25	0.030 (0.349)	0.95 (0.182)	0.99 (0.944)	0.034 (0.332)	0.95 (0.212)	0.99 (0.920)
400	50	0.012 (0.155)	0.98 (0.004)	0.99 (0.434)	0.058 (0.279)	0.94 (0.020)	0.96 (0.438)
400	150	0.006 (0.077)	0.99 (0)	0.99 (0.006)	2.664 (6.291)	0.71 (0.070)	0.72 (0.238)
400	250	0.480 (3.865)	0.97 (0.012)	0.98 (0.030)	16.87 (17.85)	0.07 (0.518)	0.07 (1.038)
500	25	0.008 (0.261)	0.95 (0.162)	0.99 (0.958)	0.052 (0.379)	0.93 (0.200)	0.98 (0.940)
500	50	0.002 (0.118)	0.98 (0.002)	0.99 (0.536)	0.028 (0.179)	0.96 (0.004)	0.99 (0.536)
500	150	0 (0)	1 (0)	1 (0.010)	0.146 (1.017)	0.93 (0.002)	0.94 (0.066)
500	250	0 (0)	1 (0)	1 (0)	13.23 (16.97)	0.30 (0.338)	0.30 (0.718)

Table 3: Simulation results for $\tau^0 \in \{\lfloor T\cdot 0.6 \rfloor, \lfloor T\cdot 0.8 \rfloor\}$ based on 500 replicates. Bias, RMSE and av.margin of error rounded to three decimals, coverage rounded to two decimals.

T	p	$\tau^0 = \lfloor T\cdot 0.2 \rfloor$			$\tau^0 = \lfloor T\cdot 0.4 \rfloor$		
		bias (RMSE)	$\alpha = 0.05$	$\alpha = 0.01$	bias (RMSE)	$\alpha = 0.05$	$\alpha = 0.01$
			Coverage (av. ME)			Coverage (av. ME)	
300	25	0.016 (0.245)	0.95 (0.036)	0.98 (0.772)	0.012 (0.245)	0.95 (0.143)	0.99 (0.884)
300	50	0.032 (0.219)	0.97 (0)	0.97 (0.116)	0.006 (0.134)	0.98 (0)	0.99 (0.274)
300	150	2.294 (12.89)	0.93 (0)	0.93 (0.014)	0 (0.063)	0.99 (0)	0.99 (0.002)
300	250	12.97 (32.051)	0.72 (0)	0.72 (0)	0.534 (4.560)	0.97 (0.008)	0.98 (0.016)
400	25	0.008 (0.253)	0.95 (0.024)	0.99 (0.824)	0.014 (0.279)	0.94 (0.088)	0.99 (0.898)
400	50	0.016 (0.155)	0.97 (0)	0.98 (0.102)	0.006 (0.100)	0.99 (0)	0.99 (0.324)
400	150	0.086 (1.661)	0.99 (0)	0.99 (0)	0 (0)	1 (0)	1 (0.002)
400	250	2.886 (18.42)	0.93 (0)	0.93 (0)	0 (0)	1 (0)	1 (0)
500	25	0.002 (0.272)	0.95 (0.018)	0.99 (0.834)	0.026 (0.326)	0.94 (0.068)	0.99 (0.950)
500	50	0.006 (0.118)	0.98 (0)	0.98 (0.098)	0.004 (0.110)	0.99 (0)	0.99 (0.306)
500	150	0.006 (0.077)	0.99 (0)	0.99 (0)	0.002 (0.045)	0.99 (0)	0.99 (0)
500	250	0.042 (0.475)	0.98 (0)	0.98 (0)	0 (0)	1 (0)	1 (0)

Table 4: Simulation results for $\tau^0 \in \{\lfloor T\cdot 0.2 \rfloor, \lfloor T\cdot 0.4 \rfloor\}$ based on 500 replicates. Bias, RMSE and av.margin of error rounded to three decimals, coverage rounded to two decimals.

the associated random walk process is dropping to negative infinity more quickly, i.e., the distribution of the change point estimator is concentrating on the the true change point parameter. The consequence is that for large enough values of $(\psi^2\sigma^2/\bar{\sigma})$, the quantiles at both considered levels of significance are identically zero, i.e., the interval is a point estimate. This in turn causes the coverage in some cases to be observed as exactly one, moreover, makes it more difficult to distinguish between higher significance levels.

5. Age Evolving Associations of the Gut Microbiome

Microbiome studies are becoming increasingly important, due to recent findings on interactions of human microbiota with several human health outcomes for e.g., (Sharma and Tripathi, 2019; Svoboda, 2020), which has in turn led to considerable scientific interest in the related area of probiotic pharmaceuticals. Large scale microbiome data have become available in the last decade and are obtained by 16s rRNA sequencing technology. The resulting data correspond to operational taxonomic units (OTUs) which represent counts of observed microbial taxa identified by their genetic signature. For our analyses we consider the publicly available global human gut microbiome data of Yatsunenko et al. (2012).

It has been discussed in the literature that the gut microbiome undergoes a significant transformation from infancy/adolescence to adulthood. This transition point is often determined based on domain knowledge or other significant life events. Lozupone et al. (2013) suggest this transition age at around two years due to a switch over to solid food. This cutoff age has also been employed in Kaul et al. (2017a) for geographical classification of subjects based on their microbiota. Lane et al. (2019) suggest that such a transition point may occur well into adolescence of an individual, due to various social interactions, including those with siblings, early exposure to antibiotics amongst others. We aim to pursue this question quantitatively to estimate this transition point based on microbiome data from a second order perspective. We employ the model in (1.1), and estimate the transition age in the second order association structure of the taxa.

The Global Gut data set contains measurements of individuals from several geographical locations, with an associated age variable distributed over (0.08 years, 57 years). Our analysis is carried out at the second to finest, i.e., the *genus* level of bacterial taxonomy. We subset the analysed set of *genera* by retaining only those present in at least 35% of the samples. This limits the number of *genera* to $p = 166$ for model (1.1), with $T = 490$ observations. A further pre-processing of the data set is carried out via a log-relative abundance transformation of the raw OTU data, in order to switch over from a count to a continuous scale. The reference group chosen for this transformation is *Bifidobacterium*[7] due to it being a highly observed taxa which is present in all analyzed samples. This transformation is motivated by the compositional structure of the data set, see, e.g., Aitchison (1982) and is often adopted in the microbiome literature, see, e.g., Kaul et al. (2017b,a).

To study the age evolution of the association structure, all specimens are first sorted according to the age variable. Model (1.1) is then implemented with Algorithm 2 of Section 3, with a preliminary search grid over $\check{\tau} \in \{0.25, 0.5, 0.75\}$, which is the same as that used in the simulation studies of Section 4. All other computations such as tuning parameters selection, drift and asymptotic variance computation are as described in Section 4.

7. Phylogeny: *Bacteria*→*Actinobacteria*→*Actinobacteria* →*Bifidobacteriales*→*Bifidobacteriaceae*→*Bifidobacterium*

The estimated change point and the corresponding confidence intervals are obtained in the integer scale associated with index numbers of observations $\{1...., T\}$. We choose the significance level at $\alpha = 0.05, 0.01$, i.e., a coverage of 95% and 99%, respectively. These estimated values are then mapped back to the age variable to obtain the transition point in the age scale. The results of our analyses are discussed below.

The estimated change point is $\tilde{\tau} = 246$, with a jump size $\hat{\psi} = 0.49$. Upon mapping back to the age variable yields a transition age of $15yrs$. Confidence intervals are constructed under both vanishing and non-vanishing regimes and presented for both the index level and the age level in Table 5. At a coverage level of 99%, and under the vanishing jump regime, the associated confidence interval at the index level is $[243.31, 248.68]$, which yields an interval $[14yrs, 15rs]$ for the age of transition. The non-vanishing interval at the same coverage is found to be $[243, 249]$, at the index level and the same upon mapping to the age level. These results provide quantitative evidence to the hypothesis of gut microbial evolution over age and also appear to support the findings of Lane et al. (2019) from a second order shift in associations perspective. From a change point perspective, both non-vanishing and vanishing intervals appear very close to each other, which again supports our theoretical result on regime adaptation given the small jump size in this application.

	$\alpha = 0.05$		$\alpha = 0.01$	
	Vanishing	Non-Vanishing	Vanishing	Non-vanishing
Index level	[244.42, 247.57]	[244, 248]	[243.31, 248.68]	[243, 249]
Age level	[15, 15]	[15, 15]	[14, 15]	[14, 15]

Table 5: Confidence intervals under vanishing and non-vanishing jump size regimes at 95% and 99% coverage. Intervals presented at both index level and corresponding age level.

References

John Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982.

Jaromir Antoch, Marie Husova, and Noel Veraverbeke. Change-point problem and bootstrap. *Journaltitle of Nonparametric Statistics*, 5(2):123–144, 1995.

Alexander Aue, Siegfried Hörmann, Lajos Horváth, Matthew Reimherr, et al. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087, 2009.

Valeriy Avanesov, Nazar Buzun, et al. Change-point detection in high-dimensional covariance structure. *Electronic Journal of Statistics*, 12(2):3254–3294, 2018.

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.

Peiliang Bai, Abolfazl Safikhani, and George Michailidis. Multiple change points detection in low rank and sparse high dimensional vector autoregressive models. *IEEE Transactions on Signal Processing*, 68:3074–3089, 2020.

Matteo Barigozzi, Haeran Cho, and Piotr Fryzlewicz. Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics*, 206(1): 187–225, 2018.

Sumanta Basu, William Duren, Charles R Evans, Charles F Burant, George Michailidis, and Alla Karnovsky. Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data. *Bioinformatics*, 33(10):1545–1553, 2017.

A Belloni, V Chernozhukov, and C Hansen. Inference on treatment effects after selection amongst high-dimensional controls. arxiv, 2011. forthcoming. *The Review of Economic Studies*, 2014.

Alexandre Belloni, Victor Chernozhukov, and Lie Wang. Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806, 2011.

Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Common change point estimation in panel data from the least squares and maximum likelihood viewpoints. *arXiv preprint arXiv:1708.05836*, 2017.

Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Change point estimation in panel data with temporal and cross-sectional dependence. *arXiv preprint arXiv:1904.11101*, 2019.

Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Change point estimation in a dynamic stochastic block model. *The Journal of Machine Learning Research*, 21(1):4330–4388, 2020.

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

Melanie Birke and Holger Dette. A note on testing the covariance matrix for large dimension. *Statistics & probability letters*, 74(3):281–289, 2005.

Leland Bybee and Yves Atchadé. Change-point computation for large graphical models: A scalable algorithm for gaussian graphical models with change-points. *The Journal of Machine Learning Research*, 19(1):440–477, 2018.

Haeran Cho and Claudia Kirch. Bootstrap confidence intervals for multiple change points based on moving sum procedures. *arXiv preprint arXiv:2106.12844*, 2021.

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.

Rick Durrett. *Probability: theory and examples.* Cambridge university press, 2010.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407 – 499, 2004. doi: 10.1214/009053604000000067. URL `https://doi.org/10.1214/009053604000000067`.

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

Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014.

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.

Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.

Abhishek Kaul and George Michailidis. Inference for change points in high dimensional mean shift models. *arXiv preprint arXiv:2107.09150*, 2021.

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

Abhishek Kaul, Siddhartha Mandal, Ori Davidov, and Shyamal D Peddada. Analysis of microbiome data in the presence of excess zeros. *Frontiers in microbiology*, 8:2114, 2017b.

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.

Abhishek Kaul, Stergios B Fotopoulos, Venkata K Jandhyala, Abolfazl Safikhani, et al. Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics*, 15(1):71–134, 2021.

Hossein Keshavarz and George Michailidis. Online detection of local abrupt changes in high-dimensional gaussian graphical models. *arXiv preprint arXiv:2003.06961*, 2020.

Hossein Keshavarz, George Michailidis, and Yves Atchadé. Sequential change-point detection in high-dimensional gaussian graphical models. *Journal of machine learning research*, 21(82):1–57, 2020.

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.

Yan Lan, Moulinath Banerjee, and George Michailidis. Change-point estimation under adaptive sampling. *The Annals of Statistics*, 37(4):1752–1791, 2009.

Avery A Lane, Michelle K McGuire, Mark A McGuire, Janet E Williams, Kimberly A Lackey, Edward H Hagen, Abhishek Kaul, Debela Gindola, Dubale Gebeyehu, Katherine E Flores, et al. Household composition and the infant fecal microbiome: The inspire study. *American journal of physical anthropology*, 169(3):526–539, 2019.

Sokbae Lee, Myung Hwan Seo, and Youngki Shin. The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):193–210, 2016.

Yu-Ning Li, Degui Li, and Piotr Fryzlewicz. Detection of multiple structural breaks in large covariance matrices. *arXiv preprint*, 2019.

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`.

Catherine A Lozupone, Jesse Stombaugh, Antonio Gonzalez, Gail Ackermann, Doug Wendel, Yoshiki Vázquez-Baeza, Janet K Jansson, Jeffrey I Gordon, and Rob Knight. Meta-analyses of studies of the human microbiota. *Genome research*, 23(10):1704–1714, 2013.

Euan Thomas McGonigle and Hankui Peng. Subspace change-point detection via low-rank matrix factorisation. *arXiv preprint arXiv:2110.04044*, 2021.

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.

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

Alessandro Rinaldo, Daren Wang, Qin Wen, Rebecca Willett, and Yi Yu. Localizing changes in high-dimensional regression models. *arXiv preprint arXiv:2010.10410*, 2020.

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.

Emilio Seijo and Bodhisattva Sen. A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics*, 5:421–439, 2011.

Sapna Sharma and Prabhanshu Tripathi. Gut microbiome and type 2 diabetes: where we are and where to go? *The Journal of nutritional biochemistry*, 63:101–108, 2019.

Christine Sinoquet. *Probabilistic graphical models for genetics, genomics, and postgenomics.* OUP Oxford, 2014.

Elizabeth Svoboda. Could the gut microbiome be linked to autism? *Nature*, 577(7792): S14–S15, 2020.

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

Sara Van de Geer, Peter Bühlmann, Ya'acov 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, Kevin Lin, and Rebecca Willett. Statistically and computationally efficient change point localization in regression settings. *arXiv preprint arXiv:1906.11364*, 2019a.

Daren Wang, Yi Yu, Alessandro Rinaldo, et al. Optimal covariance change point localization in high dimensions. *Bernoulli*, 27(1):554–575, 2021.

Runmin Wang, Stanislav Volgushev, and Xiaofeng Shao. Inference for change points in high dimensional data. *arXiv preprint arXiv:1905.08446*, 2019b.

Tengyao Wang and Richard J. Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83, 2018.

W Whitt. Stochastic-process limits. springer series in operations research. 2002. URL http://www.columbia.edu/~ww2040/fcltchno.pdf.

Tanya Yatsunenko, Federico E Rey, Mark J Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N Baldassano, Andrey P Anokhin, et al. Human gut microbiome viewed across age and geography. *nature*, 486 (7402):222–227, 2012.

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

Appendix

Appendix A. Proofs of results in Section 2

The following notations is used throughout this Section. In addition to $\xi_{2,2}$ defined in (1.3), 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 $\hat{\eta}_{(j)} = \hat{\mu}_{(j)} - \hat{\gamma}_{(j)}$, $j = 1, ..., p$. We also recall the definition of r.v.'s ε_{tj} from (2.3),

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

Towards obtaining the rate of estimation of the proposed estimator we require a lemma that is instrumental to our argument. Define for $\mu, \gamma \in \mathbb{R}^{p(p-1)}$ and $\tau \in \{1, ..., (T-1)\}$,

$$\mathcal{U}(\tau, \mu, \gamma) = \Big(Q(\tau, \mu, \gamma) - Q(\tau^0, \mu, \gamma)\Big), \tag{A.1}$$

where $\tau^0 \in \{1, ..., (T-1)\}$ is the change point parameter and $Q(\tau, \mu, \gamma)$ is the squared loss defined earlier. For any non-negative sequences $0 \le v_T \le u_T \le 1$, define the collection

$$\mathcal{G}(u_T, v_T) = \Big\{\tau \in \{1, ..., (T-1)\};\ Tv_T \le \big|\tau - \tau^0\big| \le Tu_T\Big\} \tag{A.2}$$

Then, the following lemma provides a uniform lower bound on the expression $\mathcal{U}(\tau, \hat{\mu}, \hat{\gamma})$, over the collection $\mathcal{G}(u_T, v_T)$.

Lemma 15 *Suppose Conditions A, B and C hold and let $0 \le v_T \le u_T \le 1$ be any non-negative sequences. For any $0 < a < 1$, let $c_{a1} = 4{\cdot}48c_{a2}$, with $c_{a2} \ge \surd(1/a)$, and*

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

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

$$\inf_{\tau \in \mathcal{G}(u_T, v_T)} \mathcal{U}(\tau, \hat{\mu}, \hat{\gamma}) \ge \kappa\xi_{2,2}^2\Big[v_T - c_{a3}\max\Big\{\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}},\ \frac{u_T}{T^b}\Big\}\Big] \tag{A.3}$$

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

Proof of Lemma 15 For any fixed $\tau \ge \tau^0$ consider,

$$\begin{aligned}
&\mathcal{U}(\tau, \hat{\mu}, \hat{\gamma} = Q(\tau, \hat{\mu}, \hat{\gamma}) - Q(\tau^0, \hat{\mu}, \hat{\gamma}) \qquad \text{(A.4)}\\
&= \frac{1}{T}\sum_{t=1}^{\tau}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat{\mu}_{(j)})^2 + \frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat{\gamma}_{(j)})^2\\
&-\frac{1}{T}\sum_{t=1}^{\tau^0}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat{\mu}_{(j)})^2 - \frac{1}{T}\sum_{t=\tau^0+1}^{T}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat{\gamma}_{(j)})^2\\
&= \frac{1}{T}\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat{\mu}_{(j)})^2 - \frac{1}{T}\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}(z_{tj} - z_{t,-j}^T\hat{\gamma}_{(j)})^2\\
&= \frac{1}{T}\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}\big(z_{t,-j}^T\hat{\eta}_{(j)}\big)^2 - \frac{2}{T}\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\hat{\eta}_{(j)}\\
&+\frac{2}{T}\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}(\hat{\gamma}_{(j)} - \gamma_{(j)}^0)^T z_{t,-j}z_{t,-j}^T\hat{\eta}_{(j)}.
\end{aligned}$$

The expansion in (A.4) provides the following relation,

$$\begin{aligned}
\inf_{\substack{\tau \in \mathcal{G}(u_T, v_T);\\ \tau \ge \tau^0}} \mathcal{U}(\tau, \hat{\mu}, \hat{\gamma} \ge &\inf_{\substack{\tau \in \mathcal{G}(u_T, v_T);\\ \tau \ge \tau^0}} \frac{1}{T}\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}\big(z_{t,-j}^T\hat{\eta}_{(j)}\big)^2\\
&-2\sup_{\substack{\tau \in \mathcal{G}(u_T, v_T);\\ \tau \ge \tau^0}} \frac{1}{T}\Big|\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\hat{\eta}_{(j)}\Big|
\end{aligned}$$

$$
\begin{aligned}
&-2\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}}\frac{1}{T}\Big|\sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p}(\hat{\gamma}_{(j)}-\gamma^0_{(j)})^T z_{t,-j}z_{t,-j}^T\hat{\eta}_{(j)}\Big|\\
&= R1-R2-R3 \qquad\qquad (A.5)
\end{aligned}
$$

Bounds for the terms $R1, R2$ and $R3$ are provided in Lemmas 24 and 25, respectively. 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 $1-a-o(1)$. The first inequality follows from Lemma 24 and the final inequality follows by using the bounds in Lemma 25. Next, we obtain upper bounds for the terms $R2/\kappa\xi_{2,2}^2$ and $R3/\kappa\xi_{2,2}^2$. For this purpose, first note that $(\xi_{2,1}/\xi_{2,2})\leq\surd p$, consequently $(\xi_{2,1}/\xi_{2,2}^2)\leq 1/\psi$. Next, consider

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

with probability at least $1-a-o(1)$. As before, the first inequality follows from Lemma 24 and the final inequality follows by using the bounds of Lemma 25. Similarly we can also obtain,

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

with probability at least $1-a-o(1)$. Substituting these bounds in (A.5) and applying a union bound over these events yields the bound (A.3) uniformly over the set $\{\mathcal{G}(u_T,v_T);\,\tau\geq\tau^0\}$. The mirroring case of $\tau\leq\tau^0$ follows with similar arguments. ■

The main idea of the proof of Theorem 1 is to use a contradiction argument. Using Lemma 15 recursively, we show that any value of τ lying outside an $O(c_{a3}^2)$ neighborhood of τ^0 satisfies, $\mathcal{U}(\tau,\hat{\mu},\hat{\gamma})>0$, with probability at least $1-3a-o(1)$. Upon noting that by definition, we have, $\mathcal{U}(\tilde{\tau},\hat{\mu},\hat{\gamma})\leq 0$, shall yield the desired result.

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

$$
\big|\tilde{\tau}-\tau^0\big|\leq c_{a3}^2, \qquad (A.6)
$$

holds with probability at least $1 - 3a - o(1)$. The proof to follow relies on a recursive argument on Lemma 15, 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 with any $v_T > 0$, and applying Lemma 15 on the set $\mathcal{G}(1, v_T)$ to obtain,

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

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 $\inf_{\tau \in \mathcal{G}(1,v_T)} \mathcal{U}(\tau, \hat{\mu}, \hat{\gamma}) > 0$, thus implying that $\tilde{\tau} \notin \mathcal{G}(1, v_T)$, i.e., $\big|\tilde{\tau} - \tau^0\big| \leq T v_T^*$, with probability at least $1 - 3a - o(1)$[8]. Now reset $u_T = v_T^*$ and reapply Lemma 15 for any $v_T > 0$ to obtain,

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

Again choosing any

$$v_T > v_T^* = \max\Big\{ \frac{c_{a3}^{g_2}}{T^{u_2}}, \ \frac{c_{a3}^2}{T^{v_2}} \Big\}, \tag{A.7}$$

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 $\inf_{\mathcal{G}(u_T,v_T)} \mathcal{U}(\tau, \hat{\mu}, \hat{\gamma}) > 0$, with probability at least $1 - 3a - o(1)$. Consequently $\tilde{\tau} \notin \mathcal{G}(u_T, v_T)$, i.e., $\big|\tilde{\tau} - \tau^0\big| \leq T v_T^*$. Note the 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 15, we obtain for the m^{th} recursion,

$$\begin{aligned}
&\big|\tilde{\tau} - \tau^0\big| \leq T \max\Big\{ \frac{c_{a3}^{g_m}}{T^{u_m}}, \ \frac{c_{a3}^m}{T^{v_m}} \Big\} := T \max\{R_{1m}, R_{2m}\}, \quad \text{where,} \\
&g_m = \sum_{k=0}^{m-1} \frac{1}{2^k}, \ u_m = \frac{1}{2} + \frac{u_{m-1}}{2} = \frac{b}{m} + \sum_{k=1}^{m} \frac{1}{2^m}, \ \text{and} \\
&v_m = b + v_{m-1} \geq mb, \ \text{with} \ u_1 = v_1 = b.
\end{aligned}$$

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 $1/T$.[9] Consequently for m large enough we have $\big|\tilde{\tau} - \tau^0\big| \leq T R_{1m}$, 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,

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

8. Since by construction of $\tilde{\tau}$ we have, $\mathcal{U}(\tilde{\tau}, \hat{\gamma}, \hat{\gamma}) \leq 0$.
9. 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.

with probability at least $1 - 3a - o(1)$. This proves the bound (A.6). 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 15 arises from stochastic upper bounds of Lemma 20, Lemma 21, Lemma 22 and Lemma 36, 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. ■

The limiting distributions of Theorem 2 and Theorem 5 are presented in more conventional *argmax* notation instead of the *argmin* of the problem setup in Section 1. This is purely notational and all results can equivalently be stated in the *argmin* language.

For a clear presentation of the proofs below we use the following additional notation. Let $\mathcal{U}(\tau, \theta_1, \theta_2)$ be as in (A.1) and consider,

$$\mathcal{C}(\tau, \mu, \gamma) \;\; = \;\; -Tp^{-1}\mathcal{U}(\tau, \mu, \gamma) \tag{A.8}$$

The multiplication of $\mathcal{U}$ with the product Tp^{-1} is only meant for notational convenience later on. Then, we can re-express the change point estimator $\tilde{\tau}(\mu, \gamma)$ defined in (1.5) as,

$$\tilde{\tau}(\mu, \gamma) = \arg\max_{\tau \in \{1,\dots,(T-1)\}} \mathcal{C}(\tau, \mu, \gamma)$$

The proofs of Theorem 2 and Theorem 5 below are applications of the Argmax Theorem (reproduced as Theorem 47). The arguments here are largely an exercise in verification of requirements of this theorem.

Proof of Theorem 2 In the vanishing jump regime $\psi \to 0$, the argmax theorem requires verification of the following conditions (see, page 288 of Vaart and Wellner (1996)).

- (i) The sequence $\psi^2(\tilde{\tau} - \tau^0)$ is uniformly tight.
- (ii) For any $r \in [-c_u, c_u] \subseteq \mathbb{R}$ we have, $\mathcal{C}(\tau^0 + r\psi^{-2}, \hat{\mu}, \hat{\gamma}) \Rightarrow Z(r)$.
- (iii) The process $Z(r)$ satisfies suitable regularity conditions.[10].

We begin by noting that the sequence of r.v.'s under consideration here is $\psi^2(\tilde{\tau} - \tau^0)$, which are supported on $\mathbb{R}$, which forms the underlying indexing metric space for the limiting process under consideration for this vanishing jump size case. Now Part (i) follows from the result of Theorem 1 and Part (iii) follows from well known properties of Brownian motion's. Thus, it only remains to prove Part (ii). For this purpose, let $\tau^* = \tau^0 + r\psi^{-2}$, with $r \in (0, c_1]$, then using Lemma 28 we have,

$$p^{-1} \sum_{\tau^0+1}^{\tau^*} \sum_{j=1}^{p} \eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T \eta_{(j)}^0 \to_p r\sigma_2^2. \tag{A.9}$$

Also, let $\zeta_t = \sum_{j=1}^p \zeta_{tj} = \sum_{j=1}^p \varepsilon_{tj} z_{t,-j}^T \eta_{(j)}^0$, then from Condition D(ii) we have that $\zeta_t^* = \xi_{2,2}^{-2} p^{-1} \mathrm{var}(\zeta_t) \to \sigma_2^{*2}$, thus the sequence $\{\zeta_t^*\}$ are finite variance i.i.d. random variables[11],

10. Almost all sample paths $\zeta \to \{2\sigma_\infty W(\zeta) - |\zeta|\}$ are upper semicontinuous and posses a unique maximum at a (random) point $\arg\max_{\zeta \in \mathbb{R}}\{2\sigma_\infty W(\zeta) - |\zeta|\}$, which as a random map in the indexing metric space is tight.
11. More precisely, sequence $\{\zeta_t^*\}$ forms an i.i.d triangular array

now applying the function central limit theorem (see, e.g., Theorem 4.3.2 of Whitt (2002))) on the sequence $\{\zeta_t^*\}$ in t, we obtain,

$$\begin{aligned} p^{-1}\sum_{t=\tau^0+1}^{\tau^*}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}\eta^0_{(j)} &= \psi\sum_{t=\tau^0+1}^{\tau^*}\psi^{-1}p^{-1}\sum_{j=1}^{p}\zeta_{tj} = \psi\sum_{t=\tau^0+1}^{\tau^*}\big\{\xi_{2,2}^{-1}p^{-1/2}\sum_{j=1}^{p}\zeta_{tj}\big\} \\ &= \psi\sum_{t=\tau^0+1}^{\tau^*}\zeta_t^* \Rightarrow \sigma_2^* W_2(r), \end{aligned} \tag{A.10}$$

where $W_2(r)$ is a Brownian motion on $[0,\infty)$. Next, define the process

$$G(r) = \begin{cases} 2\sigma_1^* W_1(r) + \sigma_1^2 r & \text{if } r < 0, \\ 0, & \text{if } r = 0, \\ 2\sigma_2^* W_2(r) - \sigma_2^2 r & \text{if } r > 0, \end{cases} \tag{A.11}$$

and consider the function $\mathcal{C}$ evaluated at τ^* and at the known nuisance parameters

$$\begin{aligned} \mathcal{C}(\tau^*,\mu^0,\gamma^0) &= -p^{-1}\sum_{t=\tau^0+1}^{\tau^*}\sum_{j=1}^{p}\big(z_{tj}-z_{t,-j}^T\mu^0_{(j)}\big)^2 + p^{-1}\sum_{t=\tau^0+1}^{\tau^*}\sum_{j=1}^{p}\big(z_{tj}-z_{t,-j}^T\gamma^0_{(j)}\big)^2 \\ &= 2p^{-1}\sum_{t=\tau^0+1}^{\tau^*}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\eta^0_{(j)} - p^{-1}\sum_{t=\tau^0+1}^{\tau^*}\sum_{j=1}^{p}\eta^{0T}_{(j)}z_{t,-j}z_{t,-j}^T\eta^0_{(j)} \\ &\Rightarrow \big\{2\sigma_2^* W_2(r) - \sigma_2^2 r\big\}, \end{aligned} \tag{A.12}$$

where convergence in distribution follows from (A.9) and (A.10). Next, from Lemma 26 we have that,

$$\sup_{\tau\in\mathcal{G}\big((c_1T^{-1}\psi^{-2}),0\big)}\big|\mathcal{C}(\tau,\hat\mu,\hat\gamma)-\mathcal{C}(\tau,\mu^0,\gamma^0)\big| = o_p(1). \tag{A.13}$$

Combining the results of (A.13) and (A.12) we obtain,

$$\mathcal{C}(\tau^*,\hat\mu,\hat\gamma) \Rightarrow \big\{2\sigma_2^* W_2(r) - \sigma_2^2 r\big\}$$

Symmetrical arguments for the case of $r<0$ yields an analogous result. Finally, a change of variable yields the relation, $\arg\min_r G(r) =^d \big(\sigma_1^{*2}/\sigma_1^4\big)\arg\min_r Z(r)$, where $Z(r)$ is as defined in (2.2) and $=^d$ represents equality in distribution, see, e.g. proof of Proposition 3 of Bai (1997). This completes the proof of Part (ii) and the statement of this theorem now follows as an application of the argmax theorem. ■

Proof of Theorem 5 The broad structure of the argument of this proof is similar to that of the proof of Theorem 2 in the sense that it is also an application of the argmax theorem.

The first important distinction is that the sequence of r.v's under consideration $\big(\tilde\tau-\tau^0\big)$, are supported on the set of integers $\mathbb{Z}$. Consequently, the underlying indexing metric space for the limiting process for this non-vanishing jump size framework is the set of integers $\mathbb{Z}$.

Now consider any $c_u > 0$ and $r \in \{-c_u, -c_u+1, ..., 0, 1, ..., c_u\} \subseteq \mathbb{Z}$. Let $\tau^* = \tau^0 + r$, then the requirements of the argmax theorem requires verification of the following conditions.

(i) The sequence $\big(\tilde{\tau} - \tau^0\big)$ is uniformly tight.

(ii) $\mathcal{C}(\tau^*, \hat{\mu}, \hat{\gamma}) \Rightarrow \mathcal{C}_\infty(r)$.

(iii) The process $\mathcal{C}_\infty(r)$ satisfies suitable regularity conditions.

Part (i) follows directly from the result of Theorem 1. Part (iii) is provided in Lemma 16. A verification of Part (ii) is provided below. Let $r > 0$ and consider $\mathcal{C}$ evaluated at τ^* and at the known nuisance parameters, i.e.,

$$\begin{aligned}
\mathcal{C}(\tau^*, \mu^0, \gamma^0) &= -p^{-1}\sum_{t=\tau^0+1}^{\tau^0+r}\sum_{j=1}^{p}\big(z_{tj} - z_{t,-j}^T\mu_{(j)}^0\big)^2 + p^{-1}\sum_{t=\tau^0+1}^{\tau^0+r}\sum_{j=1}^{p}\big(z_{tj} - z_{t,-j}^T\gamma_{(j)}^0\big)^2 \\
&= 2p^{-1}\sum_{t=\tau^0+1}^{\tau^0+r}\sum_{j=1}^{p}\varepsilon_{tj} z_{t,-j}^T\eta_{(j)}^0 - p^{-1}\sum_{t=\tau^0+1}^{\tau^0+r}\sum_{j=1}^{p}\eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T\eta_{(j)}^0 \\
&= p^{-1}\sum_{t=\tau^0+1}^{\tau^0+r}\sum_{j=1}^{p}\Big\{2\varepsilon_{tj} z_{t,-j}^T\eta_{(j)}^0 - \eta_{(j)}^{0T} z_{t,-j} z_{t,-j}^T\eta_{(j)}^0\Big\} \\
&\Rightarrow \sum_{t=1}^{r}\mathcal{L}\big(-\psi_\infty^2\sigma_2^2,\ \bar{\sigma}_2^2\big) \qquad \text{(A.14)}
\end{aligned}$$

The weak convergence follows directly from Condition B′ and since $r \le c_u$, which in turn is due to the non-vanishing jump size regime under consideration. Next, from Lemma 26,

$$\sup_{\tau\in\mathcal{G}\left((c_1T^{-1}\psi^{-2}),0\right)}\big|\mathcal{C}(\tau, \hat{\mu}, \hat{\gamma}) - \mathcal{C}(\tau, \mu^0, \gamma^0)\big| = o_p(1).$$

This result together with (A.14) yields the statement of Part (ii). Repeating the same argument for $r < 0$ yields the symmetric result. An application of the argmax theorem now yields the statement of this theorem. ■

Lemma 16 (Regularity conditions of $\arg\max \mathcal{C}_\infty(r)$**)** *Let $\mathcal{C}_\infty(r)$ be as defined in (2.6) and suppose Condition B′ holds. Then, the map $r \to \mathcal{C}_\infty(r)$ is continuous with respect to the domain space $\mathbb{Z}$. Additionally assume that Condition D holds and that the jump size is non-vanishing, i.e, $0 < \psi_\infty < \infty$. Then $\arg\max_{r\in\mathbb{Z}}\mathcal{C}_\infty(r)$ possesses an almost sure unique maximum at ω_∞, which as a random map in $\mathbb{Z}$ is tight.*

Proof of Lemma 16 From Condition B′, each side of the random walk $\mathcal{C}(r)$ has increments supported on $\mathbb{R}$, thus the first assertion on the continuity of the map $r \to \mathcal{C}_\infty(r)$ follows trivially since the domain space of this map is restricted to only the integers $\mathbb{Z}$ ($\epsilon - \delta$ definition of continuity). To prove the remaining assertions note that from Condition B′, Condition D and the assumed framework of the non-vanishing jump size, we have that each side of $\mathcal{C}_\infty(r)$ has i.i.d increments with a negative drift of $-\psi_\infty^2\sigma_1^2$ or $-\psi_\infty^2\sigma_2^2$. Consequently, we have $\mathcal{C}_\infty(r) \to -\infty$, as $r \to \infty$ almost surely (strong law of large numbers). Using elementary properties of random walks, this implies that $\max_r \mathcal{C}_\infty(r) < \infty$, a.s. (follows from the Hewitt-Savage 0-1 law, see, e.g. (1.1) and (1.2) on Page 172, 173 of Durrett (2010)). Additionally $\omega_\infty \geq 0$, from the construction of $\mathcal{C}_\infty(r)$. Thus, we have $0 \leq \omega_\infty < \infty$, a.s. which directly implies that when ω_∞ is well defined (unique) then it must be tight. To show that ω_∞ is unique, note that since by assumption (Condition B$'$) the increments are continuously distributed and supported on $\mathbb{R}$, therefore $\max \mathcal{C}_\infty(r)$ is continuously distributed on $(0,\infty)$, with some additional probability mass at the singleton zero. Hence, the probability of $\max \mathcal{C}_\infty(r)$ attaining any two identical values is zero. Consequently ω_∞ is unique a.s. This completes the proof of this lemma. ∎

Proof of Theorem 7 The main idea of this proof is first to prove the weak convergence of the two underlying stochastic processes, i.e., the two sided random walk and the Brownian motion on the lhs and rhs of (2.9), repsectively, followed by an application of continuous mapping type results to obtain the weak convergence of the desired *argmax*.

The first immediate roadblock towards this approach the incoherence of the indexing spaces of the stochastic processes on the lhs and rhs of (2.9). To alleviate this incoherence one may consider representing the lhs of (2.9) as $\arg\max_{r\in\mathbb{R}}\mathcal{C}_\infty(\lfloor r\rfloor)$. This representation is however not well defined due to the non-uniqueness of the *argmax* functional in this case. Thus, *argmax* needs to re-defined as the smallest maximizer: $\text{sargmax} f(x) = \min\{x;\ f(x) \geq f(y)\ \forall\ y\}$. The functional sargmax has been studied in the literature, e.g., Lan et al. (2009) and Seijo and Sen (2011) whose motivations are exactly the same that arise here. Under this definition one can re-write the lhs of (2.9) as,

$$\begin{aligned} \arg\max_{r\in\mathbb{Z}}\mathcal{C}_\infty\big(r,\psi_T,\sigma_T^2,\bar\sigma_T\big) &=^d \text{sargmax}_{r'\in\mathbb{R}}\mathcal{C}_\infty\big(\lfloor r'\rfloor,\psi_T,\sigma_T^2,\bar\sigma_T^2\big)\\ &=^d \lfloor\psi_T^{-2}\rfloor \text{sargmax}_{r\in\mathbb{R}}\mathcal{C}_\infty\big(\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor,\psi_T,\sigma_T^2,\bar\sigma_T^2\big) \qquad \text{(A.15)}\end{aligned}$$

where the second equality follows directly from a change of variables $r' = r\lfloor\psi^{-2}\rfloor$. Next consider the random walk in the rhs of (A.15) as per the defining relation (2.6).

$$\begin{aligned}\mathcal{C}_\infty\big(\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor,\psi_T,\sigma_T^2,\bar\sigma_T^2\big) &= \begin{cases}\sum_{t=1}^{-\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor}\mathcal{L}\big(-\psi_T^2\sigma_{1T}^2,\ \bar\sigma_{1T}^2\big), & r\in\mathbb{R}^-\\ 0, & r=0\\ \sum_{t=1}^{\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor}\mathcal{L}\big(-\psi_T^2\sigma_{2T}^2,\ \bar\sigma_{2T}^2\big), & r\in\mathbb{R}^+.\end{cases}\\ &=^d \begin{cases}\sum_{t=1}^{-\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor}\Big[\mathcal{L}\big(0,\ \bar\sigma_{1T}^2\big)-\psi_T^2\sigma_{1T}^2\Big], & r\in\mathbb{R}^-\\ 0, & r=0\\ \sum_{t=1}^{\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor}\Big[\mathcal{L}\big(0,\ \bar\sigma_{2T}^2\big)-\psi_T^2\sigma_{2T}^2\Big], & r\in\mathbb{R}^+.\end{cases} \qquad \text{(A.16)}\end{aligned}$$

here the second equality follows from the additive invariance of $\mathcal{L}$ w.r.t scalar addition. Now consider the positive arm ($r>0$) of this process, we have,

$$\sum_{t=1}^{\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor}\Big[\mathcal{L}\big(0,\ \bar\sigma_{2T}^2\big)-\psi_T^2\sigma_{2T}^2\Big] =^d \lfloor\psi_T\rfloor\sum_{t=1}^{\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor}\mathcal{L}\big(0,\ \lfloor\psi_T^{-2}\rfloor\bar\sigma_{2T}^2\big)-\psi_T^2\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor\sigma_{2T}^2$$

$$\Rightarrow \quad 2\sigma_2^* W_2(r) - \sigma_2^2 r \tag{A.17}$$

The equality follows from the invariance of $\mathcal{L}$ w.r.t scalar multiplication. The weak convergence follows from the functional central limit theorem, together with the limit assumptions on the underlying sequences, specifically, $\psi_T \to 0$, $\sigma_{2T}^2 \to \sigma_2^2$, and $\psi_T^{-2}\bar{\sigma}_{2T}^2 \to 4\sigma_2^{*2}$. Here we have also utilized the elementary result $\psi_T^2\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor \to r$. The relation (A.17) together with a symmetric result on the negative arm ($r < 0$) of this process yields,

$$\mathcal{C}_\infty\big(\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor, \psi_T, \sigma_T^2, \bar{\sigma}_T^2\big) \Rightarrow G(r, \sigma^2, \sigma^{*2}) := \begin{cases} 2\sigma_1^* W_1(r) + \sigma_1^2 r & \text{if } r < 0, \\ 0, & \text{if } r = 0, \\ 2\sigma_2^* W_2(r) - \sigma_2^2 r & \text{if } r > 0, \end{cases} \tag{A.18}$$

Applying the continuous mapping theorem for the *sargmax* functional (Lemma 3.1 of Lan et al. (2009) or Theorem 3.1 of Seijo and Sen (2011)) we obtain,

$$\begin{aligned} \text{sargmax}_{r\in\mathbb{R}}\mathcal{C}_\infty\big(\lfloor r\lfloor\psi_T^{-2}\rfloor\rfloor, \psi_T, \sigma_T^2, \bar{\sigma}_T^2\big) &\Rightarrow \arg\max_{r\in\mathbb{R}} G(r, \sigma^2, \sigma^{*2}) \\ &= \big(\sigma_1^{*2}/\sigma_1^4\big)\arg\max_{r\in\mathbb{R}} Z(r, \sigma^2, \sigma^{*2}), \end{aligned}$$

The last equality follows from a change of variables (also see, proof of Theorem 2). Also note that *sargmax* of the rhs has been replaced by *argmax*, since the rhs possesses a unique maximizer. Finally, the statement of the theorem now follows by a back substitution to the relation (A.15) and noting that $\psi_T^2\lfloor\psi_T^{-2}\rfloor \to 1$. This completes the proof of this theorem. ■

Appendix B. Proofs of results in Section 3

The main result of Section 3 is Theorem 10, which forms the basis of the subsequent corollaries. This result 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.

Proof of Theorem 10 Consider any $\tau \in \mathcal{G}(u_T, 0)$, and w.l.o.g. assume that $\tau \ge \tau^0$. Then for any $j = 1, .., p$, by construction of the estimator $\hat{\mu}_{(j)}(\tau)$, we have the basic inequality,

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

An algebraic rearrangement of this inequality yields,

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

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

$$\frac{1}{\tau}\sum_{t=1}^{\tau}\big(z_{t,-j}^T(\hat{\mu}_{(j)} - \mu_{(j)}^0)\big)^2 + \lambda_j\|\hat{\mu}_{(j)}(\tau)\|_1 \quad \le \quad \lambda_j\|\mu_{(j)}^0\|_1 + \frac{2}{\tau}\sum_{t=1}^{\tau}\varepsilon_{tj}z_{t,-j}^T(\hat{\mu}_{(j)} - \mu_{(j)}^0) \tag{B.1}$$

$$
\begin{aligned}
& -\frac{2}{\tau}\sum_{t=\tau^0+1}^{\tau}(\mu^0_{(j)}-\gamma^0_{(j)})z_{t,-j}z^T_{t,-j}(\hat\mu_{(j)}-\mu^0_{(j)})\\
\le\;& \lambda\|\mu^0_{(j)}\|_1+\frac{2}{\tau}\Big\|\sum_{t=1}^{\tau}\varepsilon_{tj}z^T_{t,-j}\Big\|_\infty\|\hat\mu_{(j)}-\mu^0_{(j)}\|_1\\
& +\frac{2}{\tau}\Big\|\sum_{t=\tau^0+1}^{\tau}(\mu^0_{(j)}-\gamma^0_{(j)})z_{t,-j}z^T_{t,-j}\Big\|_\infty\|\hat\mu_{(j)}-\mu^0_{(j)}\|_1
\end{aligned}
$$

Using the bounds in Lemma 29 we obtain

$$
\begin{aligned}
\frac{1}{\tau}\Big\|\sum_{t=1}^{\tau}\varepsilon_{tj}z_{t,-j}\Big\|_\infty &\le c_u\sigma^2\sqrt{(1+\nu^2)}\Big\{\frac{\log(p\vee T)}{T\ell_T}\Big\}^{\frac12}=\lambda_{1j}\\
\frac{1}{\tau}\Big\|\sum_{t=\tau^0+1}^{\tau}\eta^{0T}_{(j)}z_{t,-j}z^T_{t,-j}\Big\|_\infty &\le c_u(\sigma^2\vee\phi)\|\eta^0_{(j)}\|_2\max\Big\{\frac{\log(p\vee T)}{T\ell_T},\ \frac{u_T}{\ell_T}\Big\}=\lambda_{2j},
\end{aligned}
$$

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

$$
\frac{1}{\tau}\sum_{t=1}^{\tau}\big(z^T_{t,-j}(\hat\mu_{(j)}-\mu^0_{(j)})\big)^2+\lambda_j\|\hat\mu_{(j)}(\tau)\|_1\le\lambda_j\|\mu^0_{(j)}\|_1+(\lambda_{1j}+\lambda_{2j})\|\hat\mu_{(j)}(\tau)-\mu^0_{(j)}\|_1,
$$

with probability at least $1-c_{u2}\exp\{-c_{u3}\log(p\vee T)\}$. Choosing $\lambda_j\ge 2(\lambda_{1j}+\lambda_{2j})$, leads to $\big\|\big(\hat\mu_{(j)}(\tau)\big)_{S^c_{1j}}\big\|_1\le 3\big\|\big(\hat\mu_{(j)}(\tau)-\mu^0_j\big)_{S_{1j}}\big\|_1$, and thus by definition $\hat\mu_{(j)}-\mu^0_{(j)}\in\mathcal{A}_{1j}$, with the same probability. This proves the first assertion of this theorem. Next, applying the restricted eigenvalue condition of (37) to the l.h.s. of the inequality (B.1), we also have that,

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

This directly implies that $\|\hat\mu_{(j)}(\tau)-\mu^0_{(j)}\|_2\le 3\sqrt{s}(\lambda_j/\kappa)$, which yields the desired ℓ_2 bound. To complete 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 $\tau\le\tau^0$, and the corresponding results for $\hat\gamma_{(j)}(\tau)-\gamma^0_{(j)}$ can be obtained by symmetrical arguments. ■

Proof of Corollary 11 All we need to show here is that the mean estimates $\hat\mu(\hat\tau)_{(j)},\hat\gamma(\hat\tau)_j$, $j=1,...,p$ of Step 2 of Algorithm 1 satisfy all requirements of Condition C. Then the statement of the Corollary follows from direct applications of the corresponding results of Section 2.

Towards this, Part (i) of Condition C now holds directly as a consequence of Theorem 10. Additionally from Condition (3.2) we have $\hat\tau\in\mathcal{G}(u_T,0)$, with probability at least $1-o(1)$, where $u_T=c_uT^{-1}\psi^{-2}\log(p\vee T)$. Substitute this choice of u_T in λ_{2j}, $j=1,...p$, of Theorem 10 to obtain,

$$
\lambda_{2j}\;=\;c_u(\sigma^2\vee\phi)\|\eta^0_{(j)}\|_2\max\Big\{\frac{\log(p\vee T)}{T\ell_T},\ \frac{1}{\psi^2}\frac{\log(p\vee T)}{T\ell_T}\Big\}
$$

$$\begin{aligned}
&\le \; O(1)\max\Big\{\frac{\log(p\vee T)}{T\ell_T},\; \Big(\frac{\max_j\|\eta^0_{(j)}\|_2}{\psi\sqrt{s}}\Big)\Big(\frac{1}{\psi}\Big)\Big(\frac{s\log(p\vee T)}{T\ell_T}\Big)^{1/2}\Big(\frac{\log(p\vee T)}{T\ell_T}\Big)^{1/2}\Big\}\\
&\le \; o(1)\Big\{\frac{\log(p\vee T)}{T\ell_T}\Big\}^{\frac{1}{2}}
\end{aligned}$$

Here the final inequality follows since by Condition A$'$ together with the condition $\Big(\max_j\|\eta^0_{(j)}\|_2/(\psi\sqrt{s})\Big) = O(1)$. Consequently $\lambda_{2j}\le\lambda_{1j}$, $j=1,...,p$, and thus applying Theorem 10 we obtain,

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

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. Thereby the mean estimates of Step 2 of Algorithm 1 satisfy Condition C, which completes the proof of this corollary. ■

The following lemma obtains ℓ_2 error bounds for the Step 1 edge estimates by utilizing the initializing Condition E and Theorem 10.

Lemma 17 *Suppose Conditions A$'$, B and E hold. Select regularizers λ_j, $j=1,...,p$, as prescribed in Theorem 10, with $u_T=\big(c_ul_T\kappa\big)/\big(sT^k(\sigma^2\vee\phi)\big)$. Then, edge estimates $\check\mu_{(j)}$, $j=1,...,p$ of Step 1 of Algorithm 2 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 17 We begin by noting that Part (ii) of the initializing Condition E of Algorithm 2 guarantees that $\check\tau$ satisfies,

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

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

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

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

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 E. The equality of (B.3) follows directly upon substituting the choice of u_T, and the inequality follows from assumption A$'$ 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 10 we obtain,

$$\sum_{j=1}^p\frac{\sqrt{s}}{\kappa}(\lambda_{1j}+\lambda_{2j})\;\;\le\;\;c_u\sigma^2\sqrt{(1+\nu^2)}\frac{p}{\kappa}\Big\{\frac{s\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}}$$

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

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 10. 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$'$. Part (ii) is now a direct consequence. ■

Lemma 18 *Suppose Condition A$'$, B and E hold and let $\check{\mu}_{(j)}$ and $\check{\gamma}_{(j)}$, $j=1,...,p$ be edge estimates of Step 1 of Algorithm 2. Additionally, let $\log(p\vee T)\le Tv_T\le Tu_T$ be non-negative sequences. Then,*

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

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

Proof of Lemma 18 The structure of this proof is similar to that of Lemma 15, 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.5) we have that,

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

Where $R1,R2$ and R_3 are as defined in (A.5) 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 34 we obtain,

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

$$\geq \quad \kappa\xi_{2,2}^2\Big[v_T - \frac{c_u\sigma^2}{\kappa}\Big\{\frac{u_T\log(p\vee T)}{T}\Big\}^{\frac{1}{2}} - c_u(\sigma^2\vee\phi)\frac{u_T}{T^k\kappa}\Big]$$

with probability $1-o(1)$. Where the final inequality follows from Lemma 17. Next we obtain upper bounds for the terms $R2\big/\kappa\xi_{2,2}^2$ and $R3\big/\kappa\xi_{2,2}^2$. Consider,

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

with probability $1-o(1)$. Here the first and second inequalities follow from Lemma 34 and Lemma 17, respectively. Similarly we can also obtain,

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

with probability $1-o(1)$. Recalling from Lemma 48 that $\psi<\infty$, then substituting these bounds in (B.4) and applying a union bound over these three events yields the bound of the statement of this lemma uniformly over the set $\{\mathcal{G}(u_T,v_T);\,\tau\geq\tau^0\}$. The mirroring case of $\tau\leq\tau^0$ can be obtained by similar arguments. ■

Proof of Theorem 13 This proof relies on the same recursive argument as that of Theorem 1, the distinction being that recursions are made on the bound of Lemma 18 instead of Lemma 15. Consider any $Tv_T>\log(p\vee T)$, and apply Lemma 18 on the set $\mathcal{G}(u_T,v_T)$ to obtain,

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

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

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

with probability at least $1-o(1)$. Recall that w.l.o.g $k<b<(1/2)$, 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., $|\check{\tau} - \tau^0| \le T v_T^*$, with probability at least $1 - o(1)$. Next, reset $u_T = v_T^*$ and reapply Lemma 15 for any $v_T > 0$ to obtain

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

Again choosing any

$$v_T > v_T^* = \max\Big\{ c_m^{g_2} \Big(\frac{\log(p \vee T)}{T}\Big)^{u_2},\ c_m^2 \Big(\frac{\log(p \vee T)}{T}\Big)^{v_2} \Big\}, \tag{B.4}$$

where

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

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., $|\hat{\tau} - \tau^0| \le T v_T^*$. Continuing these recursions by resetting u_T to the bound of the previous recursion, and applying Lemma 15, we obtain for the l^{th} recursion,

$$\begin{aligned} |\tilde{\tau} - \tau^0| &\le T \max\Big\{ c_m^{g_l} \Big(\frac{\log(p \vee T)}{T}\Big)^{u_l},\ c_m^l \Big(\frac{\log(p \vee T)}{T}\Big)^{v_l} \Big\} \\ &:= T \max\{R_{1l}, R_{2l}\},\ \text{ where,} \\ g_l &= \sum_{j=0}^{l-1} \frac{1}{2^j},\ u_l = \frac{1}{2} + \frac{u_{l-1}}{2} = \frac{k}{l} + \sum_{j=1}^{l} \frac{1}{2^j},\ \text{ and} \\ v_l &= k + v_{l-1} \ge lk,\ \text{ with } u_1 = v_1 = k. \end{aligned}$$

Next, it is straightforward to observe that for l large enough, $R_{2l} \le R_{1l}$, for T sufficiently large. Consequently for l large enough we have $|\tilde{\tau} - \tau^0| \le T R_{1m}$, 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,

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

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

Proof of Corollary 14 Note that from Theorem 13 we have that $\hat{\tau}$ of Step 1 of Algorithm 2 satisfies the bound (3.2). Proceeding identically to the proof of Corollary 11 yields the statement of this Corollary. ■

Appendix C. Deviation bounds used for proofs in Section 2

Lemma 19 *Suppose Condition B holds and let ε_{tj} be as in (2.3). Then, (i) the r.v. $\varepsilon_{tj} z_{t,-j,k}$ is sub-exponential with parameter $\lambda_1 = 48\sigma^2 \sqrt{(1+\nu^2)}$, for each $j = 1, ..., p$, $k = 1, ..., p-1$ and $t = 1, ..., T$. (ii) The r.v. $\zeta_t = \sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j}^T \eta_{(j)}^0$ is sub-exponential with parameter $\lambda_2 = 48\sigma^2 \xi_{2,1} \sqrt{(1+\nu^2)}$, 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 19 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 41. We begin by noting that the following r.v's are mean zero, $E(\varepsilon_{tj}) = 0$, $E(z_{t,-j}^T\eta_{(j)}^0) = 0$ and $E\big(\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0\big) = 0$. Also note that for $t \le \tau^0$, we have,

$$\varepsilon_{tj} = z_{tj} - z_{t,-j}^T\mu_{(j)}^0 = \big(1, -\mu_{(j)}^{0T}\big)\big(z_{tj}, z_{t,-j}^T\big)^T$$

Using Lemma 48 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 $t > \tau^0$. Similarly, $z_{t,-j}\eta_{(j)}^0 \sim \text{subG}(\sigma_2)$ with $\sigma_2 = \sigma\|\eta_{(j)}^0\|_2$. Recall that if $Z \sim \text{subG}(\sigma)$, then the rescaled variable $Z/\sigma \sim \text{subG}(1)$. Next observe that,

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

where $\Phi(v) = \|v\|_2^2 - E\big(\|v\|_2^2\big)$. Using Lemma 42 and Lemma 44 we have that $T1 \sim \text{subE}(64)$, $T2 \sim \text{subE}(16)$, and $T3 \sim \text{subE}(16)$. Applying Lemma 43 and rescaling with σ_1, and σ_2 we obtain that $\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 \sim \text{subE}(48\sigma_1\sigma_2)$. Another application of Lemma 43 yields $\zeta_t = \sum_{j=1}^p \varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0 \sim \text{subE}(\lambda_2)$ where

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

This completes the proof of Part (ii). ■

Lemma 20 *Suppose Condition B holds and let ε_{tj} be as defined in (2.3). Additionally, let u_T, v_T be any non-negative sequences satisfying $0 \le v_T \le u_T \le 1$. Then for any $0 < a < 1$, choosing $c_{a1} = 4\cdot 48c_{a2}$, with $c_{a2} \ge \surd(1/a)$, we have for $T \ge 2$,*

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

with probability at least $1-a$.

Proof of Lemma 20 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. Consequently, the sum of interest is over indices t in an empty set, and is thus trivially zero. Now, let $\zeta_t = \sum_{j=1}^p \varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0$, then using Lemma 19 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 19, 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 τ. Applying Kolmogorov's inequality (Theorem 46) with any $d > 0$,

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

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

Lemma 21 *Suppose Condition B holds and let* ε_{tj} *be as defined in (2.3) and let* $0 \leq v_T \leq u_T \leq 1$ *be any non-negative sequences. Then for any* $c_{u2} > 3$ *and* $c_{u1} \geq 96c_{u2}$*, we have for* $T \geq 2$,

$$(i) \qquad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T)\\ \tau\geq\tau^0}} \frac{1}{T}\Big\| \sum_{t=\tau^0+1}^{\tau} \varepsilon_{tj} z_{t,-j}^T \Big\|_\infty \leq c_{u1}\sigma^2\surd(1+\nu^2)\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}} \log(p\vee T),$$

$$(ii) \qquad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T)\\ \tau\geq\tau^0}} \frac{1}{T}\Big| \sum_{t=\tau^0+1}^{\tau}\sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j}^T\big(\hat\eta_{(j)} - \eta^0_{(j)}\big)\Big| \leq$$

$$c_{u1}\sigma^2\surd(1+\nu^2)\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\log(p\vee T)\sum_{j=1}^{p}\|\hat\eta_{(j)} - \eta^0_{(j)}\|_1,$$

each with probability at least $1 - 2\exp\big\{-(c_{u2}-3)\log(p\vee T)\big\}$.[12]

Proof of Lemma 21 Part (ii) is a direct consequence of Part (i), thus we only prove Part (i). Without loss of generality, we can assume $v_T \geq (1/T)$. This follows since the only additional distinct element τ in the set $\mathcal{G}(u_T, 0)$ in comparison to $\mathcal{G}(u_T, (1/T))$ is τ^0, 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 19 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 45) for any fixed $\tau \in \mathcal{G}(u_T, v_T)$ satisfying $\tau \geq \tau^0$, we have for any $d > 0$,

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

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

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

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

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

12. Here $\Big\|\sum\varepsilon_{tj}z_{t,-j}^T\Big\|_\infty = \max_{j,k}\Big|\sum\varepsilon_{tj}z_{t,-j,k}\Big|$.

w.p. 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 τ for $\tau \in \mathcal{G}(u_T, v_T)$, we obtain,

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

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

Lemma 22 *Suppose Condition B holds and let u_T, v_T be any non-negative sequences satisfying $0 \leq v_T \leq u_T \leq 1$. Then for any $0 < a < 1$, choosing $c_{a1} = 64c_{a2}$, with $c_{a2} \geq \sqrt{(1/a)}$, we have for $T \geq 2$,*

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

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

with probability at least $1 - a$.

Proof of Lemma 22 As before in Lemma 20, 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 44 and Lemma 43 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 19$\big)$ 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 46) we obtain,

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

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

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

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

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

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

Lemma 23 *Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.1) and suppose Condition B and C(ii) hold. Let u_T,v_T be any non-negative sequences satisfying $0\le v_T\le u_T\le 1$. Then for any $0<a<1$, choosing $c_{a1}=64c_{a2}$, with $c_{a2}\ge\surd(1/a)$, we have for $T\ge 2$,*

$$(i)\qquad \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\ge\tau^0}}\Phi(\eta^0,\eta^0)\ge v_T\kappa\xi_{2,2}^2-c_{a1}\sigma^2\xi_{2,2}^2\Big(\frac{u_T}{T}\Big)^{\frac12}$$

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

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

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

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

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

Proof of Lemma 23 Part (i) and Part (iii) of this lemma are a direct consequence of Lemma 22. To prove Part (ii), first note that,

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

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 36, we have,

$$\sup_{\tau\in\mathcal{G}(u_T,v_T)}\Phi(\hat\eta-\eta^0,\hat\eta-\eta^0)$$

$$\leq c_u(\sigma^2 \vee \phi) \log(p \vee T) u_T \Big(\sum_{j=1}^p \|\hat\eta_{(j)} - \eta^0\|_2^2 + \sum_{j=1}^p \|\hat\eta_{(j)} - \eta^0\|_1^2 \Big)$$
$$\leq c_u(\sigma^2 \vee \phi) s \log(p \vee T) u_T \sum_{j=1}^p \|\hat\eta_{(j)} - \eta^0_{(j)}\|_2^2$$

with probability at least $1-o(1)$. Here the final inequality follows by using (C.2). The proof of Part (iv) is an application of the Cauchy-Schwartz inequality together with the bounds of Part (ii) and Part (iii),

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \big|\Phi(\hat\eta - \eta^0, \eta^0)\big| \;\; \leq \;\; \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \big\{\Phi\big(\hat\eta - \eta^0, \hat\eta - \eta^0\big)\big\}^{\frac{1}{2}} \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \big\{\Phi\big(\eta^0, \eta^0\big)\big\}^{\frac{1}{2}}.$$

This completes the proof of this lemma. ■

Lemma 24 *Suppose Condition B and C(ii) hold. Let u_T, v_T be any non-negative sequences satisfying $0 \leq v_T \leq u_T \leq 1$. Then for any $0 < a < 1$, choosing $c_{a1} = 4{\cdot}48 c_{a2}$, with $c_{a2} \geq \surd(1/a)$, and for $u_T \geq c_{a1}^2\sigma^4/(T\phi^2)$, we have for $T \geq 2$,*

$$(i) \quad \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \frac{1}{T} \sum_{t=\tau^0+1}^{\tau} \sum_{j=1}^p \big(\hat\eta_{(j)}^T z_{t,-j}\big)^2 \geq$$
$$\kappa\xi_{2,2}^2 \Big[v_T - \frac{c_{a1}\sigma^2}{\kappa}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}} - c_u(\sigma^2 \vee \phi)\frac{u_T}{\kappa\xi_{2,2}}\Big\{s\log(p\vee T)\sum_{j=1}^p \|\hat\eta_{(j)} - \eta^0_{(j)}\|_2^2\Big\}^{\frac{1}{2}}\Big]$$

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

$$(iii) \quad \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \frac{1}{T}\Big| \sum_{t=\tau^0+1}^{\tau} \sum_{j=1}^p \varepsilon_{tj} z_{t,-j}^T \hat\eta_{(j)}\Big| \leq$$
$$c_{a1}\surd(1+\nu^2)\sigma^2\xi_{2,1}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}} + c_u\surd(1+\nu^2)\sigma^2\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\log(p\vee T)\sum_{j=1}^p \|\hat\eta_{(j)} - \eta^0_{(j)}\|_1,$$

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

Proof of Lemma 24 Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.1). 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 23 we obtain,

$$\inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\hat\eta, \hat\eta) \;\; \geq \;\; \inf_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} \Phi(\eta^0, \eta^0) - 2 \sup_{\substack{\tau\in\mathcal{G}(u_T,v_T);\\ \tau\geq\tau^0}} |\Phi(\hat\eta - \eta^0, \eta^0)|$$

$$\geq \quad \upsilon_T\kappa\xi_{2,2}^2 - c_{a1}\sigma^2\xi_{2,2}^2\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}$$
$$-c_u(\sigma^2\vee\phi)u_T\xi_{2,2}\Big(s\log(p\vee T)\sum_{j=1}^p\|\hat\eta_{(j)}-\eta^0_{(j)}\|_2^2\Big)^{\frac{1}{2}}$$

with probability at least $1-a-o(1)$. To prove Part (ii), note that using identical arguments as in the proof of Lemma 23 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-Schwartz inequality yields,

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

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

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

Now using Lemma 20 we have for any $0<a<1$, $R1\leq c_{a1}\surd(1+\nu^2)\sigma^2\xi_{2,1}\big(u_T/T\big)^{1/2}$, with probability at least $1-a$. Also, using Lemma 21 we have,

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

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

Lemma 25 *Suppose Condition A and C hold. Then we have,*

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

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

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

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

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

Proof of Lemma 25 Part (i) can be obtained as,

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

with probability at least $1-o(1)$. Here the final inequality follows from (1.6). 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,

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

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,

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

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

Lemma 26 *Let $\mathcal{C}(\tau,\mu,\gamma)$ be as defined in (A.8) and suppose Condition A, B and C hold. Additionally assume that the relation (2.4) holds. Then, for any $c_u > 0$, we have,*

$$\sup_{\tau\in\mathcal{G}\big((c_uT^{-1}\psi^{-2}),0\big)}\big|\mathcal{C}(\tau,\hat\mu,\hat\gamma)-\mathcal{C}(\tau,\mu^0,\gamma^0)\big| = o_p(1)$$

Proof of Lemma 26 For any $\tau \ge \tau^0$, first define the following,

$$\begin{aligned}
R_1 &= p^{-1}\sum_{\tau^0+1}^{\tau}\sum_{j=1}^{p}\|z_{t,-j}^T\hat\eta_{(j)}\|_2^2 - 2p^{-1}\sum_{\tau^0+1}^{\tau}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\hat\eta_{(j)}\\
&\quad +2p^{-1}\sum_{\tau^0+1}^{\tau}\sum_{j=1}^{p}(\hat\gamma_{(j)}-\gamma_{(j)}^0)^Tz_{t,-j}z_{t,-j}^T\hat\eta_{(j)}\\
&= R_{11}-2R_{12}+2R_{13},\\
R_2 &= p^{-1}\sum_{\tau^0+1}^{\tau}\sum_{j=1}^{p}\|z_{t,-j}^T\eta_{(j)}^0\|_2^2 - 2p^{-1}\sum_{\tau^0+1}^{\tau}\sum_{j=1}^{p}\varepsilon_{tj}z_{t,-j}^T\eta_{(j)}^0\\
&= R_{21}-2R_{22}. \qquad\qquad (C.3)
\end{aligned}$$

Then we have the following algebraic expansion,

$$\begin{aligned}
\big(\mathcal{C}(\tau,\hat\mu,\hat\gamma)-\mathcal{C}(\tau,\mu^0,\gamma^0)\big) &= -Tp^{-1}\Big(Q(z,\tau,\hat\mu,\hat\gamma)-Q(z,\tau^0,\hat\mu,\hat\gamma)\Big)\\
&\quad +Tp^{-1}\Big(Q(z,\tau,\mu^0,\gamma^0)-Q(z,\tau^0,\mu^0,\gamma^0)\Big)\\
&= \big(R_2-R_1\big)\\
&= \Big\{\big(R_{21}-2R_{22}\big)-\big(R_{11}-2R_{12}+2R_{13}\big)\Big\}. \qquad (C.4)
\end{aligned}$$

Lemma 27 shows that the expressions $\big|R_{21}-R_{11}\big|$, $\big|R_{22}-R_{12}\big|$, and $|R_{13}|$ are $o_p(1)$ uniformly over the set $\{\mathcal{G}\big(c_1T^{-1}\psi^{-2},0\big)\}\cap\{\tau\ge\tau^0\}$. The same result can be obtained symmetrically on the set $\{\mathcal{G}\big(c_uT^{-1}\psi^{-2},0\big)\}\cap\{\tau\le\tau^0\}$, thereby yielding $o_p(1)$ bounds for these terms uniformly over $\mathcal{G}\big(c_uT^{-1}\psi^{-2},0\big)$ Consequently,

$$\begin{aligned}
\sup_{\tau\in\mathcal{G}\big((c_uT^{-1}\psi^{-2}),0\big)}\big|\mathcal{C}(\tau,\hat\mu,\hat\gamma)-\mathcal{C}(\tau,\mu^0,\gamma^0)\big| &\le \sup_{\tau\in\mathcal{G}\big((c_uT^{-1}\psi^{-2}),0\big)}|R_{21}-R_{11}|\\
&\quad + \sup_{\tau\in\mathcal{G}\big((c_uT^{-1}\psi^{-2}),0\big)}2|R_{22}-R_{12}|\\
&\quad + \sup_{\tau\in\mathcal{G}\big((c_uT^{-1}\psi^{-2}),0\big)}2|R_{13}|\\
&= o_p(1)
\end{aligned}$$

This completes the proof of this lemma. ■

Lemma 27 *Suppose Condition A, B and C hold and additionally assume that relation (2.4) holds. Let* R_{11}, R_{12}, R_{13}, *and* R_{21}, R_{22} *be as defined in (C.3). Let* $0 < c_u < \infty$ *be any constant, then we have the following bounds.*

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

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

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

Proof of Lemma 27 Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.1) and consider,

$$\begin{aligned}
&\sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} |R_{11} - R_{21}| \\
&\quad= \sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} p^{-1}\Big| \sum_{\tau^0+1}^{\tau} \sum_{j=1}^{p} \Big(\|z_{t,-j}^T \hat{\eta}_{(j)}\|_2^2 - \|z_{t,-j}^T \eta_{(j)}^0\|_2^2 \Big) \Big| \\
&\quad= \sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} p^{-1}\Big| \sum_{\tau^0+1}^{\tau} \sum_{j=1}^{p} (\hat{\eta}_{(j)} - \eta_{(j)}^0)^T z_{t,-j} z_{t,-j}^T (\hat{\eta}_{(j)} + \eta_{(j)}^0) \Big| \\
&= \sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} \Big| Tp^{-1}\Phi(\hat{\eta} - \eta^0, \hat{\eta} - \eta^0) + 2Tp^{-1}\Phi(\hat{\eta} - \eta^0, \eta^0) \Big|. \qquad (C.5)
\end{aligned}$$

Now from Part (ii) of Lemma 23 we have

$$\begin{aligned}
&\sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} Tp^{-1}\Phi(\hat{\eta} - \eta^0, \hat{\eta} - \eta^0) \\
&\qquad\qquad \leq c_u c_1 (\sigma^2 \vee \phi)\psi^{-2} p^{-1} s \log(p \vee T) \sum_{j=1}^{p} \|\hat{\eta}_{(j)} - \eta_{(j)}^0\|_2^2 \\
&\qquad\qquad = O\Big(\frac{s^2 \log^2(p \vee T)}{\psi^{-2} T l_T} \Big) = o(1), \qquad (C.6)
\end{aligned}$$

with probability at least $1 - o(1)$. Also, from Part (iv) of Lemma 23, we have for $u_T \geq c_{a1}^2 \sigma_x^4 / (T\phi^2)$,

$$\sup_{\substack{\tau \in \mathcal{G}(u_T,0);\\ \tau \geq \tau^0}} 2Tp^{-1}\big|\Phi(\hat{\eta} - \eta^0, \eta^0)\big|$$

$$\leq c_u(\sigma^2 \vee \phi) T u_T p^{-1} \xi_{2,2} \Big(s \log(p \vee T) \sum_{j=1}^{p} \|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2^2 \Big)^{\frac{1}{2}} \tag{C.7}$$

with probability at least $1 - a - o(1)$. Upon choosing $a = \big(64^2\psi^2\sigma^4\big)/(c_1\phi^2) \to 0$, we have $c_1 T^{-1}\psi^{-2} = c_{a1}^2\sigma_x^4/(T\phi^2)$, consequently from (C.7) we have,

$$\begin{aligned}
&\sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} 2T\big|\Phi(\hat{\eta} - \eta^0, \eta^0)\big| \\
&\quad \leq c_u c_1 (\sigma^2 \vee \phi) \frac{\xi_{2,2}}{p\psi^2} \Big(s \log(p \vee T) \sum_{j=1}^{p} \|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2^2 \Big)^{\frac{1}{2}} \\
&\quad = c_u c_1 (\sigma^2 \vee \phi) \frac{1}{\xi_{2,2}} \Big(s \log(p \vee T) \sum_{j=1}^{p} \|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2^2 \Big)^{\frac{1}{2}} \\
&\quad \leq O\Big(\frac{1}{\psi} \frac{s \log(p \vee T)}{\sqrt{(T l_T)}} \Big) = o(1) \qquad \text{(C.8)}
\end{aligned}$$

with probability at least $1 - a - o(1) = 1 - o(1)$. Substituting this uniform bound together with (C.6) back in (C.5) yields Part (i) of this lemma. To prove Part (ii), note that

$$\begin{aligned}
\sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} |R_{12} - R_{22}| &= \sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} p^{-1} \Big| \sum_{\tau^0+1}^{\tau} \sum_{j=1}^{p} \varepsilon_{tj} z_{t,-j}^T (\hat{\eta}_{(j)} - \eta^0_{(j)}) \Big| \\
&= O\Big(p^{-1}\psi^{-1} \log(p \vee T) \sum_{j=1}^{p} \|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_1 \Big) \\
&\leq O\Big(\frac{s \log^{3/2}(p \vee T)}{\psi\sqrt{(T l_T)}} \Big) = o(1),
\end{aligned}$$

with probability at least $1 - o(1)$. Here the second equality follows from Part (ii) of Lemma 21. 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.6) and (C.8), respectively. Now applications of the Cauchy-Schwartz inequality yields the following bound for the term $|R_{13}|$.

$$\begin{aligned}
\sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} |R_{13}| &= \sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} \Big| \sum_{\tau^0+1}^{\tau} \sum_{j=1}^{p} (\hat{\gamma}_{(j)} - \gamma^0_{(j)})^T z_{t,-j} z_{t,-j}^T \hat{\eta}_{(j)} \Big| \\
&\leq \sup_{\substack{\tau \in \mathcal{G}\big((c_1 T^{-1}\psi^{-2}),0\big);\\ \tau \geq \tau^0}} T\Big\{ \big|\Phi(\hat{\gamma} - \gamma^0, \hat{\eta} - \eta^0)\big| + \big|\Phi(\hat{\gamma} - \gamma^0, \eta^0)\big| \Big\} = o(1),
\end{aligned}$$

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

Lemma 28 *Suppose Condition B holds and that $\psi \to 0$. Then for any constant $r > 0$, we have,*

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

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

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

Proof of Lemma 28 From Lemma 43 and Lemma 44 we have,

$$p^{-1}\psi^{-2}\sum_{j=1}^{p}\Big(\big\|z_{t,-j}^T\eta_{(j)}^0\big\|_2^2 - E\big\|z_{t,-j}^T\eta_{(j)}^0\big\|_2^2\Big) \sim \text{subE}(\lambda),\;\; \lambda = 16\sigma^2. \tag{C.9}$$

Now upon applying Bernstein's inequality (Lemma 45), we obtain for any $d > 0$,

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

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 $(d/\lambda) \le 1$ for T large, we obtain,

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

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

Appendix D. Deviation bounds used for proofs in Section 3

Lemma 29 *Suppose Condition A' and B holds, and $c_{u1} > 0$ be any constant. Then uniformly over $j = 1,...,p$, we have,*

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

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,

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

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 29 We begin with proving the first bound. Using Lemma 19 we have that $\varepsilon_{tj}z_{t,-j,k}\sim\text{subE}(\lambda_1)$, with $\lambda_1=48\sigma^2\surd(1+\nu^2)$. For any $\tau\geq c_{u1}T\ell_T$, applying Lemma 45 we have for $d>0$,

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

Choose $d=c_u\lambda_1\surd\{\log(p\vee T)/\tau\}$, and recall that by choice we have $\tau\geq c_{u1}T\ell_T$, and from Condition A$'$ we have $\log(p\vee T)\leq c_{u1}T\ell_T$. Thus, $d/\lambda_1\leq 1$, and consequently $(d^2/\lambda_1^2)\leq(d/\lambda_1)$. Using these relations the above probability bound yields,

$$\frac{1}{\tau}\Big|\sum_{t=1}^{\tau}\varepsilon_{tj}z_{t,-j,k}\Big|\leq(c_u/\surd c_{u1})\lambda_1\Big\{\frac{\log(p\vee T)}{T\ell_T}\Big\}^{\frac{1}{2}}$$

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 τ.

To prove the second bound, first note that using similar arguments as in Lemma 19 we have that $\eta_{(j)}^{0T}z_{t,-j}z_{t,-j,k}-E\big(\eta_{(j)}^{0T}z_{t,-j}z_{t,-j,k}\big)\sim\text{subE}(\lambda_1)$, with $\lambda_1=48\sigma^2\|\eta_{(j)}^0\|_2$. For any $\tau\in\mathcal{G}(u_T,0)$, satisfying $\tau\geq c_{u1}T\ell_T$, applying a union bound over $k=1,...,p-1$, on the Bernstein's inequality (Lemma 19) yields the following probability bound,

$$\begin{aligned}pr\Big\{\Big\|\sum_{t=\tau^0+1}^{\tau}\big(\eta_{(j)}^{0T}z_{t,-j}z_{t,-j}-\eta_{(j)}^{0T}\Delta_{-j,-j}\big)\Big\|_\infty>d(\tau-\tau^0)\Big\}&\\ \leq 2p\exp\Big\{-\frac{(\tau-\tau^0)}{2}\Big(\frac{d^2}{\lambda_1^2}\wedge\frac{d}{\lambda_1}\Big)\Big\}&\qquad\text{(D.1)}\end{aligned}$$

Now upon choosing,

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

it can be verified that [13],

$$\begin{aligned}d\frac{(\tau-\tau^0)}{\tau}&\leq\frac{c_u}{c_{u1}}\lambda_1\max\Big\{\frac{\log(p\vee T)}{T\ell_T},\ \frac{u_T}{\ell_T}\Big\},\quad\text{and,}\\ \frac{(\tau-\tau^0)}{2}\Big(\frac{d^2}{\lambda_1^2}\wedge\frac{d}{\lambda_1}\Big)&=\frac{(c_u\wedge c_u^2)}{2}\log(p\vee T)\qquad\text{(D.2)}\end{aligned}$$

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

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

13. See, Remark 30

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,

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

Using this relation in the probability bound now yields,

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

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 a union bound over the at most T values of τ, and similarly over $j = 1, ..., p$, by using another union bound. This completes the proof of the lemma. ∎

Remark 30 *Consider,*

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

observe that when $\log(p\vee T)\big/(\tau-\tau^0) \ge 1$, *then the maximum of the two terms in the expression (D.3) is* $\log(p\vee T)\big/(\tau-\tau^0)$. *In this case,*

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

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

Lemma 31 *Suppose Condition B holds and let* ε_{tj} *be as defined in (2.3). Let* $\log(p\vee T) \le Tv_T \le Tu_T$ *be any non-negative sequences. Then for any* $c_u > 0$, *we have,*

$$\sup_{\substack{\tau\in\mathcal{G}(u_T,v_T)\\ \tau\ge\tau^0}}\frac{1}{T}\Big\|\sum_{t=\tau^0+1}^{\tau}\varepsilon_{tj}z_{t,-j}^T\Big\|_\infty \le 48\surd(2c_u)\sigma^2\surd(1+\nu^2)\Big(\frac{u_T\log(p\vee T)}{T}\Big)^{\frac{1}{2}},$$

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

Proof of Lemma 31 The proof of this result is very similar to that of Lemma 21, the difference being utilization of the additional assumption $Tv_T \ge \log(p\vee T)$, in order to obtain this sharper bound. Proceeding as in (C.1) we have,

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

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

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

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

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

with probability $1-2\exp\{-c_{u1}\log(p\vee T)\}$, 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 at most T values of τ for $\tau\in\mathcal{G}(u_T,v_T)$, yields the statement of the lemma. ∎

Lemma 32 *Let* $\Phi(\cdot\,,\cdot)$ *be as defined in (C.1) and suppose Condition B holds. Additionally, let* $0\leq u_T,\ v_T\leq 1$ *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 32 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 45) we obtain,

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

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

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

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

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

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 τ, yields the bound (D.5) uniformly over τ. The statements of this lemma are now a direct consequence. ■

Lemma 33 *Let* $\Phi(\cdot,\cdot)$ *be as defined in (C.1) and suppose Condition B holds. Let* $\check\mu_{(j)}$ *and* $\check\gamma_{(j)}$, $j=1,...,p$ *be Step 1 edge estimates of Algorithm 2, and* $0\leq u_T, v_T \leq 1$ *be any non-negative sequences satisfying* $\log(p\vee T)\leq Tv_T \leq Tu_T$. *Then,*

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

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

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

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

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

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

Proof of Lemma 33 Part (i) and Part (iii) are a direct consequence of Lemma 32. To prove Part (ii), first note from Theorem 10 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 $\|\check\eta_{(j)}-\eta_{(j)}^0\|_1 \leq c_u\surd s\|\check\eta_{(j)}-\eta_{(j)}^0\|_2$. (see, e.g. (C.2)). Now applying Part (ii) of 36 yields,

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

with probability at least $1-o(1)$. Part (iv) follows by an application of the Cauchy-Schwartz inequality together with the bounds of Part (ii) and Part (iii) (see, (C.3)). This completes the proof of this lemma. ■

Lemma 34 *Suppose Condition B holds. Let* $\check{\mu}_{(j)}$, $\check{\gamma}_{(j)}$, $j = 1, ..., p$ *be Step 1 estimates of Algorithm 2, and assume* $0 \leq u_T, v_T \leq 1$ *satisfy* $\log(p \vee T) \leq T v_T \leq T u_T$. *Then,*

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

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

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

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

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

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

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

Proof of Lemma 34 Let $\Phi(\cdot\,,\cdot)$ be as defined in (C.1). 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 33 we obtain,

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

with probability at least $1 - o(1)$. To prove Part (ii), note that using identical arguments as in the proof of Lemma 33 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 $\Phi\big(\check\gamma-\gamma^0,\check\eta\big)\le\big|\Phi(\check\gamma-\gamma^0,\check\eta-\eta^0)\big|+\big|\Phi(\check\gamma-\gamma^0,\eta^0)\big|$, together with applications of the Cauchy-Schwartz inequality yields,

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

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

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

Now using Lemma 31, we have

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

w.p. at least $1-o(1)$. Part (iv) now follows by combining bounds for terms $R1$ and $R2$. ■

Appendix E. Uniform (over τ) Restricted Eigenvalue Condition

Lemma 35 *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* $s\ge1$, *let* $\mathcal{K}_p(s)=\{\delta\in\mathbb{R}^p;\ \|\delta\|_1\le1,\ \|\delta\|_0\le s\}$. *Then for non-negative* $0\le v_T\le u_T\le1$, *and any* $d_1>0$, *we have* $T\ge2$,

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

Proof of Lemma 35 Consider any fixed $\delta \in \mathbb{R}^p$, with $\|\delta\|_2 \le 1$, then from Lemma 44 we have $\|z_t^T\delta\|_2^2 - E\|z_t^T\delta\|_2^2 \sim \text{subE}(\lambda)$, with $\lambda = 16\sigma^2$. Now, for any fixed $\tau \in \mathcal{G}(u_T, v_T)$, $\tau \ge \tau^0$ applying Lemma 45 (Bernstein's inequality) we have,

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

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

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

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 $s \ge 1$,

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

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 τ for $\tau \in \mathcal{G}(u_T, v_T)$, thus yielding the statement of this lemma. ■

Lemma 36 *Suppose Condition B holds and let $0 \le v_T \le u_T \le 1$ be any non-negative sequences. Then for all $\delta_{(j)} \in \mathbb{R}^{p-1}$, $j = 1, ..., p$, and $T \ge 2$, we have,*

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

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 $\delta_{(j)} \in \mathbb{R}^{p-1}$, $j = 1, ..., p$,

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

$$c_u(\sigma^2 \vee \phi)u_T\Big(\sum_{j=1}^{p}\|\delta_{(j)}\|_2^2 + \sum_{j=1}^{p}\|\delta_{(j)}\|_1^2\Big)$$

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

Proof of Lemma 36 w.l.o.g. assume $v_T \geq (1/T)$ $\big($see, Lemma 21$\big)$. Now for any $s\geq 1$, 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 35 over the components $j=1,...,p$ we obtain,

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

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 $\delta_{(j)}\in\mathbb{R}^{p-1}$, $j=1,...,p$,

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

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,

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

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 by 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 the result. ■

Lemma 37 *Suppose Condition A' and B hold, then for $i=1,2$,*

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

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

Lemma 37 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 37 has additional uniformity over τ, i and j, which can be follow by dditional union bounds.

Appendix F. Auxiliary results

In the following Definition's 38, 39, and Lemma's 40-45, we list well known properties of subgaussian and subexponential distributions. These are largely reproduced from Vershynin (2019) and Rigollet (2015). Theorem 46 and 47 below reproduce Kolmogorov's inequality and the argmax theorem. Lemma 48 provides an upper bound for the ℓ_2 norm of the parameter vectors defined in Section 1.

Definition 38 Sub-gaussian r.v.: *A random variable $X \in \mathbb{R}$ is said to be sub-gaussian with parameter $\sigma > 0$ $\big($denoted by $X \sim \mathrm{subG}(\sigma)\big)$ if $E(X) = 0$ and its moment generating function*

$$E(\mathrm{e}^{tX}) \leq \mathrm{e}^{t^2\sigma^2/2}, \qquad \forall\, t \in \mathbb{R} \tag{F.1}$$

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

Definition 39 Sub-exponential r.v.: *A random variable $X \in \mathbb{R}$ is said to be sub-exponential with parameter $\sigma > 0$ $\big($denoted by $X \sim \mathrm{subE}(\sigma)\big)$ 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 40 *[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 40 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,

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

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,

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

   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 41 (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 41 Consider $X \sim \mathrm{subG}(\sigma)$, and w.l.o.g assume that $\sigma = 1$ (else define $X^* = X/\sigma$). 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 40. 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 42 *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 42 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 43 *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 43 is analogous to that of Lemma 42 and is thus omitted.

Lemma 44 *Let* $X \sim \mathrm{subG}(\sigma)$ *then the random variable* $Z = X^2 - E[X^2]$ *is sub-exponential:* $Z \sim \mathrm{subE}(16\sigma^2)$.

The next result is Bernstein's inequality, reproduced from Lemma 1.13 of Rigollet (2015).

Lemma 45 (Bernstein's inequality) *Let* $X_1, X_2, ..., X_T$ *be independent random variables such that* $X_t \sim \mathrm{subE}(\sigma)$. *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 Kolmogorov's inequality reproduced from Hájek and Rényi (1955)

Theorem 46 (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* $\mathrm{var}(\xi_k) = D_k^2$ $(k = 1, 2, ...)$, *we have, for any $\varepsilon > 0$,*

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

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

Theorem 47 (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 $K \subseteq H$. 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 $\mathcal{M}_n(\hat{h}_n) \ge \sup_h \mathcal{M}_n(h) - o_p(1)$, then $\hat{h}_n \Rightarrow \hat{h}$ in H.*

Lemma 48 *Suppose condition B holds, and let $\mu^0_{(j)}$ and $\gamma^0_{(j)}$, be as defined in (1.2). Then we have,*

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

Consequently we also have that $\psi \le 2\nu < \infty$, where ψ is as defined in (1.3).

Proof of Lemma 48 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 $1/\phi \le \max_j |\Omega_{jj}| \le 1/\kappa$. Now note that the ℓ_2 norm of the rows (or columns) of Ω are bounded above, i.e., $\|\Omega_{j\cdot}\|_2 = \|\Omega e_j\|_2 \le 1/\kappa$. This finally implies that

$$\|\mu^0_{(j)}\|_2 = \| - \Omega_{-j,j}/\Omega_{jj}\|_2 \le \|\Omega_{j\cdot}\|_2/|\Omega_{jj}| \le \frac{\phi}{\kappa} = \nu \tag{F.3}$$

Since the r.h.s. in (F.3) is free of j, this implies that $\max_j \|\mu^0_{(j)}\| \le \nu$. Identical arguments can be used to show that $\max_j \|\gamma^0_{(j)}\| \le \nu$. These two statements together imply the statement of the lemma. ■