Efficient Oracle-based Estimation and Inference for Multiple Change Points in High-Dimensional Gaussian Graphical Models

Xianru Wang[1], Bin Liu*[2], Xinsheng Zhang[2], and Yufeng Liu[3]

[1]Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, China;

[2]Department of Statistics and Data Science, School of Management at Fudan University, Shanghai, China.

[3]Department of Statistics and Operations Research, Department of Biostatistics, Carolina Center for Genome Sciences, Linberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, U.S.A

Abstract

In this paper, we study efficient multiple change point estimation and inference for high-dimensional Gaussian Graphical Models (GGMs). Leveraging Nodewise Lasso for graph structure estimation, we propose a two-step change point detection framework and develop two algorithms: the Refined Dynamic Programming Algorithm (RDPA) and the Refined Wild Binary Segmentation (RWBS). Our estimation method

*Corresponding author: liubin0145@gmail.com

is formulated based on a penalized loss function, enabling accurate detection of change points. Under mild model assumptions, we establish the theoretical guarantees for both algorithms, including consistency in estimating the number of change points, oracle-level accuracy in locating change points, and optimal convergence rates for the underlying coefficient vectors. Finally, extensive numerical studies and an application to the S&P 500 dataset demonstrate the computational efficiency and statistical accuracy of our proposed methods.

Keywords: Graphical models, High dimension, Change point detection, Efficient computation.

1 Introduction

The structure of multivariate dependence is prevalent in datasets across various fields, including social sciences, finance and genetics. Networks or graphs serve as potent instruments for modeling the dependence relationships between nodes (variables). A standard network consists of a set of graphs $G = (V, E)$, where $V = \{1, \ldots, p\}$ is a vertex set and $E \subseteq V \times V$ is an edge set. Each node in V corresponds to an element of a p-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$. A pair (a, b) is contained in the edge set E if and only if X_a is conditionally dependent on X_b, given all remaining variables $\boldsymbol{X}_{V\setminus a,b} = \{X_k; k \in V \setminus a, b\}$. The dependency structure of a p-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ can be further described by a Gaussian graphical model(GGM), where we assume $\boldsymbol{X} \sim \mathbb{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Under the GGM, it is well known (Lauritzen 1996) that there is a deterministic relationship between the conditional independence and the $p \times p$ precision matrix $\boldsymbol{\Omega} = (\omega_{ab})_{1\le a,b\le p}$, where $\boldsymbol{\Omega} := \boldsymbol{\Sigma}^{-1}$. In particular, $X_a \perp X_b | \boldsymbol{X}_{\setminus\{a,b\}} \Leftrightarrow \omega_{ab} = 0$.

In this paper, we are interested in high dimensional GGMs with multiple change points as follows. Given a p-variate time-ordered observations:

$$\boldsymbol{X}^t = (X_1^t, ..., X_p^t)^\top \sim \mathbb{N}(\boldsymbol{\mu}, (\boldsymbol{\Omega}^t)^{-1}),\ t = 1, ..., n. \tag{1.1}$$

let $\widetilde{k} \geq 0$ be the true number of unknown change points along with the true location vector $\widetilde{\boldsymbol{\tau}} = (\widetilde{\tau}_0, \widetilde{\tau}_1 ..., \widetilde{\tau}_{\widetilde{k}}, \widetilde{\tau}_{\widetilde{k}+1})^\top$ with $0 = \widetilde{\tau}_0 < \widetilde{\tau}_1 < \widetilde{\tau}_2 < \ldots < \widetilde{\tau}_{\widetilde{k}} < \widetilde{\tau}_{\widetilde{k}+1} = 1$. Then, the unknown $\widetilde{k}$ change points divide the n ordered observations into $\widetilde{k}+1$ intervals and the underlying precision matrix $\boldsymbol{\Omega}^t$ have the following form:

$$\boldsymbol{\Omega}^t = \begin{cases} \boldsymbol{\Omega}_0, & \text{if } \widetilde{k} = 0, \\ \boldsymbol{\Omega}_0(j), & \text{if } \widetilde{\tau}_{j-1} < t/n \leqslant \widetilde{\tau}_j, j = 1, \ldots, \widetilde{k}+1. \end{cases}$$

For the above dynamic networks, the objectives of this paper are: (a) **Estimating Change Points**: estimating the number of change points $(\widetilde{k})$ and the locations of change points $(\widetilde{\boldsymbol{\tau}})$; (b) **Estimating Precision Matrices**: Recover the precision matrices for each segment $\boldsymbol{\Omega}^0(j)$, $j = 1, ..., \widetilde{k}+1$; (c) **Deriving Asymptotic Distribution**: Establish the limiting distribution of the estimated change points; (d) **Making Confidence Intervals**: Construct confidence intervals for change point estimates. Note that we consider a high dimensional framework with the diverging change point number in the sense that p and $\widetilde{k}$ are allowed to scale with the sample size n. Based on this, our main goal is to develop a complete framework of change point detection for high-dimensional GGMs, aiming to accurately estimate the number of change points $(\widetilde{k})$, the locations of change points $(\widetilde{\boldsymbol{\tau}})$ and the underlying precision matrices $\boldsymbol{\Omega}^0(j)$ in each segmentation achieving good statistical convergence rates as both p and n tend towards infinity.

Recent advances in change point detection have focused on models with mean vectors and covariance matrices, addressing variations in these components. Key contributions in both low- and high-dimensional settings are found in studies such as Jirak (2015); Cho (2016); Wang et al. (2021a); Wang and Feng (2023). Additionally, change point detection and localization in regression models have been thoroughly explored Leonardi and Bühlmann (2016); Lee et al. (2016); Wang et al. (2021b); Xu et al. (2023). In contrast, change point detection in Gaussian Graphical Models (GGMs) is more complex, as it requires estimating graphs with $p \times p$ parameter matrix rather than p dimensional vector,

adding both computational and theoretical challenges. Efficient algorithms are crucial for obtaining satisfactory change point estimators. Dynamic programming methods (Leonardi and Bühlmann, 2016) aim for global optimal partitions, providing high statistical efficiency but at a computational cost of $O(n^2)$ evaluations. In contrast, binary segmentation algorithms offer greater computational efficiency, requiring only $O(n \log n)$ evaluations Leonardi and Bühlmann (2016); Cho (2016). Variants such as wild binary segmentation Fryzlewicz et al. (2014) and seeded binary segmentation Kovács et al. (2023) have further enhanced binary segmentation. Notably, the computational cost per evaluation varies depending on the complexity of the considered problem (mean vectors, covariance matrices, regression models, and so on), which can be particularly time-consuming when both p and n are large. For change point detection in GGMs, even a single fit can have an expensive computational cost. Hence, it is desirable to propose computionally efficient algorithms to detect change points in GGMs under a high dimensional framework.

Recent studies on GGMs, particularly in high-dimensional frameworks, have emerged due to advances in analyzing high-throughput data. These studies often assume i.i.d. samples from p-dimensional Gaussian distributions, where p exceeds the sample size. Notably, Meinshausen and Bühlmann (2006) introduced neighborhood selection for structural recovery, and other methods, such as ℓ_1 penalized likelihood Banerjee et al. (2008); Friedman et al. (2008), are widely used for estimating graphical structures. Moving away from likelihood-based techniques, Liu (2013) applied a multiple testing approach, while Cai et al. (2011) developed a constrained ℓ_1 minimization strategy for precision matrix estimation under broader distributional assumptions.

While much research has focused on time-invariant networks, the study of time-varying networks remains underdeveloped. Low-dimensional settings have seen some progress, such as in Kolar and Xing (2012) and Harchaoui and Lévy-Leduc (2010), but high-dimensional time-varying networks pose new challenges. Existing methods, primarily designed for sin-

gle change point detection (Liu et al., 2021; Bybee and Atchadé, 2018), are limited in practical applications, as they assume at most one change point occurs, potentially missing structural shifts and affecting estimation accuracy. Based on the analysis in Section 5, we applied our method to estimate change points in the S&P 500 graph structure, treating stock prices as nodes. The period from 2019 to 2021, divided by two change points, aligns with different COVID-19 pandemic stages, with the estimated graph structures showing significant differences across phases (as shown in Figure 1). Though methods like Gibberd and Nelson (2017) and Gibberd and Roy (2017) use the group-fused graphical lasso for piecewise constant structures, they are computationally expensive and lack optimal convergence. Londschien et al. (2021) introduced a more efficient algorithm for change point detection in GGMs with missing data but lacks theoretical guarantees.

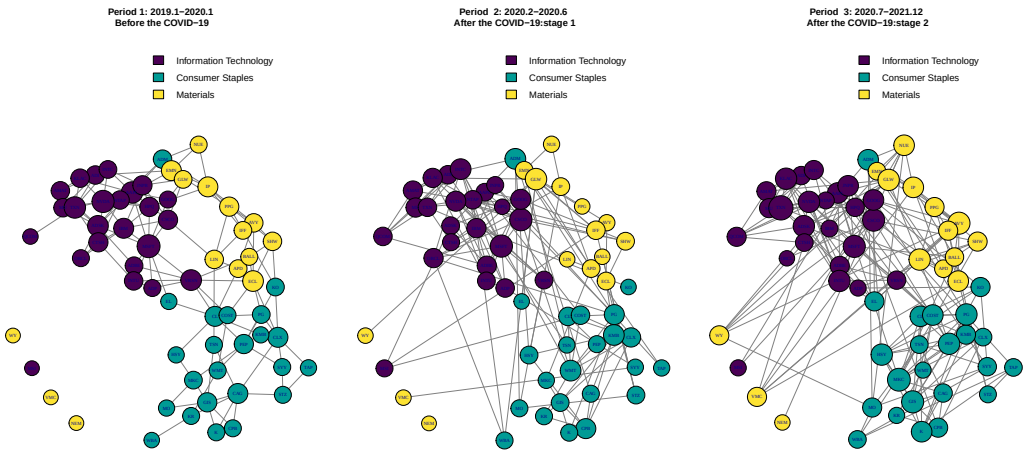

Figure 1: Network representation of 62 stocks from three categories in S&P500 dataset during different phases of the pandemic.

In change point analysis, estimation focuses on identifying the number and locations of change points, while inference extends further by quantifying the uncertainty in their positions through limit distributions and confidence intervals. This distinction is particularly critical in applications like finance, where differentiating structural shifts from random fluctuations informs decision-making. However, the stringent convergence rate requirements for change point estimators have led to relatively limited research on inference. Even in the case of traditionally low-dimensional settings, research is relatively scarce and primar-

ily focuses on mean vector parameters. Seminal contributions are Yao and Au (1989); Aue et al. (2009). Recent studies have extended change point analysis to high-dimensional mean structures, as seen in Wang and Shao (2023); Kaul et al. (2021); Kaul and Michailidis (2025). Bai (2010) investigated the construction of confidence intervals for change points in both mean and variance within panel data, utilizing the least squares loss function under the assumption of independence among components. Wang and Shao (2023) estimated the L_2-norm of mean jumps using U-statistics and constructed confidence intervals for a single change point, though their approach primarily focuses on the dense alternative hypothesis. In ultra-high-dimensional settings, Kaul et al. (2021); Kaul and Michailidis (2025) proposed confidence interval constructions for mean vectors under single and multiple change points, respectively, using the least squares loss function. However, compared to mean vectors, research on statistical inference for change points in ultra-high-dimensional Gaussian graphical models is relatively scarce, with one of the most recent works being Kaul et al. (2023). However, Kaul et al. (2023) studied confidence interval construction for a single change point in GGMs, yet inference for multiple change points remains largely unexplored. However, statistical inference for multiple change points poses greater challenges compared to a single change point. These challenges include accurately identifying the number of change points while ensuring that the estimated locations achieve theoretical optimality. Moreover, the estimation of ultra-high-dimensional parameters and the computational complexity of Gaussian graphical models further complicate the problem. Hence, we cannot directly extend the method and theory of Kaul et al. (2023) to the multiple change-point model setting.

To our knowledge, no existing work simultaneously addresses both estimation and inference in the multiple change point setting for GGMs. Building upon the existing research of change point detection in high dimensional GGMs, this paper aims to extend the current methodologies from single to multiple change point scenarios for oracle change point esti-

mation and inference, including deriving limit distribution and confident interval. While foundational studies such as those by Bybee and Atchadé (2018) and Liu et al. (2021) have significantly advanced single change point estimation, and others like Avanesov and Buzun (2018) and Kaul et al. (2023) have refined hypothesis testing and confidence interval estimations respectively, these primarily focus on **single change point**. Gibberd and Nelson (2017)'s work on multiple change point detection using the group fused graphical lasso method marks a significant step in handling multiple change points but is often constrained by **large computational demands** in high-dimensional settings particularly in cases where the number of observation and the number of nodes in the graph are both large. Combining all considerations, we would like to develop a new approach satisfying the following four properties: (1) adaptive to the change point number including $\widetilde{k} = 0$ and $\widetilde{k} \gg 0$; (2) sensitive for small signal jump size; (3) equipped with deeper inference, establishing limiting distributions and confidence intervals; (4) balancing statistical accuracy and computational efficiency in high-dimensional data. Our main contributions can be summarized as follows:

(1) Estimation. For high-dimensional graphical models (1.1) with multiple change points, where $(p \gg n)$, we propose a novel two-step refinement procedure for estimating both the change points and the underlying graphical model coefficients, building on the Nodewise Lasso approach for graphical structure estimation introduced by **?**. To implement this, we introduce a refined dynamic programming algorithm, RDPA, which serves as a benchmark for comparison. The refinement is carried out through a simplified dynamic programming procedure that enables global refinement of the change point locations. We demonstrate that this global refinement can deduce the local refinement procedure, leading to enhanced accuracy in refining the change points. Importantly, this refined procedure eliminates the need for repeated estimation of the graph structure through point-by-point scanning, thus significantly reducing the computational burden, making the refinement step

computationally negligible as compared to the initial estimation.

(2), **Algorithm.** Considering that RDPA is computationally expensive, with a complexity of $O(n^2\text{Graph}(n,p))$, where $\text{Graph}(n,p)$ represents the estimation complexity of the graph with sample size n and data dimension p, we aim to strike a balance between computational efficiency and statistical accuracy. To this end, we introduce a novel statistic based on nodewise Lasso, termed the gain function. This statistic quantifies the change in the loss function between large and small intervals, providing a more refined approach to change point detection. Building on this, we propose the Refined Wild Binary Segmentation (RWBS) algorithm, which utilizes the gain function to produce refined estimates while maintaining computational efficiency. The computational complexity of RWBS is $O(Mn\log(n)\text{Graph}(n,p))$, where M denotes the number of randomly generated intervals. Our simulation results in Section 4 show that RWBS delivers competitive statistical accuracy with significantly lower computational cost (for small values of M).

(3) **Optimal Theory.** Under the minimum signal-to-noise ratio (SNR) condition, we presents both RDPA and RWBS yield identical theoretical results in terms of change points estimation and inference. First, we proves that the estimator of change-point number is consistent in the sense of $\mathbb{P}\{\widehat{k} = \widetilde{k}\} \to 1$, when $n, p \to \infty$, where $\widetilde{k}$ can diverge with increasing sample n. Second, the estimator of change-point locations enjoy the oracle property in the sense that change-point estimator achieve the same convergence rate $O_p(\frac{1}{n})$ as the oracle estimator. Finally, the estimation of the coefficient vector can reach minimax optimization rate with $O_P(\sqrt{\log(p)/n})$ (**?**).

(4) **Change point inference.** For both RDPA and RWBS, based on the optimal theoretical properties of their refined change point estimates, we establish that under the assumption of vanishing signal strength, the limiting distribution of each detected change point follows a shifted two-sided Brownian motion. Consequently, we construct confidence intervals for each change point while allowing the number of change points to grow with

the sample size. Unlike existing studies that rely on local refinement methods, we first formulate a joint loss function for multiple change points and minimize it to obtain the optimal refined estimates. This approach eliminates the need for manually specifying local refitting intervals (as required in Xu et al. (2023) and Kaul et al. (2023)) and instead directly optimizes the global objective function, allowing the algorithm to automatically determine the optimal local intervals. Furthermore, although our final refined multiple change point estimates are obtained by minimizing the global objective function across the entire sample, we prove that the limiting distributions of the refined change points are asymptotically independent. While this result may seem counterintuitive, it enables the construction of joint confidence intervals for multiple change points. In contrast, Kaul et al. (2023) do not consider the multiple change point setting and therefore do not derive similar results. Our method is designed for multiple change points, allowing the number, locations, and signal strength of change points to scale with the sample size. We provide a unified theoretical framework for estimation and inference in the multiple change point setting, with theoretical results that encompass the existing single change point case. Consequently, our work establishes a comprehensive methodological and theoretical foundation for change point inference in high-dimensional Gaussian graphical models.

A cross-validation approach is developed in Supplementary Section A to select the optimal tuning parameters in a fully data-driven manner. Extensive numerical experiments across various settings validate the computational and statistical effectiveness of our method, demonstrating strong adaptability and reliability for change point estimation and inference. Application to the S&P 500 data further supports the validity and effectiveness of the proposed techniques in real-world scenarios.

The paper is organized as follows. Section 2 presents our methodology for identifying multiple change points and recovering networks in each segment. Section 3 provides theoretical results on change point identification, structure recovery, and confidence intervals.

Section 4 validates our method through extensive numerical studies. In Section 5, we apply the method to the S&P 500 dataset to analyze networks during the financial crisis. Conclusions are in Section 6, and the proofs are in the appendix.

We end this section with some notations. For any $n > 0$, define $\Gamma(n) = \{1, \ldots, n\}$. We set X_a^i as the i-th observation for coordinate a with $1 \leq i \leq n$ and $1 \leq a \leq p$. Denote $\mathbf{X}$ as the $n \times p$ observation matrix. We set $\mathbf{X}_i = (X_1^i, \ldots, X_p^i)^\top$ as the i-th row of $\mathbf{X}$ with $1 \leq i \leq n$ and $\mathbf{X}^a = (X_1^a, \ldots, X_n^a)^\top$ as the a-th column of $\mathbf{X}$ with $1 \leq a \leq p$. Given an interval $(u, v) \subset [0, 1]$ such that $u, v \in V_n = \{i/n, i = 1, ..., n, n \in \mathbb{N}\}$, we denote by $\mathbf{X}_{(u,v)}^{-a}$ as the $(v-u)n \times (p-1)$ dimensional matrix $\left(\boldsymbol{X}_{(u,v)}^1, \ldots, \boldsymbol{X}_{(u,v)}^{a-1}, \boldsymbol{X}_{(u,v)}^{a+1}, \ldots, \boldsymbol{X}_{(u,v)}^p\right)$, where $\boldsymbol{X}_{(u,v)}^j = (X_{un+1}^j, ..., X_{vn}^j)^\top$ with $j = 1, ..., p$, and $\widehat{\boldsymbol{\theta}}_{(u,v)}$ denote the lasso estimator based on the observations $\mathbf{X}_{(u,v)}$. Let $S_* = \cup_{j=1}^{\widetilde{k}+1} \cup_{a=1}^p S_a^{(j)}$, $S_a^{(j)} = \{i, \boldsymbol{\theta}_i^a(j) \neq 0\}$, and $s_* := \#S_*$ is the cardinality of S_*.

2 Methodology

In this section, we present our methodology for multiple change points estimation and inference in high-dimensional GGMs. Section 2.1 introduces the two-step Refined Dynamic Programming Algorithm (RDPA), which simultaneously estimates change points and graphical structures. Given the substantial computational cost of RDPA, Section 2.2 proposes a more efficient alternative—the two-step Refined Wild Binary Segmentation (RWBS) algorithm.

2.1 Two-step Refined Dynamic Programming Algorithm

Step 1: Initial joint estimation. Consider the p-dimensional multivariate normal distributed random variable $\boldsymbol{X} = (X_1, \ldots, X_p) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$, where $\boldsymbol{\Omega} = (\omega_{ab})_{a,b=1}^p \in \mathbb{R}^{p \times p}$ is the precision matrix. For each node $a \in \{1, \ldots, p\}$, consider the optimal prediction of $\boldsymbol{X}_a$, given all remaining variables. Let $\widetilde{\boldsymbol{\theta}}^a \in \mathbb{R}^p$ be the vector of coefficients for optimal

prediction being defined as

$$\widetilde{\boldsymbol{\theta}}^a = \underset{\boldsymbol{\theta}\in\mathbb{R}^p:\theta_a=0}{\arg\min}\, \mathbb{E}\Big(X_a - \sum_{k\in\Gamma(p)} \theta_k X_k\Big)^2.$$

Note that $\widetilde{\boldsymbol{\theta}}^a$ can be determined by the inverse covariance matrix $\boldsymbol{\Omega} = (\omega_{ab})_{a,b=1}^p \in \mathbb{R}^{p\times p}$. Specifically, according to Meinshausen and Bühlmann (2006), we have $\widetilde{\theta}_b^a = -\omega_{ab}/\omega_{aa}$ for $b \in V \setminus \{a\}$. Furthermore, it is also well known (Lauritzen, 1996) that for Gaussian distributions, $X_a \perp X_b | \boldsymbol{X}_{\setminus\{a,b\}} \Leftrightarrow \omega_{ab} = 0$. Define $\mathrm{NE}_a = \{b \in V \setminus \{a\} : \widetilde{\theta}_b^a \neq 0\}$ be the neighborhood of node a. Hence, the nonzero entries of $\widetilde{\boldsymbol{\theta}}^a$ correspond to the neighbors of node a. In literature, for homogeneous data with no change points, $\widetilde{\boldsymbol{\theta}}^a$ can be estimated by the nodewise Lasso estimation for high-dimensional graphical models, which is given by

$$\widehat{\boldsymbol{\theta}}^a = \underset{\theta:\theta_a=0}{\arg\min} \left(n^{-1} \left\| \mathbf{X}^a - \mathbf{X}^{-a}\boldsymbol{\theta} \right\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1 \right), \tag{2.1}$$

where $\|\boldsymbol{\theta}\|_1 = \sum_{b\in\Gamma(p)} |\theta_b|$ is the ℓ_1-norm of the coefficient vector.

Note that we consider the heterogeneous data with multiple change points $\widetilde{\boldsymbol{\tau}}$ with $\widetilde{\boldsymbol{\theta}}^a(1), \ldots, \widetilde{\boldsymbol{\theta}}^a(\widetilde{k}+1)$ being the true graphical structures for each node a of the $k+1$ segmentations. Hence, we cannot use (2.1) directly to obtain parameter estimation. For any given candidate change point partition $\boldsymbol{\tau}_k = (\tau_0, \tau_1, ..., \tau_k, \tau_{k+1})^\top$ with $k+1$ segmentations, we introduce our estimator with the Lasso penalty in each segmentation as follows:

$$\widehat{\boldsymbol{\theta}}^a(\boldsymbol{\tau}_k, j) = \underset{\boldsymbol{\theta}:\theta_a=0}{\arg\min} \left(L_n^a\left(I_j(\boldsymbol{\tau}_k), \boldsymbol{\theta}\right) + \lambda_n \sqrt{(\tau_j - \tau_{j-1})} \|\boldsymbol{\theta}\|_1 \right), \text{ for } j = 1, \ldots, k+1, \tag{2.2}$$

where $L_n^a\left(I_j(\boldsymbol{\tau}_k), \boldsymbol{\theta}\right) := \frac{1}{n}\|\mathbf{X}^a_{(\tau_{j-1},\tau_j)} - \mathbf{X}^{-a}_{(\tau_{j-1},\tau_j)}\boldsymbol{\theta}\|_2^2$, and λ_n is the non-negative regularization parameter for the sub-interval based Lasso estimation, which is discussed in Section 3.2. Note that $L_n^a(I_j(\boldsymbol{\tau}_k), \boldsymbol{\theta})$ can be regarded as the ℓ_2-loss in the j-th segmentation of $\boldsymbol{\tau}_k$ and $\widehat{\boldsymbol{\theta}}^a(\boldsymbol{\tau}_k, j)$ is the corresponding lasso estimate for node a in that segmentation.

The main challenge in estimating multiple change points in GGMs is the interdependence between the unknown change point number $\widetilde{k}$, locations $\widetilde{\boldsymbol{\tau}}$, and precision matrices in each segment. This complicates the estimation, as change point detection affects the graphical structure estimation and vice versa. To address this, we propose a refined two-step

approach. In the first step, we solve an optimization problem based on a loss function and the number of segments, providing initial estimates for both change points and graphical structures. However, due to the non-convex nature of this problem, the initial estimates have suboptimal convergence rates. The second step refines these estimates by using the initial p-dimensional coefficient vectors, improving the accuracy of the change point detection. The following sections discuss the refinement procedure and algorithm in detail.

For any candidate partition $\boldsymbol{\tau}_k = (\tau_0, \tau_1, ..., \tau_k, \tau_{k+1})^\top$ with k change points, let $\boldsymbol{\theta}^a(\boldsymbol{\tau}_k, j)$ be the coresponding coefficients of node a for the j-th segmentation of $\boldsymbol{\tau}_k$ and set $\boldsymbol{\theta}^a(\boldsymbol{\tau}_k) := (\boldsymbol{\theta}^a(\boldsymbol{\tau}_k, 1)^\top, \ldots, \boldsymbol{\theta}^a(\boldsymbol{\tau}_k, k+1)^\top)^\top \in \mathbb{R}^{p(k+1)}$ as coefficients for all $k+1$ segmentations. For the j-th segmentation with $j = 1, \ldots, k+1$, let $\boldsymbol{\Theta}(\boldsymbol{\tau}_k, j) := (\boldsymbol{\theta}^1(\boldsymbol{\tau}_k, j), \ldots, \boldsymbol{\theta}^p(\boldsymbol{\tau}_k, j))$ be the $\mathbb{R}^{p\times p}$ graphical structure. Note that by setting $\boldsymbol{\tau}_k = \widetilde{\boldsymbol{\tau}}_{\widetilde{k}}$ the underlying true regression coefficients can be correspondingly written as $\widetilde{\boldsymbol{\theta}}^a(\widetilde{\boldsymbol{\tau}}_{\widetilde{k}}) = (\widetilde{\boldsymbol{\theta}}^a(\widetilde{\boldsymbol{\tau}}_{\widetilde{k}}, 1), ..., \widetilde{\boldsymbol{\theta}}^a(\widetilde{\boldsymbol{\tau}}_{\widetilde{k}}, \widetilde{k}+1))^\top$. In this case, for simplicity, for each node a, we can abbreviate $\widetilde{\boldsymbol{\theta}}^a(\widetilde{\boldsymbol{\tau}}_{\widetilde{k}}, j)$ as $\widetilde{\boldsymbol{\theta}}^a(j)$ with $j = 1 \ldots, \widetilde{k}+1$ and abbreviate $\widetilde{\boldsymbol{\theta}}^a(\widetilde{\boldsymbol{\tau}}_{\widetilde{k}})$ as $\widetilde{\boldsymbol{\theta}}^a$. Similarly, for each segmentation $j = 1, \ldots, \widetilde{k}+1$, we can abbreviate the $\mathbb{R}^{p\times p}$ underlying true graphical structure $\widetilde{\boldsymbol{\Theta}}(\widetilde{\boldsymbol{\tau}}_{\widetilde{k}}, j)$ as $\widetilde{\boldsymbol{\Theta}}(j) := (\widetilde{\boldsymbol{\theta}}^1(j), \ldots, \widetilde{\boldsymbol{\theta}}^p(j))$.

An initial estimate of the change points and underlying parameters can be obtained using the following penalized loss function:

$$\widehat{\boldsymbol{\tau}}_{\widehat{k}} = \underset{k\in\{0,1,\ldots,k_{max}\}}{\arg\min} \underset{\boldsymbol{\tau}_k=(\tau_0,\tau_1,\ldots,\tau_k,\tau_{k+1})^\top}{\arg\min} \left\{ \sum_{j=1}^{k+1} \sum_{a=1}^{p} L_n^a \left(I_j(\boldsymbol{\tau}_k), \widehat{\boldsymbol{\theta}}^a(\boldsymbol{\tau}_k, j) \right) + \gamma(k+1) \right\}, \quad (2.3)$$

where $\widehat{\boldsymbol{\theta}}^a(\boldsymbol{\tau}_k, 1), \ldots, \widehat{\boldsymbol{\theta}}^a(\boldsymbol{\tau}_k, k+1)$ can be obtained by (2.2). Besides, based on the estimated change points $\widehat{\boldsymbol{\tau}}_{\widehat{k}}$, for each node a, we can obtain the initial estimators of the underlying coefficient vector $\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}}) := (\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}}, 1)^\top, \ldots, \widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}}, \widehat{k}+1)^\top)^\top$ by

$$\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}}, j) = \underset{\boldsymbol{\theta}:\theta_a=0}{\arg\min} \left(L_n^a \left(I_j(\widehat{\boldsymbol{\tau}}_{\widehat{k}}), \boldsymbol{\theta} \right) + \lambda_n \sqrt{\widehat{\tau}_j - \widehat{\tau}_{j-1}} \|\boldsymbol{\theta}\|_1 \right), \text{ for } j = 1, \ldots, \widehat{k}+1. \quad (2.4)$$

The optimization problem in (2.3) has two main components: the first term is the loss function for the p node-wise Lasso regressions across all segmentations, while the second term penalizes the number of segments (change points) to avoid overfitting. The objective

function balances these to optimally partition the data and estimate all coefficient vectors $\widetilde{\boldsymbol{\theta}}^a(j)_{a=1}^p$. Solving (2.3) yields joint estimates for change points and regression coefficients. The tuning parameters λ_n and γ control sparsity in the coefficient vectors and segmentation complexity, respectively. The interplay between these parameters highlights a fundamental challenge—balancing the accuracy of estimated graph structures with the precision of change point localization.

It is well established that the minimal partition problem in (2.3) can be efficiently solved using dynamic programming algorithms (DPA) (Auger and Lawrence, 1989). Upon completing Step 1, we obtain initial estimates for the number of change points, $\widehat{k}$, their locations, $\widehat{\boldsymbol{\tau}}\widehat{k}$, and the corresponding coefficient vectors, $\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}\widehat{k}, j)$, for each node $a = 1, ..., p$ and segment $j = 1, ..., \widehat{k} + 1$. These initial estimates, especially for change point locations and coefficient vectors, may not be optimal due to the challenges of jointly estimating high-dimensional parameters. To improve accuracy and convergence, a refinement procedure is introduced in the next step.

Step 2-1: Refined change point estimation. Recall $L_n^a\left(I_j(\boldsymbol{\tau}), \boldsymbol{\theta}\right) := \frac{1}{n}\|\mathbf{X}^a_{(\tau_{j-1},\tau_j)} - \mathbf{X}_{(\tau_{j-1},\tau_j)}\boldsymbol{\theta}\|_2^2$. For any candidate partition $\boldsymbol{\tau}_k = (\tau_0, \tau_1, ..., \tau_k, \tau_{k+1})^\top$ and $\boldsymbol{\theta}^a(\boldsymbol{\tau}_k) := (\boldsymbol{\theta}^a(\boldsymbol{\tau}_k, 1)^\top, \ldots, \boldsymbol{\theta}^a(\boldsymbol{\tau}_k, k+1)^\top)^\top \in \mathbb{R}^{p(k+1)}$, define

$$J_n(\boldsymbol{\tau}_k, \{\boldsymbol{\theta}^a(\boldsymbol{\tau}_k)\}_{a=1}^p) = \sum_{j=1}^{k+1}\sum_{a=1}^{p} L_n^a\left(I_j(\boldsymbol{\tau}_k), \boldsymbol{\theta}^a(\boldsymbol{\tau}_k, j)\right). \tag{2.5}$$

Note that $J_n(\boldsymbol{\tau}_k, \{\boldsymbol{\theta}^a(\boldsymbol{\tau}_k)\}_{a=1}^p)$ can be regarded as the total loss for fiting the Gaussian graphical models with $\boldsymbol{\tau}_k$ as the partition and $\{\boldsymbol{\theta}^a(\boldsymbol{\tau}_k)\}_{a=1}^p$ being the parameters in each partition. Before introducing the refined change point estimator, we first introduce the oracle change point estimator $\ddot{\boldsymbol{\tau}}$. Suppose that the true change point number $\widetilde{k}$ and true underlying regression coefficients for all segmentations are known. Then, we can put $\widetilde{k}$ and $\{\widetilde{\boldsymbol{\theta}}^a\}_{a=1}^p$ into $J_n(\boldsymbol{\tau}_k, \{\boldsymbol{\theta}^a(\boldsymbol{\tau}_k)\}_{a=1}^p)$ and define the oracle change point estimator $\ddot{\boldsymbol{\tau}}$ as follows:

$$
\begin{aligned}
\ddot{\boldsymbol{\tau}} &= \underset{\boldsymbol{\tau}_{\widetilde{k}}=(\tau_0,\ldots,\tau_{\widetilde{k}+1})^\top}{\arg\min} J_n(\boldsymbol{\tau}_{\widetilde{k}}, \{\widetilde{\boldsymbol{\theta}}^a(\widetilde{\boldsymbol{\tau}})\}_{a=1}^p), \\
&= \underset{\boldsymbol{\tau}_{\widetilde{k}}=(\tau_0,\tau_1,\ldots,\tau_{\widetilde{k}},\tau_{\widetilde{k}+1})^\top}{\arg\min} \frac{1}{n} \sum_{j=1}^{\widetilde{k}+1} \sum_{a=1}^{p} L_n^a \left(I_j(\boldsymbol{\tau}_{\widetilde{k}}), \widetilde{\boldsymbol{\theta}}^a(j) \right).
\end{aligned} \tag{2.6}
$$

The key idea behind $\ddot{\boldsymbol{\tau}}$ is that if the true number of change points and the underlying graphical structure within each segment were known, the estimation problem would reduce to finding the optimal partition that best fits the data. Under this oracle setting, the change point locations can be obtained by solving a simplified optimization problem. This allows for a more computationally efficient solution via a streamlined dynamic programming procedure. Specifically, we fix the number of change points at $\widetilde{k}$ and the corresponding coefficient vectors at $\widetilde{\boldsymbol{\theta}}^a(j)$ for $j = 1, \ldots, \widetilde{k}$, thereby focusing solely on refining the segmentation. For any $1 < s < e < n$, let

$$
L_n^a\left((s,e), \boldsymbol{\theta}^a\right) := \frac{1}{n} \|\mathbf{X}_{(s,e)}^a - \mathbf{X}_{(s,e)}^{-a} \boldsymbol{\theta}^a\|_2^2, \text{ and } L_n\left((s,e), \{\boldsymbol{\theta}^a\}_{a=1}^p\right) := \sum_{a=1}^{p} L_n^a\left((s,e), \boldsymbol{\theta}^a\right).
$$

For $v \in V_n = \{i/n : \ i = 1, ..., n\}$ and $k \geq 1$, let

$$
\begin{aligned}
&F_k(v, \{\widetilde{\boldsymbol{\theta}}^a(1)\}_{a=1}^p, ..., \{\widetilde{\boldsymbol{\theta}}^a(k+1)\}_{a=1}^p) \\
&= \min_{0<\tau_1<\cdots<\tau_k<v} \left\{ L_n((0,\tau_1), \{\widetilde{\boldsymbol{\theta}}^a(1)\}_{a=1}^p) + \cdots + L_n((\tau_k, v), \{\widetilde{\boldsymbol{\theta}}^a(k+1)\}_{a=1}^p) \right\}.
\end{aligned}
$$

Letting $F_0(v, \{\widetilde{\boldsymbol{\theta}}^a(1)\}_{a=1}^p) = L_n\big((0,v), \{\widetilde{\boldsymbol{\theta}}^a(1)\}_{a=1}^p\big)$, we have the following recurrence formula

$$
\begin{aligned}
&F_k(v, \{\widetilde{\boldsymbol{\theta}}^a(1)\}_{a=1}^p, ..., \{\widetilde{\boldsymbol{\theta}}^a(k+1)\}_{a=1}^p) \\
&= \min_{\substack{u \in V_n \\ u<v}} \left\{ F_{k-1}(u, \{\widetilde{\boldsymbol{\theta}}^a(1)\}_{a=1}^p, ..., , \{\widetilde{\boldsymbol{\theta}}^a(k)\}_{a=1}^p) + L_n\big((u,v), \{\widetilde{\boldsymbol{\theta}}^a(k+1)\}_{a=1}^p\big) \right\}.
\end{aligned} \tag{2.7}
$$

Then the final oracle estimator $\ddot{\boldsymbol{\tau}} = (0, \ddot{\tau}_1, ..., \ddot{\tau}_{\widetilde{k}}, 1)^\top$ of change points results from solving the following optimization problem by a simplified dynamic programming procedure. Specifically, for $j = \widetilde{k}, \ldots, 1$, we calculate

$$
\ddot{\tau}_j = \underset{u \in V_n, u<\ddot{\tau}_{j+1}}{\arg\min} \{F_{j-1}(u, \{\widetilde{\boldsymbol{\theta}}^a(1)\}_{a=1}^p, ..., \{\widetilde{\boldsymbol{\theta}}^a(j)\}_{a=1}^p) + L_n\big((u, \ddot{\tau}_{j+1}), \{\widetilde{\boldsymbol{\theta}}^a(j+1)\}_{a=1}^p\big), \tag{2.8}
$$

where $\ddot{\tau}_0 := 1$ and $\ddot{\tau}_{\widetilde{k}+1} := 1$. It is proven that $\ddot{\boldsymbol{\tau}}$ has some oracle propeties since it has the best convergence rates. See Lee et al. (2018). In practice, both the change point number and the underlying regression coefficients are typically unknown. However, the estimates

obtained in the first step can provide valuable information, effectively serving as a useful "prior". These initial estimates, while not definitive, offer a reasonable approximation that can guide subsequent analysis. Specifically, fixing the number of change points $\widehat{k}$ in (2.2), we put $\{\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}})\}_{a=1}^p$ obtained in (2.4) into $J_n(\boldsymbol{\tau}_k, \{\boldsymbol{\theta}^a(\boldsymbol{\tau}_k)\}_{a=1}^p)$ and obtain the improved change point estimator by:

$$\begin{aligned} \widehat{\boldsymbol{\tau}}^r &= \underset{\boldsymbol{\tau}_{\widehat{k}}=(\tau_0,\ldots,\tau_{\widehat{k}+1})^\top}{\arg\min} J_n(\boldsymbol{\tau}_{\widehat{k}}, \{\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}})\}_{a=1}^p), \\ &= \underset{\boldsymbol{\tau}_{\widehat{k}}=(\tau_0,\ldots,\tau_{\widehat{k}+1})^\top}{\arg\min} \frac{1}{n} \sum_{j=1}^{\widehat{k}+1} \sum_{a=1}^{p} L_n^a \left(I_j(\boldsymbol{\tau}_{\widehat{k}}), \widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}}, j) \right), \end{aligned} \tag{2.9}$$

where $\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}}, j)$ with $j = 1, ..., \widehat{k}+1$ are obtained by (2.4).

Step 2-1 reduces computational cost by using the estimator $\widehat{k}$ for the number of change points and the initial Lasso estimator $\{\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}_{\widehat{k}})\}_{a=1}^p$, avoiding the need for screening and recomputing the Lasso in each segmentation. This is similar to the oracle change point estimator $\ddot{\boldsymbol{\tau}}$, where a simplified dynamic programming procedure, as outlined in (2.7) - (2.8), can be used to compute $\widehat{\boldsymbol{\tau}}^r$ in (2.9). Combining Steps 1 and 2, we present the Refined Dynamic Programming Algorithm (RDPA) in Algorithm A.1 in Supplementary Section A for high-dimensional GGMs.

Step 2-2: Update the graphical structure estimation. Based on the refined change point estimator $\widehat{\boldsymbol{\tau}}^r$, for each node $a = 1, \ldots, p$, we can further obtain the final coefficients' estimator by

$$\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}^r, j) = \underset{\boldsymbol{\theta}:\theta_a=0}{\arg\min} \left(L_n^a \left(I_j(\widehat{\boldsymbol{\tau}}^r), \boldsymbol{\theta} \right) + \lambda_n^r \sqrt{\widehat{\tau}_j^r - \widehat{\tau}_{j-1}^r} \|\boldsymbol{\theta}\|_1 \right), \text{ for } j = 1, \ldots, \widehat{k}+1, \tag{2.10}$$

where λ_n^r is the non-negative regularization parameter for the refined step, which is discussed in Section 3.2.

Step 2-3: Construct confidence interval for the change point locations. In this step, we make inference on the detected change points. Let $\widetilde{\boldsymbol{\delta}}^a(j) := \widetilde{\boldsymbol{\theta}}^a(j+1) - \widetilde{\boldsymbol{\theta}}^a(j)$ be the true singal jump for node a at the j-th change point with $j = 1, \ldots, \widetilde{k}$. Define $\widetilde{\Delta}(j) = \sqrt{\sum_{a=1}^p \|\widetilde{\boldsymbol{\delta}}^a(j)\|^2}$ and $\widetilde{\psi}(j) = \widetilde{\Delta}(j)/\sqrt{p}$. Thanks to the two-step refinement, we can prove in Theorem 3.2 that

$$|\widehat{\eta}_j^{\,r} - \widetilde{\eta}_j| = O_p\big(\frac{1}{\widetilde{\psi}^2(j)}\big), \quad \text{for} \quad j = 1, \ldots, \widetilde{k}, \tag{2.11}$$

where $\widetilde{\eta}_j := n\widetilde{\tau}_j$ and $\widehat{\eta}_j^{\,r} := n\widehat{\tau}_j^{\,r}$. The convergence rate in (2.11) is known to have the best convergence rate. This further helps us to make inference for the detected change points. Before making confidence interval for the detected change points, we introduce some notations. Specifically, let $W_j^{(1)}(r)$ and $W_j^{(2)}(r)$ be two independent Brownian motions defined on $[0, \infty)$. For each $j = \widetilde{k}, \ldots, 1$, define the following process

$$Z_j(r) = \begin{cases} 2\sigma_1(j)W_j^{(1)}(r) - (\sigma_1^*(j))^2|r|, & \text{if } r > 0, \\ 0, & \text{if } r = 0, \\ 2\sigma_2(j)W_j^{(2)}(-r) - (\sigma_2^*(j))^2|r| & \text{if } r < 0, \end{cases} \tag{2.12}$$

where $0 < \sigma_1(j), \sigma_2(j), \sigma_1^*(j), \sigma_2^*(j) < \infty$, are parameters that control both the variance and the negative drift of the process $Z_j(r)$. The detailed definitions of these parameters are definded in Assumption E. Then, for each $j = 1, \ldots, \widetilde{k}$, under the vanishing regime with $\widetilde{\psi}(j) \to 0$, we can prove in Theorem 3.8 that

$$\widetilde{\psi}^2(j)\,(\widehat{\eta}_j^{\,r} - \widetilde{\eta}_j) \Rightarrow^d \arg\max_{r \in \mathbb{R}} Z_j(r). \tag{2.13}$$

Let $q_{\alpha/2}$ and $q_{1-\alpha/2}$ be the quantiles of $\arg\max_{r \in \mathbb{R}} Z_j(r)$. Let $\big(\widehat{\sigma}_1(j), \widehat{\sigma}_2(j), \widehat{\sigma}_1^*(j), \widehat{\sigma}_2^*(j), \widehat{\psi}^2(j)\big)$ be some estimators for $\big(\sigma_1(j), \sigma_2(j), \sigma_1^*(j), \sigma_2^*(j), \widetilde{\psi}^2(j)\big)$, respectively. Then, the $1-\alpha$ level empirical confidence interval for $\widetilde{\eta}_j$ can be constructed as:

$$\Big[\widehat{l}_\alpha(j), \widehat{u}_\alpha(j)\Big] := \Big[\widehat{\eta}_j^{\,r} - \frac{q_{1-\alpha/2}}{\widehat{\psi}^2(j)}, \widehat{\eta}_j^{\,r} - \frac{q_{\alpha/2}}{\widehat{\psi}^2(j)}\Big]. \tag{2.14}$$

The detailed implementation for obtaining $[\widehat{l}_\alpha(j), \widehat{u}_\alpha(j)]$ can be found in Appendix B.

We present the theoretical results of RDPA in Section 3.2. However, The Step 1 in RDPA is very time-consuming with computational complexity of $O(n^2\text{Gragh}(n,p))$, where $\text{Gragh}(n,p)$ is the computational cost for solving the high dimensional GGM with sample size n and data dimension p. Our next task is to develop a computationally efficient algorithm to detect multiple change points for high dimensional GGMs.

2.2 Refined Wild Binary Segmentation Through Local Lasso Estimator

In this section, to improve RDPA, we develop a much more efficient Refined Wild Binary Segmentation (RWBS) approach for multiple change-point estimation in high dimensional GGMs. Note that WBS is the modified version of the binary segmentation and has been used for estimating multiple change points in different senaroes. Recently, it has also been widely applied to estimate multiple change points in different high-dimensional models. For instance, Wang and Samworth (2018) employed it for high-dimensional mean vectors, Wang et al. (2021b) for regression coefficients, and Wang et al. (2021a) for covariance matrices.

It is worth noting that the majority of algorithms for estimating multiple change points based on WBS require the construction of CUSUM statistics. Unlike existing studies, ultra-high-dimensional Gaussian graphical models lack a straightforward construction method similar to the CUSUM for mean vectors. Therefore, they often need to be combined with Lasso. As is well known, Lasso estimates tend to be biased and require further bias correction. To avoid this issue, we will construct testing statistics based on the loss function and combine them with the WBS algorithm to obtain initial estimates of the change points. This step highly improves the computational efficiency for change point detection compared to Step 1 of RDPA. As for Step 2 of RWBS, we use a refined procedure with the initial estimators obtained by Step 1 of RWBS as input to improve the statistical accuracy of estimators for both change points and the graphical structures. The convergence rates of change point estimates achieves theoretical optimality by using this two-step based estimation procedure. This allows us to construct confidence intervals for multiple change points in the final step.

Next, we introduce RWBS to implement our efficient two-step estimation. For heterogeneous observations with multiple change points, we apply Wild Binary Segmentation using the Lasso estimator. To be specific, given any search interval $(s, e) \subset (0, 1)$, define

$$Z(s,e) = \begin{cases} \sum_{a=1}^{p} L_n^a\left((s,e], \widehat{\boldsymbol{\theta}}^a_{(s,e]}\right), & \text{if } (e-s)n \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.15}$$

where $\widehat{\boldsymbol{\theta}}^a_{(s,e)}$ is the Lasso estimate for the (s,e)-th sub-interval:

$$\widehat{\boldsymbol{\theta}}^a_{(s,e)} := \underset{\boldsymbol{\theta}:\theta_a=0}{\arg\min}\Big(\frac{1}{n}\sum_{i=s+1}^{e}\left\|\mathbf{X}^a_{(s,e)} - \mathbf{X}^{-a}_{(s,e)}\boldsymbol{\theta}\right\|_2^2 + \lambda_n\sqrt{(e-s)}\|\boldsymbol{\theta}\|_1\Big). \tag{2.16}$$

Define the statistics as

$$\mathcal{B}_t^{s,e} = Z(s,e) - (Z(s,t) + Z(t,e)), t \in (s,e). \tag{2.17}$$

Remark 2.1. (Insight of gain function: improvement in model fit) The gain function is defined as the difference between the loss function of a large interval (s,e) assuming no change point, and the sum of the loss functions for two smaller intervals after a potential split at t. It quantifies the improvement in model fit when assuming a change point at some location t within (s,e). When there is exactly one true change point at τ, the gain function identifies it by showing that the sum of the loss functions for the two smaller intervals $(Z(s,t)+Z(t,e))$ is minimized at $t=\tau$. Thus, the split at τ maximizes the gain function: $\tau = \arg\max_{t\in(s,e)} \mathcal{B}_t^{s,e} = Z(s,e) - (Z(s,t)+Z(t,e))$. This is similar to the CUSUM statistic, which identifies the point of maximum deviation from a reference mean. Both methods aim to locate the point where the data "fit" improves the most, with CUSUM tracking mean shifts and the gain function tracking model fit improvements.

Then we are ready to describe the main algorithm, which we call it refined Wild Binary Segmentation through Lasso Estimator. This algorithm makes use of a collection of pre-generated interval $\{s_m, e_m\}_{m=1}^M, 0 < s_m < e_m \leq 1$, as input, which is first used in Wild Binary Segmentation (WBS) of Fryzlewicz et al. (2014). To be specific, every new change point is found by screening the M intervals, which are selected randomly and independently. The key point of WBS is if M is large enough, then with high probability, there exists at least one random interval (s_m, e_m) containing one and only one change point. In other words, the "good" event

$$\mathcal{M} = \bigcap_{k=1}^{K} \{\alpha_m \in \mathcal{S}_k, \gamma_m \in \mathcal{E}_k, \text{ for some } m \in \{1, \ldots, M\}\} \tag{2.18}$$

holds with high probability, where $\mathcal{S}_k = [\widetilde{\tau}_k - 3\delta/4, \widetilde{\tau}_k - \delta/2]$ and $\mathcal{E}_k = [\widetilde{\tau}_k + \delta/2, \widetilde{\tau}_k + 3\delta/4]$, $k = 1, \ldots, K$. This result is verified in Fryzlewicz et al. (2014); Wang et al. (2021b).

Algorithm 1 : Wild Binary Segmentation for high dimensional GGM.

INPUT: $\{\boldsymbol{X}_i\}_{i=1}^n, (s, e) = [1, n], \{a_m, b_m\}_{m=1}^M, \delta > 0, \lambda_n > 0, \gamma > 0$.

For $m = 1, \ldots, M$ **do**
  $[s_m, e_m] \leftarrow [s, e] \cap [a_m, b_m]$
  if $e_m - s_m \geq 2\delta$ **then**
    $b_m \leftarrow \arg\max_{t \in (s_m + \delta, e_m - \delta)} \mathcal{B}_t^{(s_m, e_m)}$
    $a_m \leftarrow \mathcal{B}_{b_m}^{(s_m, e_m)}$
  else
    $a_m \leftarrow -1$
end for
$m* \leftarrow \arg\max_{1 \leq m \leq M} a_m$
if $a_{m*} \geq \gamma$ **then**
  Add b_{m*} to the set of estimated change point.
  Repeat WBS on $(s, b_{m*}]$ and on $(b_{m*}, e]$

OUTPUT: The initial change change-point estimator $\check{\boldsymbol{\tau}}_{\check{k}} = (0, \check{\tau}_1, ..., \check{\tau}_{\check{k}}, 1)^\top$ and the estimator of grahical structures $\{\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}})\}_{a=1}^p$.

Algorithm 1 desribes our procedure for obtaining initial change point estimators $\check{\boldsymbol{\tau}}_{\check{k}} = (0, \check{\tau}_1, ..., \check{\tau}_{\check{k}}, 1)^\top$. Besides, based on the estimated change points, for each node a, we can obtain the WBS based initial estimators of the underlying coefficient vector $\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}}) := (\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}}, 1)^\top, \ldots, \check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}}, \widehat{k} + 1)^\top)^\top \in \mathbb{R}^{p(\check{k}+1)}$ by

$$\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}}, j) = \underset{\boldsymbol{\theta}: \theta_a = 0}{\arg\min} \left(L_n^a\left(I_j(\check{\boldsymbol{\tau}}_{\check{k}}), \boldsymbol{\theta}\right) + \lambda_n \sqrt{\check{\tau}_j - \check{\tau}_{j-1}} \|\boldsymbol{\theta}\|_1 \right), \text{ for } j = 1, \ldots, \check{k} + 1. \tag{2.19}$$

For the second step of RWBS, we put $\check{\boldsymbol{\tau}}_{\check{k}}$ and $\{\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}})\}_{a=1}^p$ obtained in (2.19) into $J_n(\boldsymbol{\tau}_k, \{\boldsymbol{\theta}^a(\boldsymbol{\tau}_k)\}_{a=1}^p)$ defined in (2.5) and obtain the improved change point estimator by $\check{\boldsymbol{\tau}}_{\check{k}}^r = (0, \check{\tau}_1^r, ..., \check{\tau}_{\check{k}}^r, 1)^\top$. We can prove in Theorem 3.4 that $\check{\boldsymbol{\tau}}_{\check{k}}^r$ has the best convergence rates as that of RDPA. Lastly, using the updated change point estimation, we can further construct the confidence interval for each change point by (2.14). The detailded implementation procedure is summarized in Algorithm 2.

Algorithm 2 : Refined WBS algorithm for high dimensional GGM (RWBS).

Input: Given the data set $\{\mathbf{X}\}$, set the value of $k_{\max} = 10$.

Step 1: Obtain the initial change point estimators $\check{\boldsymbol{\tau}}_{\check{k}}$ and graphical structures $\{\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}})\}_{a=1}^p$ by WBS in Algorithm 1.

Refinement

Step 2 :

(a) Given the change point number $\check{k}$ and the underlying coefficients $\{\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}_{\check{k}})\}_{a=1}^p$, obtain the improved change-point estimator $\check{\boldsymbol{\tau}}^r(\check{k})$ by a simplified dynamic programming procedure in (2.9).

(b) Obtain the updated estimator of the coefficients $\{\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}^r(\check{k}))\}_{a=1}^p$ by (2.10).

(c) Construct confidence interval:

Output :

(1)The change point estimator $\check{\boldsymbol{\tau}}^r(\check{k}) = (0, \check{\tau}_1^r, ..., \check{\tau}_{\check{k}}^r, 1)^\top$

(2) the updated estimator of the grahical structures $\{\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}^r(\check{k}))\}_{a=1}^p$

(3) and the confidence intervals for each change point $[\check{l}_n(k), \check{u}_n(k)]$ with $k = 1, \ldots, \check{k}$.

3 Theoretical Properties

In this section, we examine the theoretical properties of our proposed approaches. In Section 3.1, we introduce some assumptions. In Section 3.2, we present the theoretical results of the change point estimator computed by our proposed algorithms. In Sections 3.3, we derive the theoretical results of multiple change point inference.

3.1 Basic Assumptions

The fundamental assumptions are provided in Section C of the supplementary materials for comprehensive reference. To facilitate a clearer understanding of our theoretical results, we highlight one key assumption regarding the parameter space here. Let $\kappa_{n,p}^2 = \min_{1\le j\le \widetilde{k}} \frac{1}{p}\sum_{a=1}^p \|\widetilde{\boldsymbol{\theta}}^a(j) - \widetilde{\boldsymbol{\theta}}^a(j+1)\|_2^2$, and $\delta = \min_{1\le j\le \widetilde{k}}(\widetilde{\tau}_{j+1} - \widetilde{\tau}_j)$ be the minimal segment length, for some $\xi > 0$ and positive constant C_{SNR}, we require $\delta\kappa_{n,p}^2 \ge C_{SNR}\frac{s_* \log^{1+\xi}(p\vee n)}{n}$, where $\widetilde{\boldsymbol{\theta}}^a(j)$ is the true regression coefficient of the j-th interva of Node a, $\widetilde{k}$ is the true

change point number. The tuning parameters λ_n is required to satisfy

$$\lambda_n \propto \sqrt{\frac{\log(p)}{n}}. \tag{3.1}$$

3.2 Main Result I: Estimation of Change Points

3.2.1 Results of RDPA

As a benchmark, we first present main results of the estimators computed by RDPA.

Theorem 3.1. (**Consistency of initial estimators**) Suppose Assumptions A-D hold and additionally assume $\lambda_n s_*/\sqrt{\delta} = o(1)$ and $s_*^2 \log p = o(\delta n)$ hold. Then we have

(1) $\mathbb{P}\{\widehat{k} = \widetilde{k}\} \to 1$, as $n, p \to \infty$;

(2) $\max_{1\le j\le \widehat{k}} |\widehat{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{\lambda_n^2 s_*}{(\kappa^*_{n,p})^2})$;

(3) $\|\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}, j) - \widetilde{\boldsymbol{\theta}}^a(j)\|_1 = O_P(\lambda_n s_*/\sqrt{\delta})$, for $a = 1, ..., p, j = 1, ..., \widehat{k}+1$.

The first and second results of Theorem 3.1 establish the consistency of the change points detected by Algorithm 1, both in terms of their total number and locations. Specifically, Result (1) asserts that Algorithm 1 can correctly identify the number of change points with high probability. It is important to note that in (3.1), we require the condition $\lambda_n \propto \sqrt{\frac{\log p}{n}}$. Consequently, Results (2) and (3) can be rewritten as, for $j = 1, ..., \widehat{k}$ and $a = 1, \ldots, p$,

$$\max_{1\le j\le \widehat{k}} |\widehat{\tau}_j - \widetilde{\tau}_j| = O_P\left(\frac{s_* \log p}{\kappa_{n,p}^2 n}\right), \ \|\widehat{\boldsymbol{\theta}}^a(\widehat{\boldsymbol{\tau}}, j) - \widetilde{\boldsymbol{\theta}}^a(j)\|_1 = O_P\left(s_*\sqrt{\frac{\log p}{\delta n}}\right). \tag{3.2}$$

From (3.2), it is evident that consistent estimators for the change points can be obtained as long as the condition $s_*^2 \log p = o(\delta n)$ holds. Moreover, Result (3) suggests that the convergence rate of the regression coefficient estimator as shown in (3.2), is nearly minimax optimal, compared to the rate of regular Lasso estimation (Raskutti et al., 2011). It is crucial to highlight that change points are estimated simultaneously, which leads to a suboptimal convergence rate for the initial change point estimator. In the following Theorem 3.1, we introduce the refining procedure in Step 2, which improves the convergence rate for the estimator of change points, eliminating the $s_* \log p$ term.

Theorem 3.2. (**Consistency of refined change point estimator**) Suppose Assumptions A-D hold and assume $\lambda_n s_*/\sqrt{\delta} = o(1)$, $s_*^2 \log p = o(\delta n)$, hold. Then we have

(1) $\mathbb{P}\{\widehat{k}^r = \widetilde{k}\} \to 1$, as $n, p \to \infty$;

(2) $\max_{1\le j\le \widehat{k}} |\ddot{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{1}{(\kappa^*_{n,p})^2 n})$;

(3) $\max_{1\le j\le \widehat{k}} |\widehat{\tau}^r_j - \widetilde{\tau}_j| = O_P(\frac{1}{(\kappa^*_{n,p})^2 n})$.

Theorem 3.2 establishes the improved convergence rate of the refined change point estimators. Specifically, Result (1) is consistent with the first result of Theorem 3.1, since the number of change points is determined in the initial step. Result (2) describes the convergence rate of the oracle estimator $\ddot{\boldsymbol{\tau}}$, defined in (2.6). In Result (3), we show that the refined change point estimator $\widehat{\boldsymbol{\tau}}^r$ achieves the same convergence rate as the oracle estimator $\ddot{\boldsymbol{\tau}}$. This suggests that our refinement procedure in Step 2, which improves variable selection and parameter estimation, significantly enhances the accuracy of change point detection and leads to an oracle-like performance. To the best of our knowledge, this result is new for high-dimensional GGMs. By combining Theorems 3.1 and 3.2, we see a clear improvement in the convergence rate of the refined estimator. These improvements are summarized in Table 1, highlighting the benefits of the refitting procedure.

Table 1: Convergence rates for our proposed estimators by RDPA and RWBS.

Signal		**RDPA**		**RWBS**	
$\kappa_{n,p}$	δ	$\|\widehat{\boldsymbol{\tau}} - \widetilde{\boldsymbol{\tau}}\|_\infty$	$\|\widehat{\boldsymbol{\tau}}^r - \widetilde{\boldsymbol{\tau}}\|_\infty$	$\|\check{\boldsymbol{\tau}} - \widetilde{\boldsymbol{\tau}}\|_\infty$	$\|\check{\boldsymbol{\tau}}^r - \widetilde{\boldsymbol{\tau}}\|_\infty$
$O(1)$	sparse $O(1)$	$s_* \frac{\log p}{n}$	$\frac{1}{n}$	$s_* \frac{\log p}{n}$	$\frac{1}{n}$
$(\frac{s_*^2 \log p}{n})^{\frac{1}{4}}$	sparse $O(1)$	$\sqrt{\frac{\log(p)}{n}}$	$\frac{1}{s_*\sqrt{n\log(p)}}$	$\sqrt{\frac{\log(p)}{n}}$	$\frac{1}{s_*\sqrt{n\log(p)}}$
$O(1)$	dense $o(1)$	$s_* \frac{\log p}{n}$	$\frac{1}{n}$	$s_* \frac{\log p}{n}$	$\frac{1}{n}$
$(\frac{s_*^2 \log p}{n})^{\frac{1}{4}}$	dense $o(1)$	$\sqrt{\frac{\log(p)}{n}}$	$\frac{1}{s_*\sqrt{n\log(p)}}$	$\sqrt{\frac{\log(p)}{n}}$	$\frac{1}{s_*\sqrt{n\log(p)}}$

Remark 3.3. Gibberd and Roy (2017), **?**, and Kaul et al. (2020) propose distinct approaches to change point estimation, differing in both methodology and theoretical guarantees. Gibberd and Roy (2017) employs a group-fused graphical lasso to estimate the

precision matrix at each time point, using it to detect both the number and locations of change points. However, under stronger signal-to-noise ratio (SNR) conditions, the convergence rate of the change point estimator in Gibberd and Roy (2017) does not achieve optimality. Moreover, its algorithm is computationally expensive, making it impractical for large-scale data applications, as demonstrated by the numerical results in Section 4. In contrast, **?** focuses on multiple change point estimation in graphical models and introduces a computationally efficient algorithm. However, while their method is fast, it lacks theoretical guarantees on estimation accuracy and convergence, limiting its applicability in settings that require rigorous statistical assurances. On the other hand, Kaul et al. (2020) specifically addresses inference for a single change point, assuming its existence a priori. Consequently, their method does not estimate the number of change points, making it fundamentally different from our approach. Notably, our work generalizes this setting by allowing for multiple change points and providing a unified framework for both estimation and inference, thereby extending the theoretical insights beyond the single change point case.

3.2.2 Results of RWBS

Next, we introduce the theoretical results of RWBS.

Theorem 3.4. (**Consistency of initial estimators**) Suppose Assumptions A-D hold and additionally assume $\lambda_n s_*/\sqrt{\delta} = o(1)$ and $s_*^2 \log p = o(\delta n)$ hold. Then we have

(1) $\mathbb{P}\{\check{k} = \widetilde{k}\} \to 1$, as $n, p \to \infty$;

(2) $\max_{1 \le j \le \widetilde{k}} |\check{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{\lambda_n^2 s_*}{(\kappa_{n,p}^*)^2})$;

(3) $\|\check{\boldsymbol{\theta}}^a(\check{\boldsymbol{\tau}}, j) - \widetilde{\boldsymbol{\theta}}^a(j)\|_1 = O_P(\lambda_n s_*/\sqrt{\delta})$, for $a = 1, ..., p, j = 1, ..., \check{k} + 1$.

Theorem 3.4 establishes the consistency and convergence rates of the estimators for the number and locations of change points, as well as the underlying coefficient structures. Notably, both RWBS and RDPA achieve the same consistency and convergence rate re-

sults for these initial estimators. Additionally, the signal-to-noise ratio (SNR) conditions required for RWBS align with those for RDPA, particularly regarding the assumptions on the minimal signal jump size $\kappa_{n,p}$ and the minimal segment length δ (Assumption C). In the following, we present the theoretical guarantees for the refined estimators, $\check{\boldsymbol{\tau}}^r$, computed using Algorithm 2.

Theorem 3.5. (**Consistency of refined change point estimator**) Suppose Assumptions A-D hold and additionally assume $\lambda_n s_*/\sqrt{\delta} = o(1)$, $s_*^2 \log p = o(\delta n)$, hold. Then we have

(1) $\max_{1\le j\le \widetilde{k}} |\check{\tau}_j^r - \widetilde{\tau}_j| = O_P(\frac{1}{(\kappa_{n,p}^*)^2 n})$;

(2) $\max_{1\le j\le \widetilde{k}} |\ddot{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{1}{(\kappa_{n,p}^*)^2 n})$.

Theorem 3.4 establishes results analogous to those in Theorem 3.2, demonstrating the consistency and convergence rates of the refined estimators $\check{\boldsymbol{\tau}}^r$ computed by RWBS under the same assumptions. This indicates that the refinement step in RWBS is as effective as the second-step estimation in RDPA in improving accuracy. In particular, the second-step procedure substantially enhances the precision and efficiency of the estimators, highlighting the critical role of this refinement step in both methods.

Remark 3.6. Both RWBS and RDPA achieve optimal theoretical performance under the same signal-to-noise ratio (SNR) conditions, ensuring reliable change point detection. RDPA, though a benchmark method with strong theoretical guarantees, has high computational complexity, scaling as $O(n^2\text{Graph}(n,p))$, which makes it impractical for high-dimensional data. In contrast, RWBS reduces the computational cost to $O(Mn\log(n)\text{Graph}(n,p))$, where M is the number of randomly sampled intervals. Numerical results in Section 4 show that a small M can still provide accurate change point estimation, making RWBS a practical alternative to RDPA, offering a good balance between efficiency and accuracy for large-scale problems.

Remark 3.7. There are three main novel contributions in the proof of our method.

(a) Multiple Change Point Estimation: This article addresses the problem of multiple change point estimation within a complex framework, using a gain function based on the loss function. Unlike traditional methods like the CUSUM-based WBS algorithm, which are based on mean models, our approach introduces a new proof structure that incorporates more detailed loss functions, leading to a more challenging theoretical analysis. We aim for consistent change point estimation, addressing potential over- or underestimation. By controlling the gain function's bounds across various mixed data scenarios(See Lemmas F.6 and F.7), we handle the complex empirical process and excess risk(See Lemma F.1), requiring advanced techniques that differ from standard mean-based methods.

(b) Simultaneous Refinement: we incorporate a refinement step in both RDPA and RWBS procedures. However, our refinement approach differs in that it is based on the nodewise Lasso estimate and loss function. The refinement procedure involves not just a single change point or independent changes but simultaneously refining all change point locations to achieve an overall optimal solution. This simultaneous refinement is crucial for obtaining more precise estimates, and it requires controlling the theoretical properties of the gain function in various conditions.

(c) Dynamic Programming for Global Optimization: our method integrates dynamic programming into the gain function framework, enabling the simultaneous refinement of all change points and ensuring a globally optimal solution. This approach This approach goes beyond local refinement by using a global scanning and updating process to enhance accuracy. In the theoretical analysis (see supplementary Section ??), we focus on controlling two critical components: the excess risk and the empirical process associated with the gain function. To achieve this, we introduce a criterion that aligns with M-estimation principles, ensuring that our refinement procedure effectively improves estimation accuracy.

3.3 Main Result II: Inference of Change Points

After giving the theoretical results for the change point estimation. We next discucss the theoretical properties for change point inference. Recall the $\widetilde{k}$ change points $\eta_0 := 0 < \widetilde{\eta}_1 < \cdots < \widetilde{\eta}_{\widetilde{k}} < n := \eta_{\widetilde{k}+1}$. Recall $\widetilde{\boldsymbol{\delta}}^a(j) := \widetilde{\boldsymbol{\theta}}^a(j+1) - \widetilde{\boldsymbol{\theta}}^a(j)$ be the singal jump for node a at the j-th change point and recall $\widetilde{\Delta}(j) = \sqrt{\sum_{a=1}^p \|\widetilde{\boldsymbol{\delta}}^a(j)\|^2}$ and $\widetilde{\psi}(j) = \widetilde{\Delta}(j)/\sqrt{p}$. For the j-th segmentation with $j = 1, \ldots, \widetilde{k}+1$, let $\boldsymbol{\Sigma}(j) := \mathrm{Cov}(\boldsymbol{X}^{\eta_j})$ the the covariance matrix for the corresponding interval. For $t = 1, \ldots, n$ and $a = 1, \ldots, p$, define the regression residuals as

$$\epsilon_a^t = X_a^t - \sum_{j=1}^{\widetilde{k}+1} (\boldsymbol{X}_i^{-a})^\top \widetilde{\boldsymbol{\theta}}^a(j) \mathbf{1}\{\widetilde{\eta}_{j-1} < t \leq \widetilde{\eta}_j\}. \tag{3.3}$$

With the above notations, we need Assumption E, which is introduced in Supplementary Section C. Next, we give the change point inference for RDPA.

Theorem 3.8. Suppose Assumptions A-D hold. Additionally, Assumption E holds. Let $\widehat{\eta}_j^r$ be the refined change point estimator of the RDPA. Then, for each $j = 1, \ldots, \widetilde{k}$, under the vanishing regime with $\widetilde{\psi}(j) \to 0$, we obtain,

$$\widetilde{\psi}^2(j)\,(\widehat{\eta}_j^r - \widetilde{\eta}_j) \Rightarrow^d \arg\max_{r \in \mathbb{R}} Z_j(r), \tag{3.4}$$

where the process $\{Z_j(r), r \in \mathbb{R}\}$ are defined in (2.12). Moreover, we can prove that $\widetilde{\psi}^2(j)\,(\widehat{\eta}_j^r - \widetilde{\eta}_j)$ with $j = 1, \ldots, \widetilde{k}$, are asymptotically independent.

Let $q_{\alpha/2}$ and $q_{1-\alpha/2}$ be the quantiles of the $\arg\max_{r \in \mathbb{R}} Z(r)$. Using Theorem 3.8, we can construct the theoretical confidence interval as

$$\Big[\widetilde{l}_\alpha(j), \widetilde{u}_\alpha(j)\Big] := \Big[\widehat{\eta}_j^r - \frac{q_{1-\alpha/2}}{\widetilde{\psi}^2(j)}, \widehat{\eta}_j^r - \frac{q_{\alpha/2}}{\widetilde{\psi}^2(j)}\Big], \text{ for } j = 1, \ldots, \widetilde{k}.$$

Moreover, Theorem 3.8 allows us to make simultaneous confidence intervals of the detected change points. Specifically, for any $\widetilde{\eta}_{k_1}, \ldots, \widetilde{\eta}_{k_M} \subset \{\widetilde{\eta}_1, \cdots, \widetilde{\eta}_{\widetilde{k}}\}$, let $[\widetilde{l}_{\alpha/M}(k_1), \widetilde{u}_{\alpha/M}(k_1)]$,..., $[\widetilde{l}_{\alpha/M}(k_M), \widetilde{u}_{\alpha/M}(k_M)]$ be the corresponding confidence intervals. Then, as $(n, p) \to \infty$, we have

$$\mathbb{P}\Big(\widetilde{\eta}_{k_1} \in [\widetilde{l}_{\alpha/M}(k_1), \widetilde{u}_\alpha/M(k_1)], \ldots, \widetilde{\eta}_{k_M} \in [\widetilde{l}_{\alpha/M}(k_M), \widetilde{u}_\alpha/M(k_M)]\Big) \to 1 - \alpha.$$

From Theorem 3.8 and the process $Z_j(r)$ in (2.12) , we observe that Assumption E primarily requires that $\big(\sigma_1(j), \sigma_2(j), \sigma_1^*(j), \sigma_2^*(j)$, take moderate values. Specifically, it indicates that the stochastic influence of the mixture of three Brownian motion processes should neither be insignificant nor overwhelming, thus striking a balance between the signal and the noise. In Supplementary Section B, we introduce how to construct the estimators for $[\widetilde{l}_\alpha(j), \widetilde{u}_\alpha(j)]$.

4 Empirical studies

In this section, we investigate the numerical performance of our three change point detection procedures in various cases.

4.1 Simulation study

In this section, extensive numerical performance is present for change points detection. We generate $\boldsymbol{X}_i \in \mathbb{R}^p$ from $\mathbb{N}(\mathbf{0}, \boldsymbol{\Omega}_i^{-1})$, $i = 1, ..., n$, where $\boldsymbol{\Omega}_i$ is the $p \times p$ precise matrix. For the computational efficiency, we consider **Model Setting i: Varying sample size** n **and data dimension** p, we vary $n \in \{200, 400, 600, 800, 1000\}$ and $p = n$, $k_0 = 1$. As for the statistical efficiency, we first consider three model settings as follows:

Model Setting 1: no change point with varying data dimension. In this setting, we fix $n = 200$, $k_0 = 0$ and $p \in \{50, 100, 200\}$.

Model Setting 2: three change points with varying signal jump size. In this setting, we fix $n = 600$, $p = 200$, $k_0 = 3$, the change-point locations are $(0.25, 0.5, 0.75)$. The precision matrix $\boldsymbol{\Omega}_i$ take the form: $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_1, 1 \leq i \leq n/4,$ or $n/2+1 \leq i \leq 3n/4, \boldsymbol{\Omega}_i = \boldsymbol{\Omega}_2, n/4+1 \leq i \leq n/2,$ or $3n/4+1 \leq i \leq n$. We set $\boldsymbol{\Omega}_1$ being the chain precision matrix with $\{20, 30, 40\}$ nodes changing compared to $\boldsymbol{\Omega}_2$.

Model Setting 3 (Inference): three change points. We consider Model Setting 2, where $\boldsymbol{\Omega}_1$ is the chain precision matrix with 30 nodes changing compared to $\boldsymbol{\Omega}_2$.

Since our main focus is on estimating multiple change points, we select the OBS method

proposed by Kovács et al. (2020) as the comparison method. Note that our proposed methods involve the selection of two parameters λ_n and γ. To save space, we provide a cross-validation approach to select tuning parameters in Supplementary Section I.

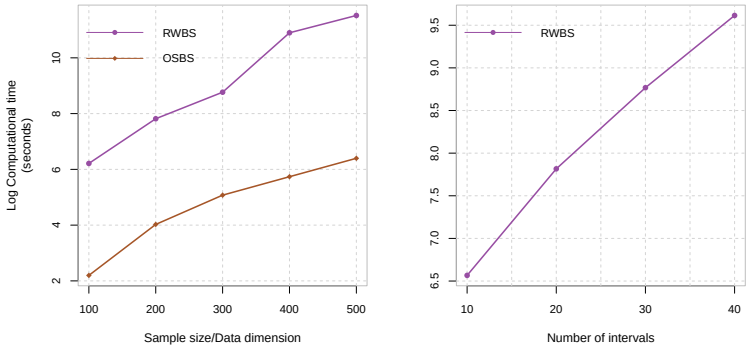

Figure 2: Computational efficiency of change point estimation under the Model setting i. The numerical results are based on one typical replication.

To demonstrate the computational efficiency of RWBS, the corresponding algorithms were implemented on a 2.50 GHz, 6-core, 4GB RAM CPU (Linux). When comparing computation times, RWBS strikes a balance between speed and statistical accuracy. As shown in Figure 2, RWBS outperforms OBS in statistical accuracy while being more efficient than the other methods. The computation time for RWBS increases with the number of intervals M, but selecting a relatively small value for M still yields satisfactory results. This shows that RWBS offers a practical and effective balance between computational efficiency and statistical accuracy when M is appropriately chosen.

Table 2: Change point detection for Model Setting 1. The numerical results are based on 100 replications.

	Methods	$p = 50$	$p = 100$	$p = 200$
Error rate	RWBS	0.00	0.00	0.00
	OBS	0.72	0.67	0.60

We present the performance of statistical accuracy in change point estimation. Table 2 compares the error rates of RWBS and OBS for detecting change points in Model Setting 1. OBS shows significantly higher error rates across varying sample sizes, while RWBS

Table 3: Multiple change point detection for Model Setting 2. The numerical results are based on 100 replications.

	Hausdorff distance (Mean|sd)		
Methods	20 nodes changes	30 nodes changes	40 nodes changes
OBS	0.402 | 0.126	0.393 | 0.118	0.389 | 0.116
RWBS-one	0.068 | 0.079	0.067 | 0.078	0.062 | 0.081
RWBS-refit	0.062 | 0.083	0.060 | 0.081	0.050 | 0.084

achieves a consistent error rate of 0.00, indicating its superior performance in detecting change points. RWBS also correctly estimates $k = 0$ when no change points are present, demonstrating its accuracy in change point testing.

Table 3 compares the Hausdorff distance for multiple change point detection across RWBS and OBS under different signal strengths. The Hausdorff distance(range from 0 to 1), measuring the similarity between true and estimated change point locations(with smaller values indicating higher estimation accuracy). It reveals that OBS consistently yields the largest distances (0.389–0.402), indicating poor accuracy. RWBS shows significant improvements, with RWBS-one ranging from 0.062 to 0.079 and RWBS-refit outperforming OBS, with Hausdorff distances close to 0.50 and small standard deviations. These results **emphasize the importance of the refinement procedure** in RWBS for optimal performance, especially when signal strength is weaker.

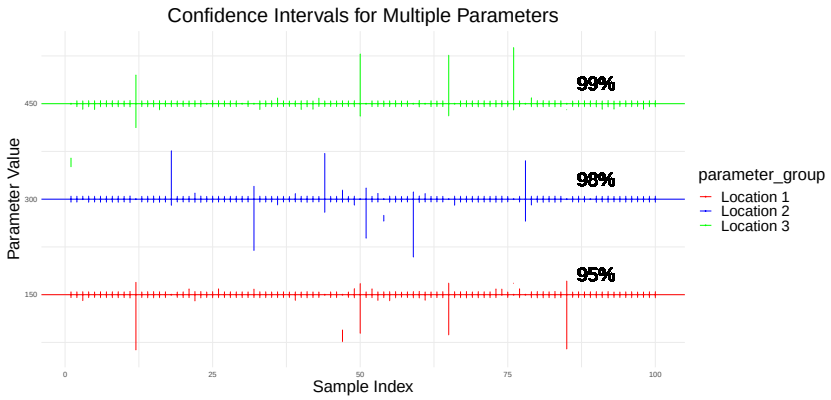

Figure 3: Multiple change point inference result under the Model setting 3. The numerical results are based on 100 replications.

As for the evaluation of change point inference, as shown in Figure 3, which illustrate the change point inference outcomes for RWBS under Model Setting 3, including

the confidence intervals for multiple change points. It is evident that the confidence intervals for each estimated change-point location consistently cover the true change points, with coverage probabilities ranging between 95% and 100%. This strongly suggests that the proposed inference method is both reliable and accurate, with the confidence intervals providing an effective estimate of the true change point locations. Furthermore, the narrowness of the intervals reinforces the precision of RWBS in detecting change points, highlighting the robustness of the method in this particular setting. Overall, these findings demonstrate the reliability and high performance of RWBS for change point detection with accurate uncertainty quantification.

5 Real data analysis for S&P 500 dataset

In this section, we analyze the S&P 500 dataset to detect structural change points in market dynamics over two critical periods: January 1, 2007, to December 31, 2011, and January 1, 2019, to December 31, 2022. The first period (2007–2011) includes the 2008 Global Financial Crisis, while the second period (2019–2022) captures the onset of the COVID-19 pandemic, both of which triggered significant market disruptions. The dataset, sourced from Yahoo! Finance (https://finance.yahoo.com/), consists of daily returns for 234 continuously listed stocks within the S&P 500 index. More specifically, for each day i, we denote $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip})^\top$, where $p = 234$ and X_{ij} represents the daily return for the jth stock. Detecting change points in financial markets is essential for identifying shifts between stable and volatile regimes, often influenced by major economic and geopolitical events. In this analysis, we apply the estimation and inference procedures detailed in Sections 2 and 3, using parameter values consistent with those in Section 4. Prior to change point detection, the data are standardized to have zero mean and unit variance to ensure comparability across different time periods.

Table 4 shows that the RWBS method effectively identifies multiple change points and

constructs confidence intervals for the S&P 500 data in two key periods. To explain these findings, we compare them with the TED spread, a measure of market risk derived from the difference between the 3-month LIBOR and the 3-month U.S. Treasury bill rate. Figure 4 shows the TED spread over time, with the orange dotted intervals marking the RWBS-estimated change point confidence intervals. In the 2007–2011 period, large fluctuations in the TED spread coincide with the estimated change points. For example, RWBS captures the "nadir of the crisis" in March 2009 and the sharp drop in stock prices in 2011, triggered by fears of the European sovereign debt crisis. These results align with findings from Cho and Fryzlewicz (2015), validating the effectiveness of our method.

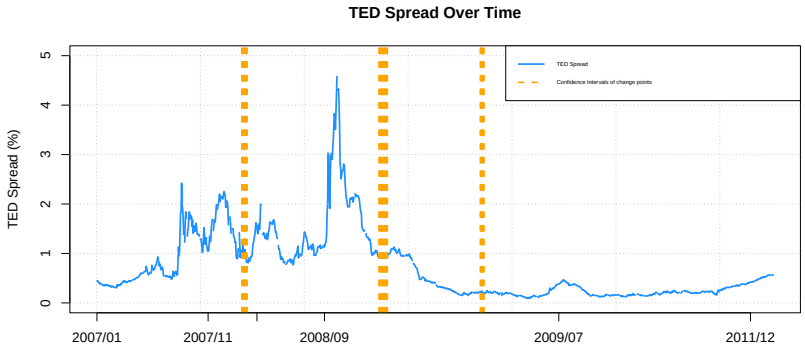

Figure 4: TED Spread during 2007 to 2011 with confidence intervals of change points (orange dotted intervals) marked by # in Table 4.

In the second period (2019–2024), our analysis identifies two important change points that correspond to key phases of the COVID-19 pandemic. The first change point, observed on January 16, 2020 (with confidence interval from 2020/01/09-2020/01/21), aligns with the early stage of the outbreak, when the virus began to spread in Europe and initial cases were reported. The second change point, estimated on June 11, 2020(with confidence interval from 2020/06/03-2020/06/23), marks the beginning of the recovery phase, when financial markets started to stabilize and investor confidence returned. Another estimated change point is December 14, 2022 (with confidence interval from 2022/12/09-2022/12/21),marking the transition into the post-COVID-19 phase. **These results demonstrate the effectiveness of our method in accurately identifying change points in financial market that reflect critical transitions in different phases of the COVID-19 pan-**

demic. Based on the estimated change points, we divided the 2019–2021 period into three intervals, each corresponding to different stages of the pandemic. The graph structures estimated for each phase, shown in Figure 1, reveal significant differences across the phases, further demonstrating the effectiveness of our method in capturing critical transitions in financial market structures.

Table 4: Change point estimation and inference results of RWBS for S&P 500 in two periods:2007/01/03-2011/12/31 and 2019/01/03-2024/12/31. Dates with TED Spread# correspond to the red dotted line in Figure 4.

Year	$\widehat{\tau}$	Date	CI	
2007-2011	285	2008/02/21	2008/02.13-2008/02/28	TED Spread#
	550	2009/03/11	2009/03/03-2009/03/25	"nadir of the crisis", TED Spread#
	741	2009/12/10	2009/12/08-2009/12/17	TED Spread#
	1020	2011/01/20	2010/12/15-2021/02/02	The European sovereign debt crisis escalated
	1167	2011/08/19	2011/08/19-2021/09/12	Global stock markets fell due to fears of contagion of the European sovereign debt crisis
2019-2024	263	2020/01/16	2020/01/09-2020/01/21	The early outbreak of COVID-19
	364	2020/06/11	2020/06/03-2020/6/23	Partial stabilization of the pandemic
	767	2022/01/14	2022/01/11-2022/01/26	
	994	2022/12/14	2022/12/09-2022/12/21	The post-COVID-19 phase
	1144	2023/07/19	2023/07/07-2023/07/25	

6 Summary

This paper presents an efficient framework for multiple change point estimation and inference in high-dimensional Gaussian Graphical Models (GGMs). Using Nodewise Lasso for graph structure estimation, we propose a two-step approach with two algorithms: the Refined Dynamic Programming Algorithm (RDPA) and the Refined Wild Binary Segmentation (RWBS). A key distinction of our work is the integration of dynamic programming into the gain function framework, allowing simultaneous refinement of all change points for an optimal global solution. We establish theoretical guarantees, including consistency in estimating the number of change points, oracle-level location accuracy, and optimal convergence rates for underlying coefficients. Compared to existing methods, our approach extends beyond single change point estimation, efficiently handling multiple change points without prior knowledge of their number.While RDPA ensures strong theoretical performance, it is computationally demanding in high dimensions. RWBS, on the other hand, provides a scalable alternative with comparable accuracy. Extensive numerical studies and

an application to the S&P 500 dataset confirm the statistical reliability and computational efficiency of our methods. Our work advances high-dimensional change point detection by offering a theoretically sound and computationally efficient solution, making it applicable to various fields, including finance, bioinformatics, and network analysis.

References

Aue, A., Gabrys, R., Horváth, L. and Kokoszka, P. (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis* **100** 2254–2269.

Auger, I. E. and Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bulletin of mathematical biology* **51** 39–54.

Avanesov, V. and Buzun, N. (2018). Change-point detection in high-dimensional covariance structure. *Electronic Journal of Statistics* **12** 3254–3294.

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics* **157** 78–92.

Banerjee, O., Ghaoui, L. E. and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research* **9** 485–516.

Bybee, L. and Atchadé, Y. (2018). Change-point computation for large graphical models: A scalable algorithm for Gaussian graphical models with change-points. *Journal of Machine Learning Research* **19** 1–38.

Cai, T., Liu, W. and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106** 594–607.

Cho, H. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics* **10** 2000–2038.

Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77** 475–507.

Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

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

Gibberd, A. J. and Nelson, J. D. (2017). Regularized estimation of piecewise constant Gaussian graphical models: The group-fused graphical lasso. *Journal of Computational and Graphical Statistics* **26** 623–634.

Gibberd, A. J. and Roy, S. (2017). Multiple changepoint estimation in high-dimensional Gaussian graphical models. *Preprint arXiv:1712.05786* .

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association* **105** 1480–1493.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics* **43** 2451–2483.

Kaul, A., Fotopoulos, S. B., Jandhyala, V. K. and Safikhani, A. (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics* **15** 71–134.

Kaul, A. and Michailidis, G. (2025). Inference for Change Points in High Dimensional Mean Shift Models. *Statistica Sinica* **35** 1–25.

Kaul, A., Zhang, H., Tsampourakis, K. and Michailidis, G. (2020). Inference on the change point for high dimensional dynamic graphical models. *arXiv preprint arXiv:2005.09711* .

Kaul, A., Zhang, H., Tsampourakis, K. and Michailidis, G. (2023). Inference on the change point under a high dimensional covariance shift. *Journal of Machine Learning Research* **24** 1–68.

Kolar, M. and Xing, E. P. (2012). Estimating networks with jumps. *Electronic Journal of Statistics* **6** 2069–2106.

Kovács, S., Bühlmann, P., Li, H. and Munk, A. (2023). Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika* **110** 249–256.

Kovács, S., Li, H., Haubner, L., Munk, A. and Bühlmann, P. (2020). Optimistic search strategy: Change point detection for large-scale data via adaptive logarithmic queries. *arXiv preprint arXiv:2010.10194* .

Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press, New York.

Lee, S., Liao, Y., Seo, M. H. and Shin, Y. (2018). Oracle estimation of a change point in high-dimensional quantile regression. *Journal of the American Statistical Association* **113** 1184–1194.

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

Leonardi, F. and Bühlmann, P. (2016). Computationally efficient change point detection for high-dimensional regression. *Preprint arXiv:1601.03704* .

Liu, B., Zhang, X. and Liu, Y. (2021). Simultaneous change point inference and structure recovery for high dimensional gaussian graphical models. *Journal of Machine Learning Research* **22** 1–62.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics* **41** 2948–2978.

Londschien, M., Kovács, S. and Bühlmann, P. (2021). Change-point detection for graphical models in the presence of missing values. *Journal of Computational and Graphical Statistics* **30** 768–779.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* **34** 1436–1462.

Raskutti, G., Wainwright, M. J. and Yu, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls. *IEEE Transactions on Information Theory* **57** 6976–6994.

Wang, D., Yu, Y. and Rinaldo, A. (2021a). Optimal covariance change point localization in high dimensions. *Bernoulli* **27** 554–575.

Wang, D., Zhao, Z., Lin, K. Z. and Willett, R. (2021b). Statistically and computationally efficient change point localization in regression settings. *Journal of Machine Learning Research* **22** 1–46.

Wang, G. and Feng, L. (2023). Computationally efficient and data-adaptive changepoint inference in high dimension. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **85** 936–958.

Wang, R. and Shao, X. (2023). Dating the break in high-dimensional data. *Bernoulli* **29** 2879–2901.

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

Xu, H., Wang, D., Zhao, Z. and Yu, Y. (2023). Change point inference in high-dimensional regression models under temporal dependence. *The Annals of Statistics, to appear* .

Yao, Y.-C. and Au, S.-T. (1989). Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics, Series A* 370–381.