Sequential Analysis
Design Methods and Applications

ISSN: 0747-4946 (Print) 1532-4176 (Online) Journal homepage: www.tandfonline.com/journals/lsqa20

Change point detection in SCAD-penalized dynamic panel models

Chao Gu & Suthakaran Ratnasingam

To cite this article: Chao Gu & Suthakaran Ratnasingam (17 Jun 2025): Change point detection in SCAD-penalized dynamic panel models, Sequential Analysis, DOI: 10.1080/07474946.2025.2510372

To link to this article: https://doi.org/10.1080/07474946.2025.2510372

View supplementary material

Published online: 17 Jun 2025.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lsqa20

SEQUENTIAL ANALYSIS
https://doi.org/10.1080/07474946.2025.2510372

Taylor & Francis
Taylor & Francis Group

Check for updates

Change point detection in SCAD-penalized dynamic panel models

Chao Gu[a*] and Suthakaran Ratnasingam[b*]

[a]Department of Mathematical Sciences, The Citadel - The Military College of South Carolina, Charleston, South Carolina, USA; [b]Department of Mathematics, California State University, San Bernardino, San Bernardino, California, USA

ABSTRACT
We propose a cumulative sum (CUSUM)-based testing procedure to sequentially monitor structural changes in smoothly clipped absolute deviation (SCAD)-penalized dynamic panel models. Initially, this approach uses historical panel data to simultaneously perform variable selection and estimation with the SCAD penalty function. Tests based on CUSUM statistics are conducted to identify any change points in subsequent monitoring data. The consistency of the method and the oracle property of the resulting regularized estimators are examined. The asymptotic properties of the test statistics are established under both the null and alternative hypotheses. Simulations are conducted to demonstrate the performance of the proposed method. Finally, a real data application is provided to illustrate the detection procedure.

ARTICLE HISTORY
Received 10 November 2024
Revised 12 May 2025
Accepted 17 May 2025

KEYWORDS
Change point analysis; online change point; panel data; penalized regression

1. INTRODUCTION

Panel data analysis is a fundamental tool in empirical research, enabling the examination of dynamic relationships between variables over time across multiple entities. It is extensively utilized in fields such as economics, political science, and biostatistics. By integrating cross-sectional and time-series dimensions, panel data models offer more efficient estimators of causal effects compared to purely cross-sectional or time-series analyses (Baltagi 2005; Wooldridge 2010). However, challenges such as time-invariant unobservable heterogeneity and endogenous regressors necessitate advanced estimation techniques (Verbeek 2008; Arellano and Bond 1991).

Dynamic regression models, including the autoregressive distributed lag (ADL) model, are particularly effective in capturing time-dependent structures within panel data. These models incorporate lagged variables to account for temporal dependencies while controlling for unobserved heterogeneity. Nonetheless, conventional estimation methods like the generalized method of moments (GMM) can become computationally complex and susceptible to overfitting, especially in high-dimensional contexts (Hansen

CONTACT Suthakaran Ratnasingam suthakaran.ratnasingam@csusb.edu Department of Mathematics, California State University, San Bernardino, San Bernardino, CA 92407, USA.
*These authors contributed equally to this work as co-first authors.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07474946.2025.2510372.
© 2025 Taylor & Francis Group, LLC

1982; Hsiao 2014). To address these challenges, penalized regression techniques, such as the least absolute shrinkage and selection operator (LASSO) and the smoothly clipped absolute deviation (SCAD) estimator can be employed to enhance variable selection and estimation efficiency (Tibshirani 1996; Fan and Li 2001).

In this context, detecting structural changes in dynamic panel data models is a critical yet challenging task. Such structural shifts can significantly affect model performance and inference, emphasizing the importance of developing efficient detection methodologies. Change point detection has been extensively studied in non-panel settings. Horváth et al. (2004) introduced a sequential monitoring approach to detect structural changes in a stationary population using weighted cumulative sums (CUSUMs) of residuals, providing a computationally efficient method to sequentially monitor changes in model parameters. The CUSUM method offers a robust alternative to likelihood-based approaches, particularly in high-dimensional settings where recalculating likelihood ratios at each step becomes computationally prohibitive (Horváth and Hušková 2012; Ratnasingam and Ning 2021).

The adaptation to panel data requires specialized approaches. Recent advances in this field include Horváth et al. (2022), who developed methods for detecting common breaks in high-dimensional cross-dependent panels, particularly relevant in financial and macroeconomic applications where multiple time-series evolve under shared structural changes. In Bai (2009), the authors introduced interactive fixed effects models to address unobserved heterogeneity in panel data, thus enabling more flexible modeling of dynamic relationships. Furthermore, Jirák (2015) proposed uniform change point tests in high-dimensional data, offering a framework for consistent detection under minimal assumptions. More recently, Horváth and Rice (2024) provided a comprehensive framework for time-series change point analysis, extending traditional detection techniques to handle complex dependencies across both time and cross-sectional units. These methodologies highlight the increasing need for scalable and interpretable change point detection techniques within panel structures. Based on these advancements, we focus on the ADL model under a simplified setting with minimal autocorrelation and heteroscedasticity. Despite this assumption, identifying change points remains challenging. We employ SCAD estimation for model training and adapt the CUSUM testing procedure for efficient structural change detection.

The remainder of this article is organized as follows: Section 2 introduces key notations, model assumptions, and the application of SCAD for variable selection in panel data. Section 3 presents the CUSUM-based test statistics and asymptotic results. Section 4 provides simulation studies, followed by a real data application in Section 5. Finally, Section 6 offers discussions and conclusions.

2. METHODOLOGY

The analysis of panel data is a significant challenge for researchers seeking to examine complex relationships among variables. By using information from both cross-sectional units and time periods, panel data models can help researchers better understand causal effects and dynamic processes. These models provide tools for understanding both time-invariant and time-varying relationships in the data.

2.1. Dynamic Regression Models

Suppose we possess a sample of panel data comprising observations on the same n entities, each with p baseline predictors, observed over two or more time periods T. The data are denoted by $(X_{i,j,t}, Y_{i,t})$, with $i = 1, ..., n$, $j = 1, ..., p$, and $t = 1, ..., T$. For simplicity, $(X_t, Y_t)_{n\times(p+1)}$ refer to the data matrices at each time t. $\mathbf{X_{t^-}} = \{X_{t-l}$: significant lagged terms of $X_t\} \subseteq \{X_1, ..., X_{t-1}\}$ and $\mathbf{Y_{t^-}} = \{Y_{t-s}$: significant lagged terms of $Y_t\} \subseteq \{Y_1, ..., Y_{t-1}\}$ at time t, respectively, with $1 \le l, s \le t-1$ and the cardinality $|\mathbf{X_{t^-}}| = L$ and $|\mathbf{Y_{t^-}}| = S$. We then consider a general ADL model with L lags of X_t and S lags of Y_t, denoted as ADL(L, S) at time t. That is

$$Y_t = \beta_{0,t} + X_t\beta_t + \mathbf{X_{t^-}}\boldsymbol{\delta}_\mathbf{L} + \mathbf{Y_{t^-}}\boldsymbol{\eta}_\mathbf{S} + U_i + V_t + \mathcal{E}_{i,t}, \tag{2.1}$$

where $i = 1, ..., n$, $t = 1, ..., T$, and

- $\beta_{0,t}$, $\beta_t = (\beta_{1,t}, ..., \beta_{p,t})^\top$, $\boldsymbol{\delta}_\mathbf{L} = (\delta_1, ..., \delta_L)^\top$ in which $\delta_L = (\delta_{1,L}, ..., \delta_{p,L})^\top$ and $\boldsymbol{\eta}_\mathbf{S} = (\eta_1, ..., \eta_S)^\top$ are true unknown coefficients,
- U_i is unobserved and the single i subscript indicates it is time-invariant and varies across entities,
- V_t is unobserved and the single t subscript implies it varies over time but not across entities,
- $\mathcal{E}_{i,t}$ is the error term which can vary over time and across entities.

In equation (2.1), U_i and V_t capture unobserved individual and time effects, respectively. To mitigate endogeneity concerns, traditional estimation approaches such as instrumental variables (IV) and the GMM have been widely employed (Stock and Watson 2003; Frees 2004; Wooldridge 2010). However, these methods introduce significant computational complexity and require carefully chosen instruments or moment conditions, which may not always be available. Additionally, GMM-based estimators often suffer from weak instrument bias and overfitting when the number of instruments grows relative to the sample size (Hsiao 2014; Keele and Kelly 2006).

Given these challenges, we simplify the model specification by assuming weak or no autocorrelation and heteroscedasticity, allowing us to adopt a penalized regression approach. Without loss of generality, we consider that the data are centered, excluding the intercept from the regression model. And for this study, our emphasis is on the streamlined ADL model as shown below in reference to Stock and Watson (2003).

$$Y_t = X_t\beta_t + \mathbf{Y_{t^-}}\boldsymbol{\eta}_\mathbf{S} + \mathcal{E}, \tag{2.2}$$

with the following assumptions:

A1. $E(\mathcal{E}|Y_{t-1}, ..., Y_{t-s}, X_t) = 0$;

A2. Y_t and Y_{t-s} become independent as s gets large, i.e.

$$\lim_{s\to\infty} \mathrm{Corr}(Y_t, Y_{t-s}) = 0;$$

A3. The random variables $(Y_t, X_{1,t}, ..., X_{p,t})$ have a stationary distribution at each time t;

A4. There is no perfect multicollinearity;

A5. $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_n)^\top$ does not vary over time and represents an $n-$vector of independent identically distributed random variables, with $E(\mathcal{E}_1) = 0$, $\mathrm{Var}(\mathcal{E}_1) = \sigma^2 < \infty$, and $E(\mathcal{E}_1^4) < \infty$;

A6. Weak or no fixed effects over time and across entities.

The first assumption refers to the concept of exogeneity between the regressors and the error term, indicating that no additional lags belong in the model and ensuring the absence of severe omitted variable bias. The second assumption reflects a weakening of linear dependence over time, but does not imply full statistical independence unless additional distributional assumptions (e.g., Gaussianity) are imposed and needed. We interpret this asymptotic uncorrelatedness as indicative of diminishing dependence rather than full independence. The third assumption specifies a stationary distribution, implying that the joint distribution of the variables, including lags, remains constant over time. These two assumptions ensure the stability of the model so that the process is not explosive. The fourth and fifth assumptions guarantee the reliability and stability of estimates. We omit the requirement for employing IV or GMM when no autocorrelation and heteroscedasticity are proposed in the error term. Furthermore, the last assumption can substantially alleviate the complexity of the ADL model, particularly when data availability is limited. Similarly, independent time series are not involved in this context because we consider the case that the past values of X_{ij} might be inaccessible, data-intensive, or relatively constant, such as gender, baseline health conditions, and genetic characteristics. Otherwise, their lags should be incorporated into the model if certain predictors exhibit evident autocorrelation.

2.2. Estimation

In general, equation (2.2) can be rewritten as a linear regression model at time t:

$$Y_t = W_t\theta_t + \mathcal{E}, \tag{2.3}$$

where $W_t = (X_t, \mathbf{Y_{t^-}})$ and $\theta_t = (\beta_t, \boldsymbol{\eta}_\mathbf{S})^\top$.

Let $\mathcal{A}_t = \{j, s : \beta_{j,t} \neq 0, \eta_S \neq 0\}$ with the cardinality $|\mathcal{A}_t| = K^* \leq K$ at time t, where $K^* = p^* + S$ indicating the total number of significant variables with nonzero coefficients, $K = p + t - 1$ forming the total number of variables, and recall $j = 1, \ldots, p$ referring to p predictors of $X_{j,t}$ and S are the true number of lags of Y_t. Under certain conditions, given $W_t^\top W_t$ is of full rank, which indicates the number of entities $n \geq K$, ordinary least squares (OLS) offer the best linear unbiased estimators (BLUE) but lacks variable selection. To address the need for selecting significant variables, we often employ the best-subset or stepwise selection procedure. However, OLS becomes ineffective when faced with both sparsity and high-dimensionality, where K^* is small and $K \gg n$, respectively.

The LASSO is introduced as a regularization procedure for simultaneous estimation and variable selection (Tibshirani 1996). It estimates equation (2.3) as

$$\hat{\theta}_t = \arg \min_{\theta_t} \| Y_t - W_t\theta_t \|_2^2 + \lambda_t \| \theta_t \|_1, \tag{2.4}$$

where $\| \cdot \|_q$ denotes the q-norm and λ_t is a nonnegative regularization (tunning) parameter. As λ_t increases, equation (2.4) continuously shrinks the corresponding $\hat{\theta}_t$ toward 0, with some estimated coefficients reaching exact 0 if λ_t is sufficiently large. This continuous shrinkage can enhance prediction accuracy by balancing bias and variance.

Fan and Li (2001) proposed *unbiasedness*, *sparsity*, and *continuity* as the three properties of a good penalty function and defined an *oracle* procedure. According to our case, it is defined as

- Identify the right model: $\hat{\mathcal{A}}_t = \{j, s : \hat{\beta}_{j,t} \neq 0, \hat{\eta}_S \neq 0\} = \mathcal{A}_t$.
- The optimal estimation rate: $\sqrt{n}(\hat{\theta}_t - \theta_t) \xrightarrow{d} N(\mathbf{0}, \Sigma)$, where Σ is the covariance matrix knowing the true model and $\hat{\theta}_t = (\hat{\beta}_t, \hat{\boldsymbol{\eta}}_{\mathbf{S}})^\top$.

Zhao and Yu (2006) showed that the LASSO does not maintain the above properties when an irrelevant predictor exhibits a high correlation with the predictors in the true model. It may fail to differentiate it from the true predictors, regardless of the value of λ_t. Additionally, Knight and Fu (2000) demonstrated that the LASSO is $n^{1/2}$-consistent under specific regularity conditions, failing to achieve simultaneous consistent variable selection and estimation. When incorporating lags of Y_t in the ADL model, we may encounter these concerns, particularly the former one. Thus, a SCAD penalty known as an *oracle* selection procedure (Fan and Li 2001) is applied to adapt (2.4), which is defined as

$$\hat{\theta}_t = \arg \min_{\theta_t} \| Y_t - W_t\theta_t \|_2^2 + \sum_{k=1}^{p+s} p_{\lambda_t}^{SCAD}(|\theta_{k,t}|), \tag{2.5}$$

where $|\cdot|$ denotes the absolute value, $p_{\lambda_t}^{SCAD}(|\cdot|)$ are the penalty functions depending on λ_t and are allowed not to be identical for all $k = 1, \ldots, K$, given by

$$p_{\lambda_t}^{SCAD}(|\theta_{k,t}|) = \begin{cases} \lambda_t|\theta_{k,t}| & \text{if } |\theta_{k,t}| \leq \lambda_t, \\ (2a\lambda_t|\theta_{k,t}| - |\theta_{k,t}|^2 - \lambda_t^2)/[2(a-1)] & \text{if } \lambda_t < |\theta_{k,t}| \leq a\lambda_t, \\ (a+1)\lambda_t^2/2 & \text{if } |\theta_{k,t}| > a\lambda_t, \end{cases} \tag{2.6}$$

where $a > 2$ is the additional tunning parameter to λ_t. Equation (2.6) forms a quadratic spline function with knots at λ_t and $a\lambda_t$. It is continuous, with the first derivative of

$$\lambda_t\left\{ I(|\theta_{k,t}| \leq \lambda_t) + \frac{(a\lambda_t - |\theta_{k,t}|)_+}{(a-1)\lambda_t} I(|\theta_{k,t}| > \lambda_t) \right\}, \tag{2.7}$$

where $(\cdot)_+ = \max\{0, \cdot\}$. Equation (2.7) implies the SCAD penalty is continuously differentiable on $\mathbb{R}$ except for being singular at the origin, with the derivatives being all zeroes outside the interval $[-a\lambda_t, a\lambda_t]$. As a result, it sets small coefficients to zero, and shrinks some coefficients toward zero while preserving large coefficients. This yields a sparse set of solutions and approximately unbiased coefficients for large coefficients. The solution to the SCAD penalty is expressed as

$$\hat{\theta}_{k,t}^{SCAD} = \begin{cases} \operatorname{sgn}(\hat{\theta}_{k,t}^{OLS})(|\hat{\theta}_{k,t}^{OLS}| - \lambda_t)_+ & \text{if } |\hat{\theta}_{k,t}^{OLS}| \le 2\lambda_t, \\ \left[(a-1)\hat{\theta}_{k,t}^{OLS} - \operatorname{sgn}(\hat{\theta}_{k,t}^{OLS}) a\lambda_t\right]/(a-2) & \text{if } 2\lambda_t < |\hat{\theta}_{k,t}^{OLS}| \le a\lambda_t, \\ \hat{\theta}_{k,t}^{OLS} & \text{if } |\hat{\theta}_{k,t}^{OLS}| > a\lambda_t, \end{cases} \tag{2.8}$$

where $\operatorname{sgn}(\cdot)$ represents the sign function and $\hat{\theta}_{k,t}^{OLS}$ is the ordinary least squares estimate. The optimal pair (λ_t, a) can be theoretically obtained through a two-dimensional grid search using criteria like cross-validation and Bayesian information criterion (BIC), which can be quite computationally expensive. Based on Bayesian statistical principles and simulation studies, Fan and Li (2001) recommended setting $a = 3.7$ as a good choice for diverse scenarios. They further showed that the choice of a does not significantly improve the performance of the whole process compared to λ_t. Moreover, Wang, Li, and Tsai (2007) demonstrated that the widely applied generalized cross-validation method often fails to appropriately select the tuning parameter, leading to significant overfitting in the model. They recommended a BIC-based tuning parameter selector, which consistently identifies the true model. In this study, we accordingly set a to be 3.7, with λ_t chosen by BIC.

2.3. High-Dimensional and Sparse Settings

In the context of the high-dimensional and sparse model, we contemplate a scenario where the majority of regression coefficients are precisely zero. This implies that only a handful of predictors exhibit non-zero regression coefficients. Recall that, we propose there are p^* significant predictors and S significant lags of Y_t in $\mathcal{A}_t$. Let $W_t = (W_t^{(1)}, W_t^{(2)})$, where $W_t^{(1)}$ is the submatrix of W_t containing those significant predictors and lags and $W_t^{(2)}$ includes those trivial predictors and lags that are intended to be excluded from the model. Analogously, we denote $\theta_t = (\theta_t^{(1)}, \theta_t^{(2)})$ and $M_t^{(u,v)} = \frac{1}{n} W_t^{(u)\top} W_t^{(v)}$, where $M_t = \frac{1}{n} W_t^\top W_t$ and $u, v = 1, 2$. In terms of equation (2.8), we have $\hat{\theta}_t^{SCAD} = (\hat{\theta}_{1,t}^{SCAD}, \ldots, \hat{\theta}_{p+t-1,t}^{SCAD})^\top$. Suppose $\mathcal{A}_t^* = \{j, s : \hat{\beta}_{j,t}^{SCAD} \neq 0, \hat{\eta}_S^{SCAD} \neq 0\}$ is the index set of the significant SCAD estimators based on the historical sample at time t with the cardinality $|\mathcal{A}_t^*| = K^{*SCAD}$. We organize $\hat{\theta}_t^{SCAD}$ as $(\hat{\theta}_t^{SCAD^{(1)}}, \hat{\theta}_t^{SCAD^{(2)}})$, accordingly. To ascertain the asymptotic properties and derive the limiting distribution of the estimators within the high-dimensional and sparse framework, we incorporate the following supplementary regularity conditions (Zhao and Yu 2006), in conjunction with $A1 - A6$:

A7. There exists a positive constant C_1 at time t such that $\frac{1}{n} W_k^\top W_k \le C_1$ for all $k = 1, \ldots, K$ and for all n;
A8. There exists a positive constant C_2 such that $\zeta^\top M_t^{(1,1)} \zeta \ge C_2$ for all $\| \zeta \|_2^2 = 1$;
A9. $K^* = O(n^{b_1})$ for some $0 < b_1 < 1$;

A10. There exist positive constants b_2 and C_3 such that $b_1 < b_2 \leq 1$ and

$$n^{(1-b_2)/2} \min_{k=1,\ldots,p^*,\ldots,p^*+S} \left|\theta_t^{(1)}\right| \geq C_3.$$

When both the predictors and the lags of Y_t are normalized, $A7$ may become trivial. $A8$ necessitates that $M_t^{(1,1)}$ maintains eigenvalues bounded from below to ensure a proper inverse matrix. The assumption $A9$ addresses the divergence rate of the variable dimension p compared to the sample size n. It preserves the stability of the penalized estimator and ensures that the number of significant predictors K^* grows at a controlled polynomial rate relative to n, preventing overfitting while maintaining a sparsity structure in the model. The last assumption mandates a gap of size n^{b_2} between the decay rates of $\theta_t^{(1)}$ and $n^{-\frac{1}{2}}$, ensuring that the estimation is not overwhelmed by the error terms.

As a consequence, in the context of high-dimensional and sparse ADL under the assumptions $A1 - A10$, we propose that the SCAD penalized estimators satisfy the *oracle* property at time t, the proof of which is deferred to the Appendix according to (Kim, Choi, and Oh 2008).

- Consistency in variable selection, $\lim_{n\to\infty} P(\mathcal{A}_t^* = \mathcal{A}_t) = 1$;
- $\sqrt{n}\mathbb{A}_t(n^{-1}W_t^{(1)\top}W_t^{(1)}/\sigma^2)^{1/2}(\hat{\theta}_t^{SCAD^{(1)}} - \theta_t^{(1)}) \xrightarrow{d} N(\mathbf{0}, H)$, where $\mathbb{A}_t$ is an arbitrary matrix with $\mathbb{A}_t\mathbb{A}_t^\top \to H$ and H is a $K^* \times K^*$ nonnegative symmetric matrix which includes the elements of M_t in the set $\mathcal{A}_t$.

3. CHANGE POINT PROBLEM

The focus of this section is to test for possible changes in the parameters of the ADL model referring to equation (2.3). We study m panels and there are n observations in each panel. In other words, we have $m \times n$ observations forming the historical sample. After training the ADL model at time $t = m$, the further incoming data $\{Y_t, W_t\}$, $t = m+1, m+2, \ldots$ are monitored sequentially. Let T_m be the monitoring horizon. The historical ADL model is

$$Y_m = W_m\theta_m + \mathcal{E}. \tag{3.1}$$

The ADL model after historical data is

$$Y_t = W_t\theta_t + \mathcal{E}, \text{ for } t = m+1, m+2, \ldots. \tag{3.2}$$

At each new time point t, our goal is to test whether the ADL model remains the same as the historical model. That is $\theta_t = \theta_m$ for all $t = m+1, m+2, \ldots$. A change point occurs when $\theta_t \neq \theta_m$ for any $t = m+1, m+2, \ldots$.

In the context of the hypothesis testing, under the null hypothesis, there is no change in the coefficients

$$H_0 : \theta_t = \theta_m, \text{ for } t = m+1, m+2, \ldots. \tag{3.3}$$

Under the alternative hypothesis, there exists $\tau \geq 1$ such that

$$H_a : \begin{cases} \theta_t = \theta_m, & \text{for } t = m+1, \ldots, m+\tau, \\ \theta_t \neq \theta_m, & \text{for } t = m+\tau+1, \ldots, m+T_m. \end{cases} \tag{3.4}$$

Following Horváth and Hušková (2012), we consider the CUSUM of residuals

$$\Gamma(m, n, \tau) = \frac{1}{\hat{\sigma}_m} \left| \sum_{t=m+1}^{m+\tau} \hat{\mathcal{E}}_t \right|, \tag{3.5}$$

where $\hat{\mathcal{E}}_t = \frac{1}{n}\sum_{i=1}^{n}(Y_{i,t} - W_{i,t}\hat{\theta}_t^{SCAD})$ at $t = m+1, m+2, \ldots,$ and $\hat{\sigma}_m^2$ is the estimated constant error variance, given by

$$\hat{\sigma}_m^2 = \frac{1}{(m\text{-}\max\{s\})(n - K^{*SCAD})} \sum_{t=\max\{s\}+1}^{m} \left[\sum_{i=1}^{n} (Y_{i,t} - W_{i,t}\hat{\theta}_t^{SCAD}) \right]^2 \tag{3.6}$$

where s are the indices in $\mathcal{A}_t^*$ at time t. Given a constant $\gamma \in [0, 1/2)$, a normalizing function $g(m, \tau, \gamma)$ is defined as

$$g(m, \tau, \gamma) = m^{1/2}\left(1 + \frac{\tau}{m}\right)\left(\frac{\tau}{\tau + m}\right)^{\gamma}, \tag{3.7}$$

where γ is the control parameter. We propose the test statistic for monitoring structural change according to Horváth and Hušková (2012),

$$\Omega = \sup_{1 \leq \tau \leq T_m} \frac{\Gamma(m, n, \tau)}{g(m, \tau, \gamma)}. \tag{3.8}$$

Set $T_m < \infty$ with $\lim_{m\to\infty} T_m/m = N$ and $N > 0$. The stopping time for the monitoring process is given by

$$\Lambda(m) = \begin{cases} \inf\{\tau \geq 1 : & \text{if } \Gamma(m, n, \tau) \geq g(m, \tau, \gamma)Q_\alpha(\gamma)\}, \\ T_m & \text{for all } \tau = 1, \ldots, T_m, \end{cases} \tag{3.9}$$

where $Q_\alpha(\gamma)$ is the $(1-\alpha)$th quantile of the asymptotic distribution such that, under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda(m) < \infty) = \alpha. \tag{3.10}$$

Under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda(m) < \infty) = 1. \tag{3.11}$$

Theorem 3.1. *Under assumptions A1-A10, when the null hypothesis holds,*

$$\lim_{m\to\infty} P(\Omega \leq Q_\alpha(\gamma)) = P\left(\sup_{0 \leq g \leq N/(N+1)} \frac{\| \mathcal{W}(g) \|_\infty}{g^\gamma} \leq Q_\alpha(\gamma) \right),$$

where $\{\mathcal{W}(g), 0 \leq g < \infty\}$ denotes the q−dimensional Wiener process, and q depicts the number of significant variables in the model in terms of historical data.

The asymptotic distribution of test statistics can be derived using Theorem 3.1. We obtain the asymptotic critical value $Q_\alpha(\gamma)$ based on,

Table 1. Asymptotic critical values for various values of N and control parameter γ.

		γ		
α	N	0.00	0.25	0.49
0.010	2	2.4471	2.8169	3.7326
	4	2.6865	2.9540	3.7450
	6	2.7858	3.0044	3.7472
	9	2.8471	3.0451	3.7499
0.025	2	2.2118	2.5620	3.4637
	4	2.4306	2.6815	3.4744
	6	2.5148	2.7352	3.4786
	9	2.5721	2.7675	3.4799
0.050	2	2.0209	2.3590	3.2573
	4	2.2224	2.4698	3.2682
	6	2.2974	2.5104	3.2717
	9	2.3515	2.5393	3.2741

$$P\left(\sup_{0\leq g\leq N/(N+1)} \frac{\| \mathcal{W}(g)\|_\infty}{g^\gamma} \geq Q_\alpha(\gamma)\right) = \alpha,$$

where $0 < \alpha < 1$ and $0 \leq \gamma < 1/2$.

Theorem 3.2. *Under assumptions A1-A10, if the alternative hypothesis is true, we have*

$$\sup_{1\leq\tau\leq T_m} \frac{\Gamma(m, n, \tau)}{g(m, \tau, \gamma)} \to \infty \quad \text{as} \quad m \to \infty.$$

The proofs of Theorems 3.1–3.2 are given in the supplementary material. Referring to Ratnasingam and Ning (2021) and Gu and Ratnasingam (2023), the asymptotic critical values for various γ and N values are given in Table 1.

Ratnasingam and Ning (2021) and Gu and Ratnasingam (2023) suggest that the sensitivity of the monitoring process increases with larger γ values. Thus, a sufficiently high value of *gamma* should be considered when we believe any change point occurs shortly after m in a simple linear model. In the following section, we conduct a simulation study to investigate this property in the ADL model.

4. SIMULATION STUDY

We consider the following three criteria that are commonly used as the determination of the quality of a sequential change point detection approach:

1. Type I error rate: Close to the nominal level,
2. Power of the test: Preferably close to 1,
3. Detection time under the alternative hypothesis: Stop as soon as possible after a change is noticed.

Below are two parameter settings for the ADL model, encompassing various conditions examined in this section. We generate data that satisfy stationarity, exogeneity, bounded eigenvalues, and the gap condition on nonzero coefficients, following assumptions A1–A10. For instance, the error term is generated as independent and identically distributed variables with finite fourth moments (A5), the predictors are normalized and constructed to avoid perfect multicollinearity (A4–A7), and the minimal signal condition (A10) holds by setting a sufficiently large gap between zero and nonzero coefficients. Furthermore, all simulation outcomes are derived from 1,000 repetitions.

Setting I (Type I error calculations)

- Under H_0, the true parameter vectors

$$\theta_t = \begin{cases} \beta_{6,t} = \beta_{8,t} = 1, \\ \beta_{5,t} = \beta_{7,t} = \beta_{9,t} = -1, \\ \eta_1 = 0.5, \\ 0, \text{Otherwise}. \end{cases} \tag{4.1}$$

- The predictor variables $X_{j,t}$ for all $j \in \{1, \ldots, p\}$ have standard normal distribution $N(0, 1)$.
- The model errors $\mathcal{E}_t$ are i.i.d. $N(0, 1)$.

Setting II (Stopping time and power analysis)

- Parameter settings under H_0, distributions of $X_{j,t}$ and $\mathcal{E}_t$ are the same as Setting I.
- Scenario 1 under H_a, the true parameter vectors after $t = \tau$

$$\theta_t = \begin{cases} \beta_{6,t} = \beta_{8,t} = 4, \\ \beta_{5,t} = \beta_{7,t} = \beta_{9,t} = -4, \\ \eta_1 = 0.5, \\ 0, \text{Otherwise}. \end{cases} \tag{4.2}$$

- Scenario 2 under H_a, the true parameter vectors after $t = \tau$

$$\theta_t = \begin{cases} \beta_{6,t} = \beta_{8,t} = 1, \\ \beta_{5,t} = \beta_{7,t} = \beta_{9,t} = -1, \\ \eta_1 = \eta_2 = 0.5, \\ 0, \text{Otherwise}. \end{cases} \tag{4.3}$$

First, we conduct the Type I error analysis for the ADL model under varying conditions. The comparisons are made across different sample sizes ($n =$ 50, 100, 200) in each panel, different values of the control parameter γ $(0, 0.25, 0.45)$, different historical panel sizes $(m = 25, 50, 100)$, different ratios of monitoring and historical data ($\frac{T_m}{m} = N = 2, 4, 6, 9$), and different dimensions $p = 10, 100$. The black-dashed horizontal line represents the nominal 5% Type I error level. In Figure 1, for $p = 10$, we observe the empirical Type I error rates are overall controlled by the nominal level of 0.05. There are obvious patterns indicating the impacts of n, m, γ, and N. When the sample size n and panel size m are both small in the historical data, the empirical Type I error rates are slightly inflated, and the simulation results suggest that a lower control parameter γ should be more appropriate to control the nominal Type I error under this condition. As n and m increase, a larger γ is recommended to prevent the empirical Type I error from being deflated. In general, the empirical Type I error moves closer to the nominal level when N is getting larger, which indicates we have more monitoring data.

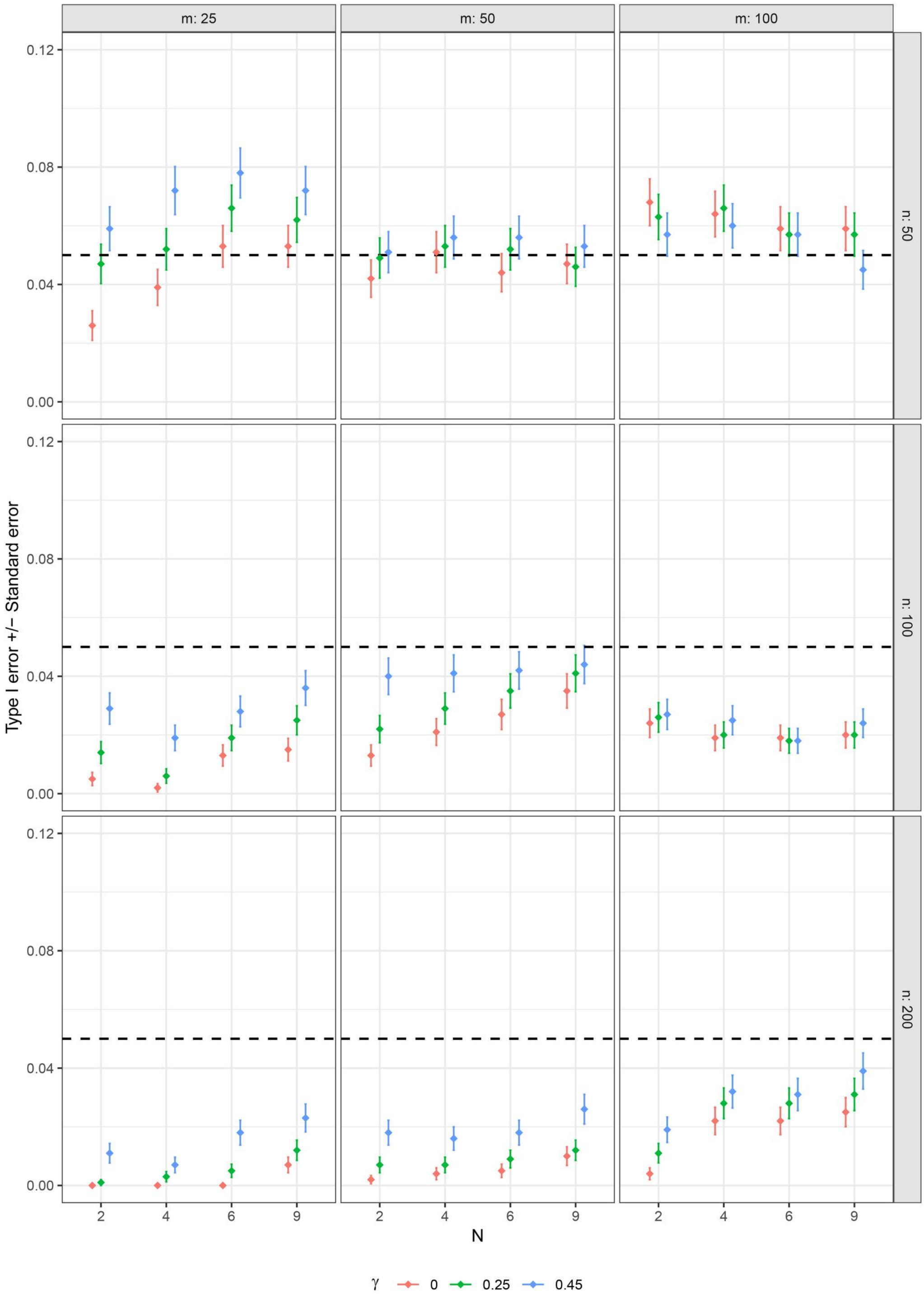

Figure 1. Comparison of Type I error rates for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, with different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, and different control parameter $\gamma \in \{0, 0.25, 0.45\}$ when $p = 10$.

Comparing Figures 1 and 2, we observe that as p increases from 10 to 100, the Type I error rates tend to decrease. This is likely because with a higher dimension p, the test statistic becomes more sensitive to deviations from the null hypothesis, leading to a more conservative test and a lower Type I error. We notice the same patterns in Figure 2 when p goes up to 100.

Figures 3 and 5 present a power comparison for various ratios of monitoring and historical data under two distinct scenarios. In both figures, the x-axis represents change points (τ), and the y-axis represents power. Figure 3 (Scenario 1) examines the power when change points are caused by changes in β_t, while the lagged effects η_S remain consistent. Figure 5 (Scenario 2) explores the power when change points are solely driven by the additional lagged effect. In Figure 3, a clear trend emerges: for a fixed N, smaller control parameters (γ) or later change points (τ) generally result in lower power. This is because smaller γ values require a larger CUSUM to exceed the threshold, and later change points leave less time for the sum to accumulate. Conversely, power increases as the sample size (n) and panel size (m) increase, aligning with expectations. Larger n and m provide more data, leading to more accurate estimates and improved detection. The figure also shows that as N increases, indicating earlier change points, the detection power significantly increases. This is because CUSUM statistics are more effective at detecting early change points due to the longer time available for the CUSUM to exceed the threshold.

Figure 5 exhibits similar trends to Scenario 1. However, a key difference is that detecting changes in Scenario 2 is generally more challenging. This is because the lagged effects are typically smaller than 1, requiring more data to achieve comparable power levels. While the overall patterns are similar, the power values in Figure 5 tend to be lower than those in Figure 3 for comparable parameter settings. This suggests that detecting changes driven solely by lagged effects is more difficult than detecting changes in the direct effect (β_{t}). For example, comparing the top-left plots ($N = 2, m = 25$) in both figures, the power in Figure 3 is noticeably higher than in Figure 5 for all change point locations. This difference highlights the impact of the change mechanism on the detection power. In both scenarios, the tradeoff between power and Type I error is crucial when selecting γ, especially when historical data are limited. As n and m increase, a larger γ becomes preferable.

Figures 4 and 6 present power comparisons for various ratios of monitoring and historical data, sample sizes, panel sizes, control parameters, and change point locations under two different scenarios, both with $p = 100$. Figure 4 (Scenario 1) examines the power when change points are caused by changes in β_{t}, while the lagged effects η_S remain consistent, and Figure 6 (Scenario 2) explores the power when change points are solely driven by the additional lagged effect. In Figure 4, a clear trend emerges: for a fixed N, smaller control parameters (γ) or later change points (τ) generally result in lower power. Conversely, power increases as the sample size (n) and panel size (m) increase. The figure also shows that as N increases, indicating earlier change points, the detection power significantly increases. Figure 6 exhibits similar trends to Scenario 1. However, a key difference is that detecting changes in Scenario 2 is generally more challenging. While the overall patterns are similar, the power values in Figure 6 tend to be lower than those in Figure 4 for comparable parameter settings. This suggests that

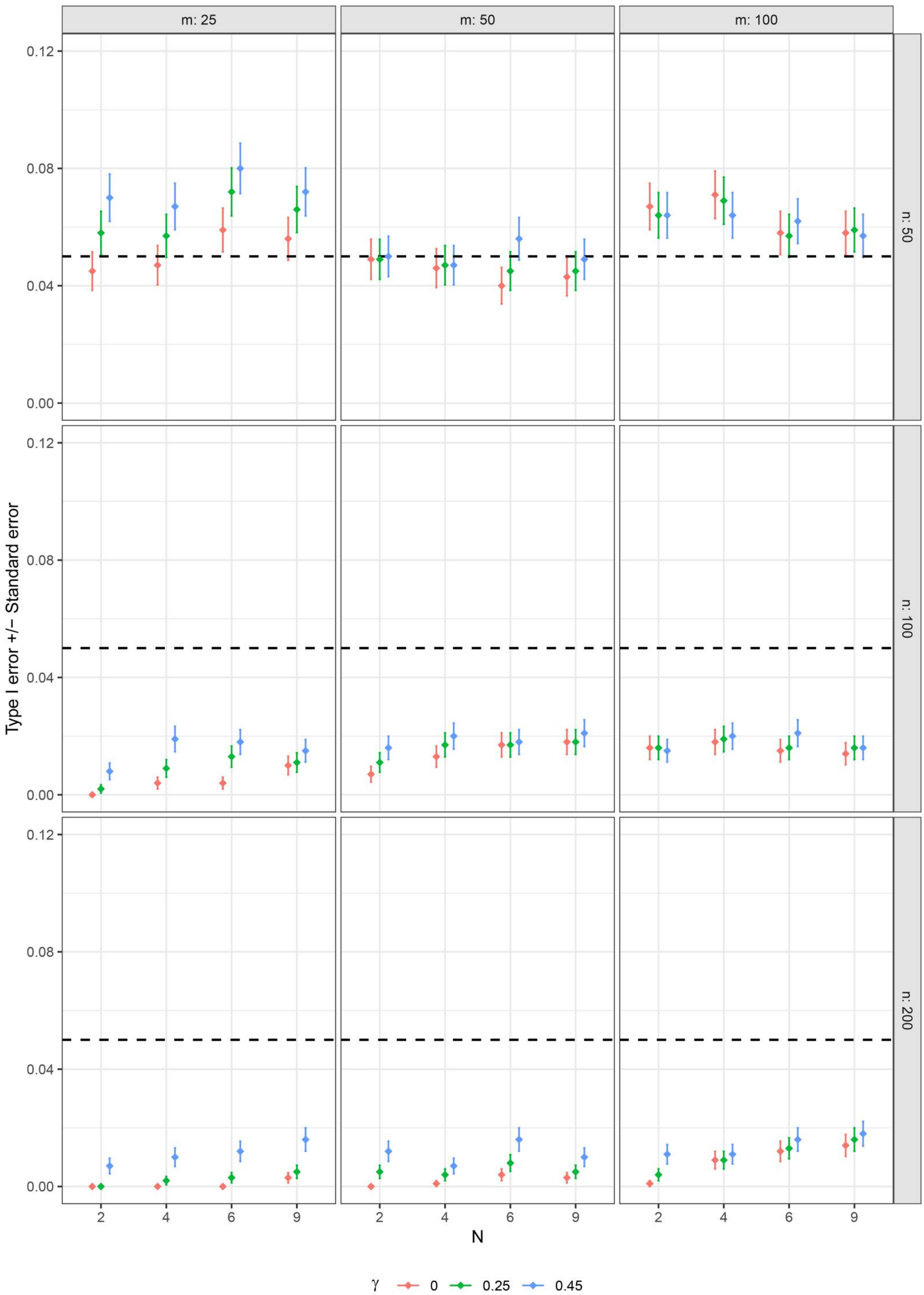

Figure 2. Comparison of Type I error rates for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, with different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, and different control parameter $\gamma \in \{0, 0.25, 0.45\}$ when $p = 100$.

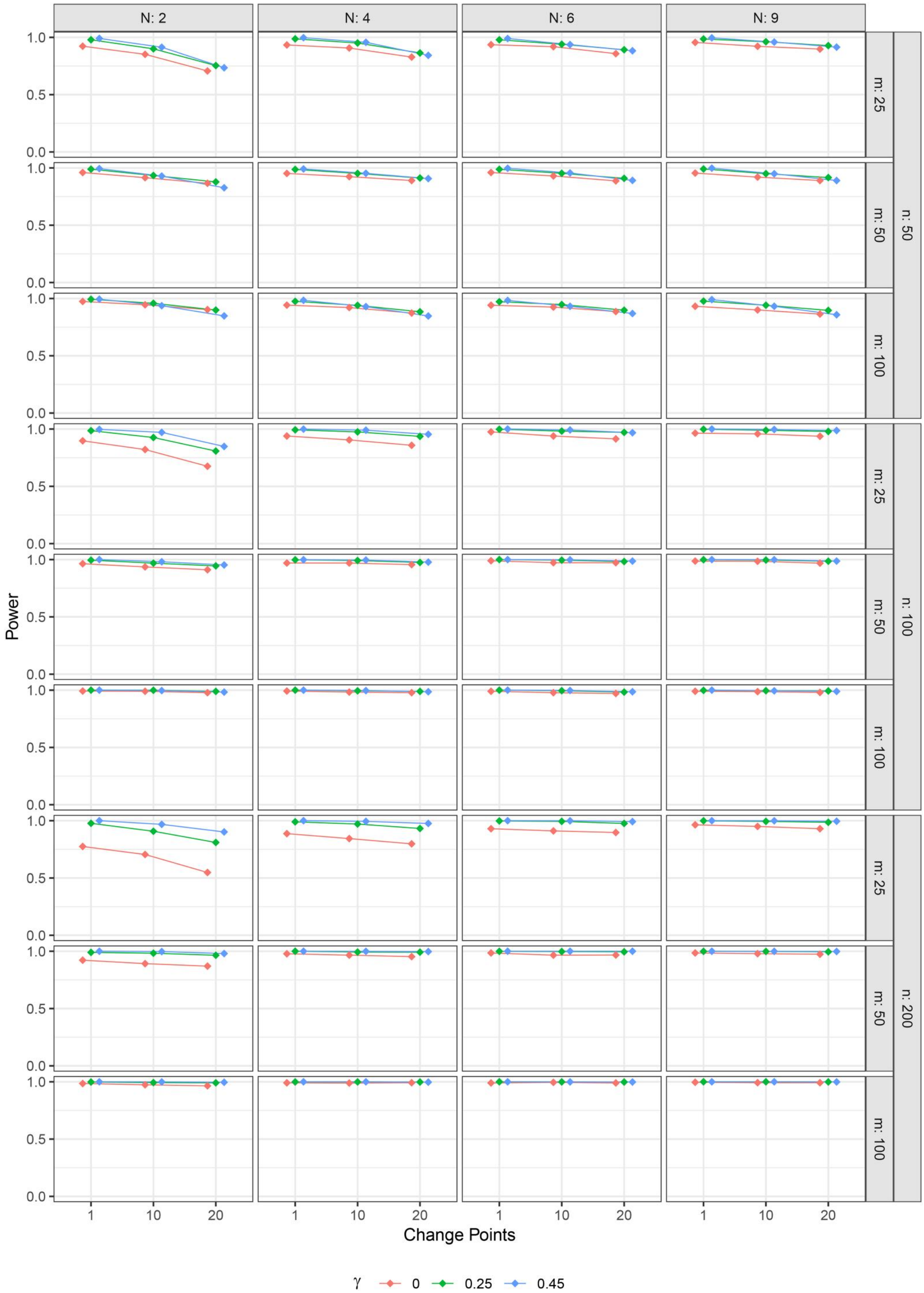

Figure 3. Scenario 1 – Power comparison for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{1, 10, 20\}$, when $p = 10$.

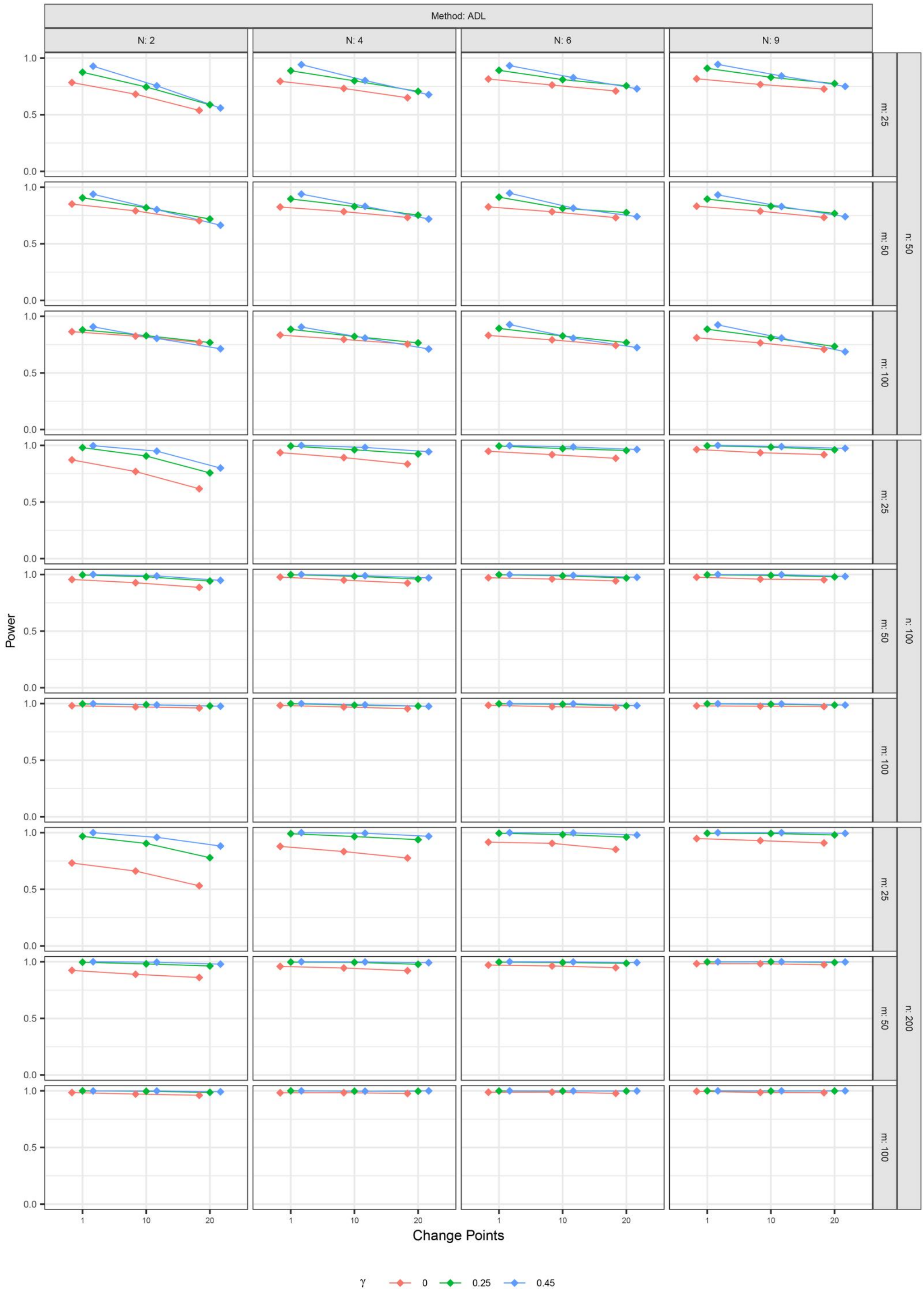

Figure 4. Scenario 1 – Power comparison for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{1, 10, 20\}$, when $p = 100$.

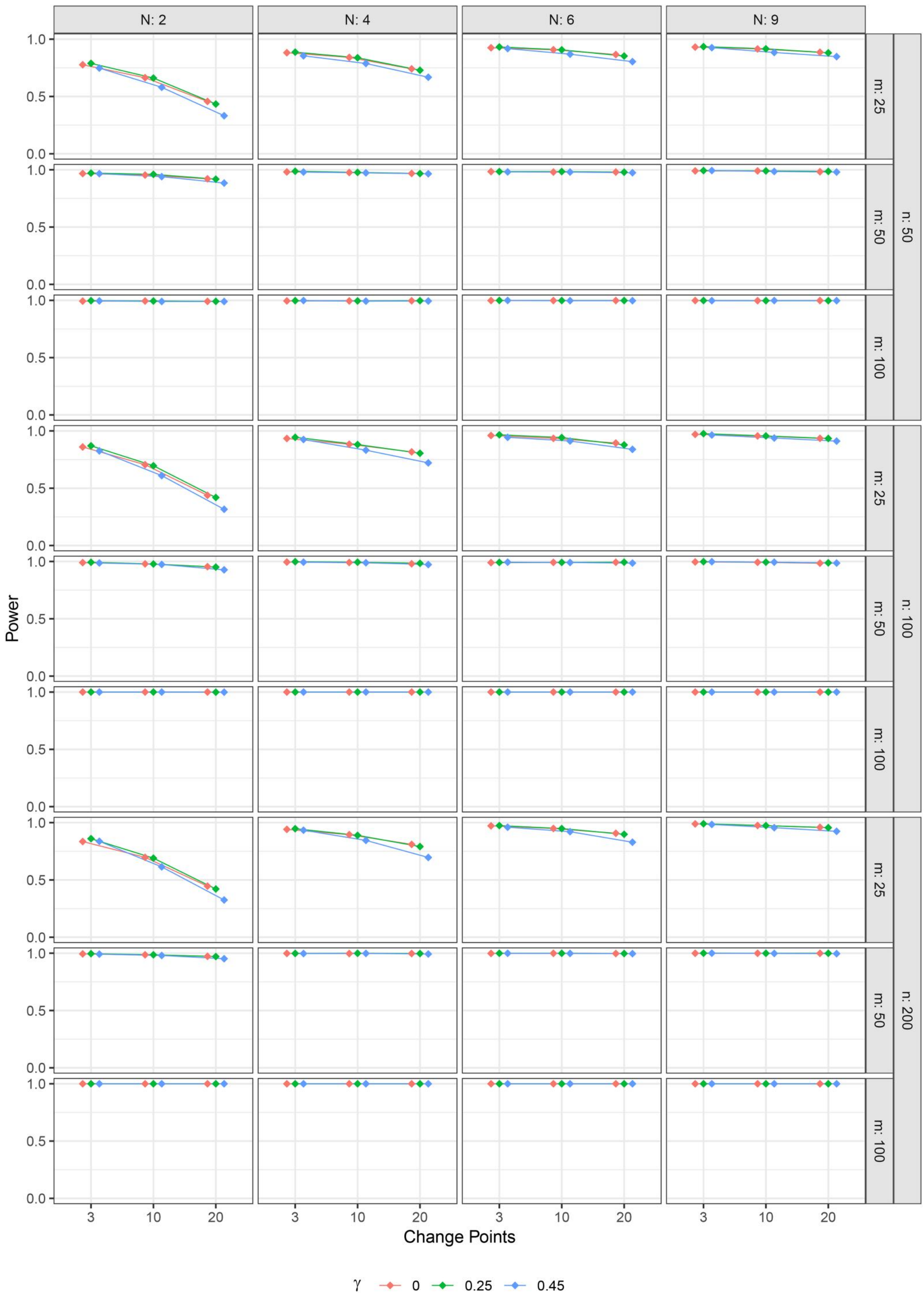

Figure 5. Scenario 2 – Power comparison for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{3, 10, 20\}$, when $p = 10$.

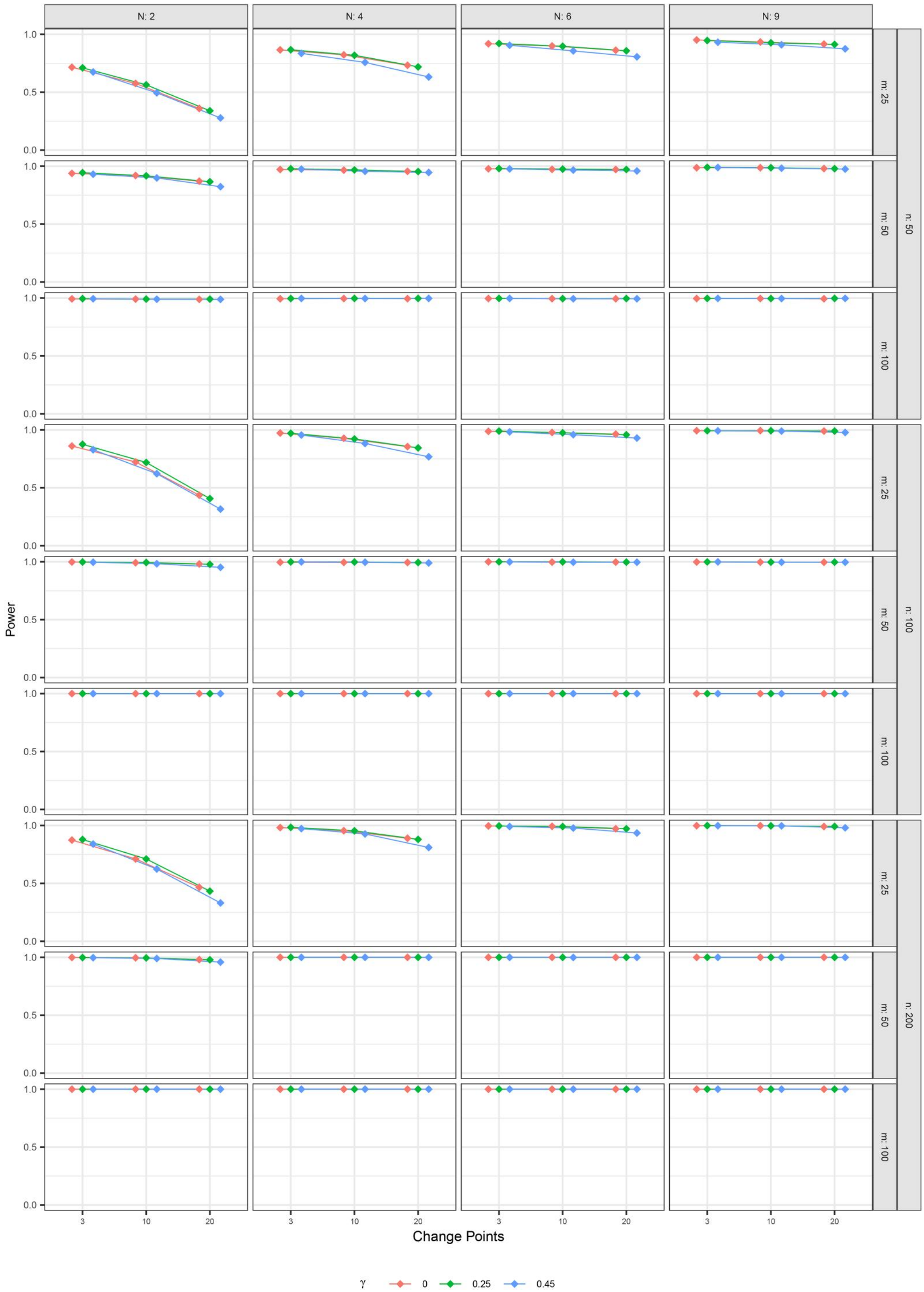

Figure 6. Scenario 2 – Power comparison for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{3, 10, 20\}$, when $p = 100$.

detecting changes driven solely by lagged effects is more difficult than detecting changes in the direct effect (β_t). Comparing Figures 3 and 5 (with $p = 10$) to Figures 4 and 6 (with $p = 100$), we observe that the overall trends remain consistent. In both sets of figures, smaller control parameters or later change points generally result in lower power, while power increases with sample size and panel size. However, the magnitude of the power values differs significantly between the two sets of figures. Specifically, the power values in Figures 3 and 5 $(p = 10)$ are generally higher than those in Figures 4 and 6 $(p = 100)$ for comparable parameter settings. This suggests that the detection power is higher when $p = 10$ compared to when $p = 100$, indicating that the choice of p significantly impacts the detection power.

In our simulation study, we use stopping time as a metric to assess the accuracy and efficacy of the proposed approach. Concurrently, we analyze the properties of different parameter configurations. Optimal settings for utilizing the approach are those where the monitored stopping time closely aligns with the actual change point. Similar to power analysis, we initially examine Scenario 1 across three distinct situations under $p = 10$, assuming change points at $\tau = 1, 10$, and 20, respectively. Figure 7 presents the median stopping time compared to the true change point. Overall, the proposed approach effectively identifies the change point across various settings. The results indicate a preference for a larger control parameter γ in training the model if promptly stopping the detection process upon a change point occurrence is desired. Moreover, this trend becomes more pronounced as sample size n, historical panel size m, and monitoring data N increase. It is worth noting that as N increases while n and m are held constant, the stopping time becomes more delayed. However, we observe that increasing n and m can notably alleviate this delay.

Under Scenario 2 with $p = 10$, as depicted in Figure 9, where only one additional lagged term of Y_t occurs, our method consistently detects this change. Comparing Figures 7 and 9, we observe that the stopping times in Scenario 2 (Figure 9) tend to be slightly higher than those in Scenario 1 (Figure 7) for similar parameter settings. This suggests that detecting changes based on lagged terms of Y_t might require a bit more time compared to detecting changes directly influenced by X_t, as in Scenario 1. The efficacy in Scenario 2 remains contingent upon the selection of γ, aligning with the aforementioned findings illustrated in Figure 9. In contrast to Scenario 1, the observed stopping time exhibits less precision due to the smaller variations in lagged effects compared to the fluctuations induced by X_t in our settings. This coincides with the argument in the power analysis, indicating that more data are required to identify a change point under this circumstance. These results are confirmed in higher dimensions of p, as demonstrated in Figures 8 and 10. We include the programming code and additional simulation results for $p = 200$ and $p = 400$ in the supplementary material.

5. REAL DATA APPLICATION

In this section, we apply the proposed sequential change point detection method to a dataset consisting of 84 features across 422 companies. As outlined in equation (2.2), we consider the following model to demonstrate the effectiveness of the proposed monitoring procedure over time t:

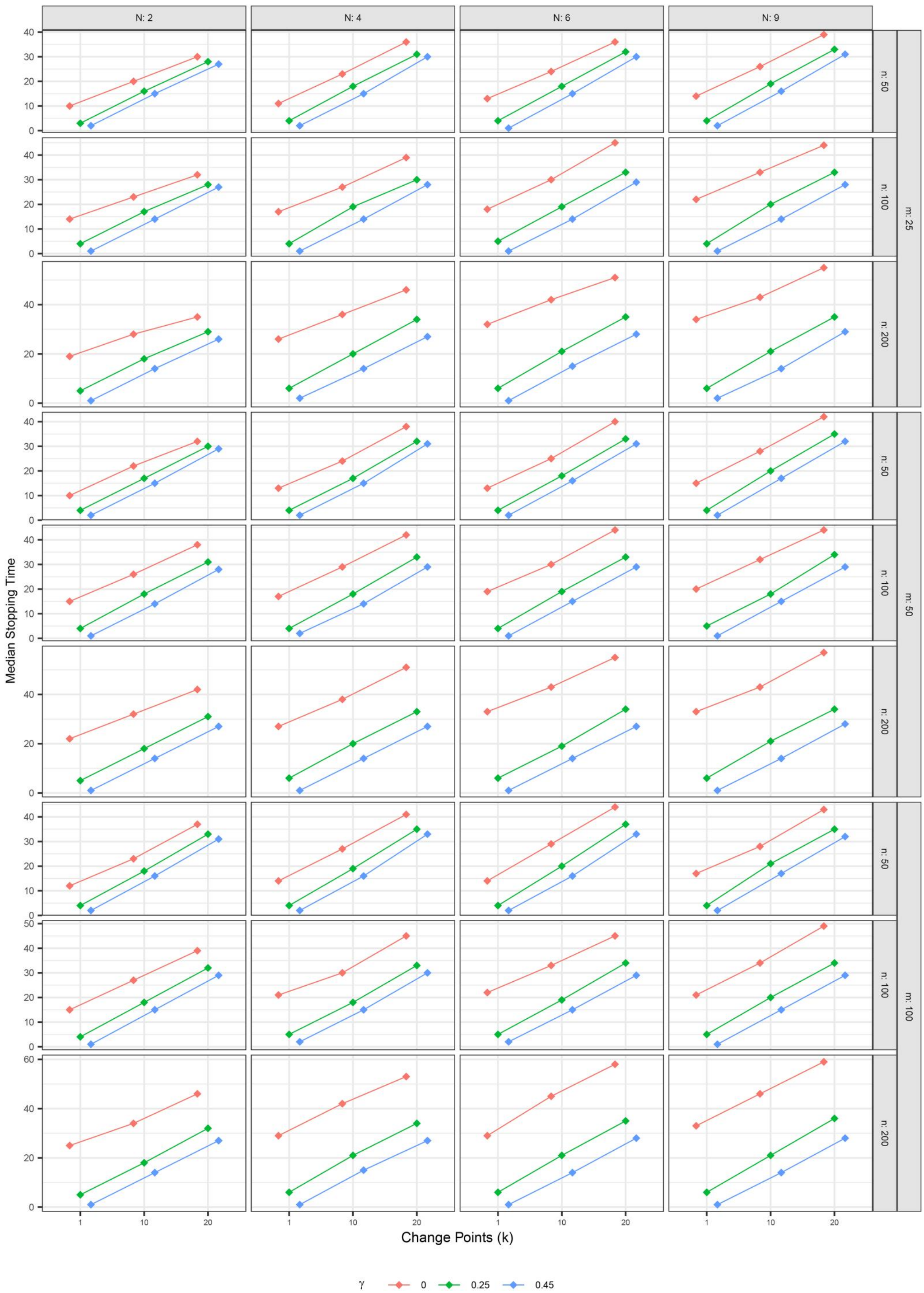

Figure 7. Scenario 1 – Median stopping time for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{1, 10, 20\}$, when $p = 10$.

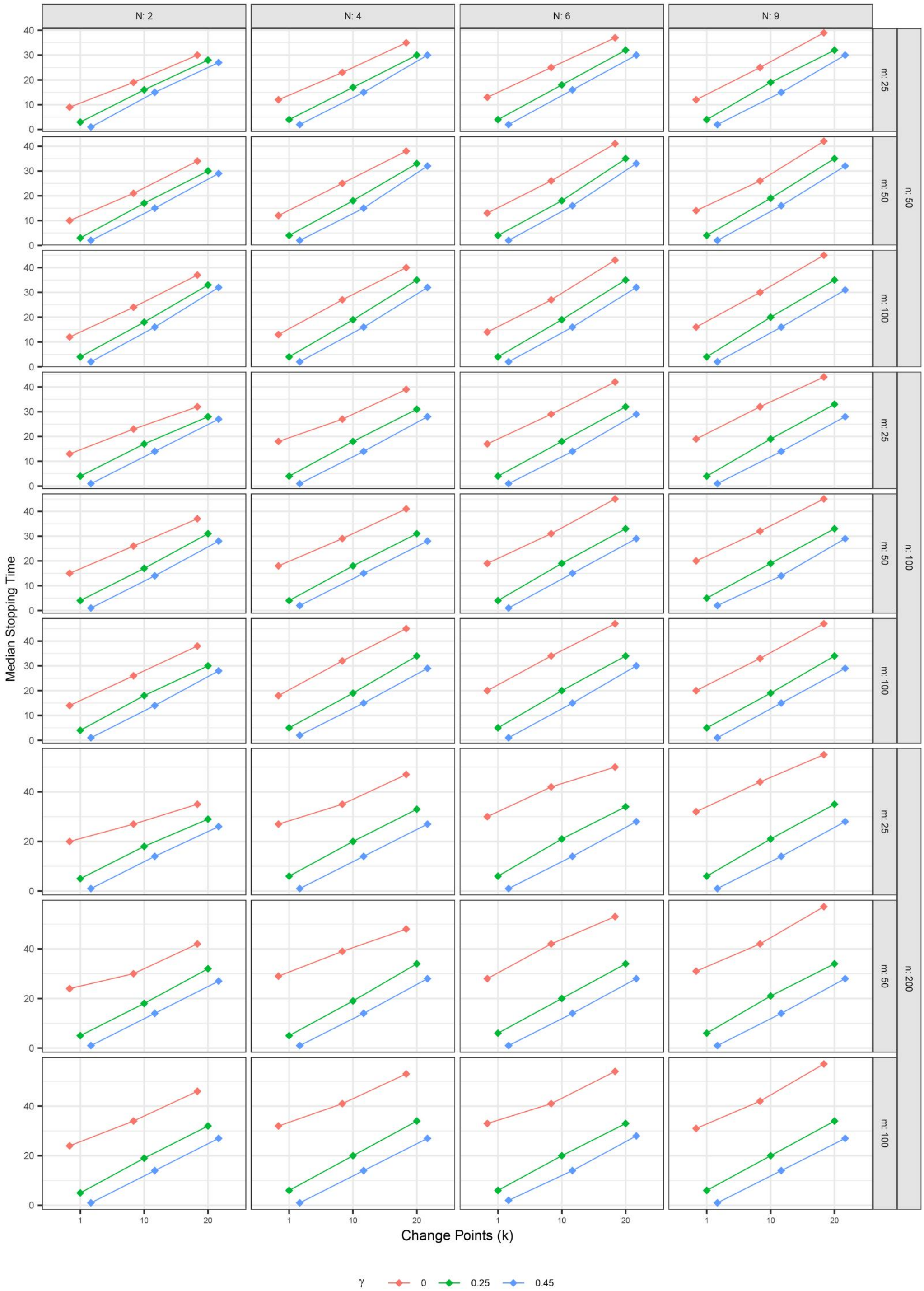

Figure 8. Scenario 1 – Median stopping time for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{1, 10, 20\}$, when $p = 100$.

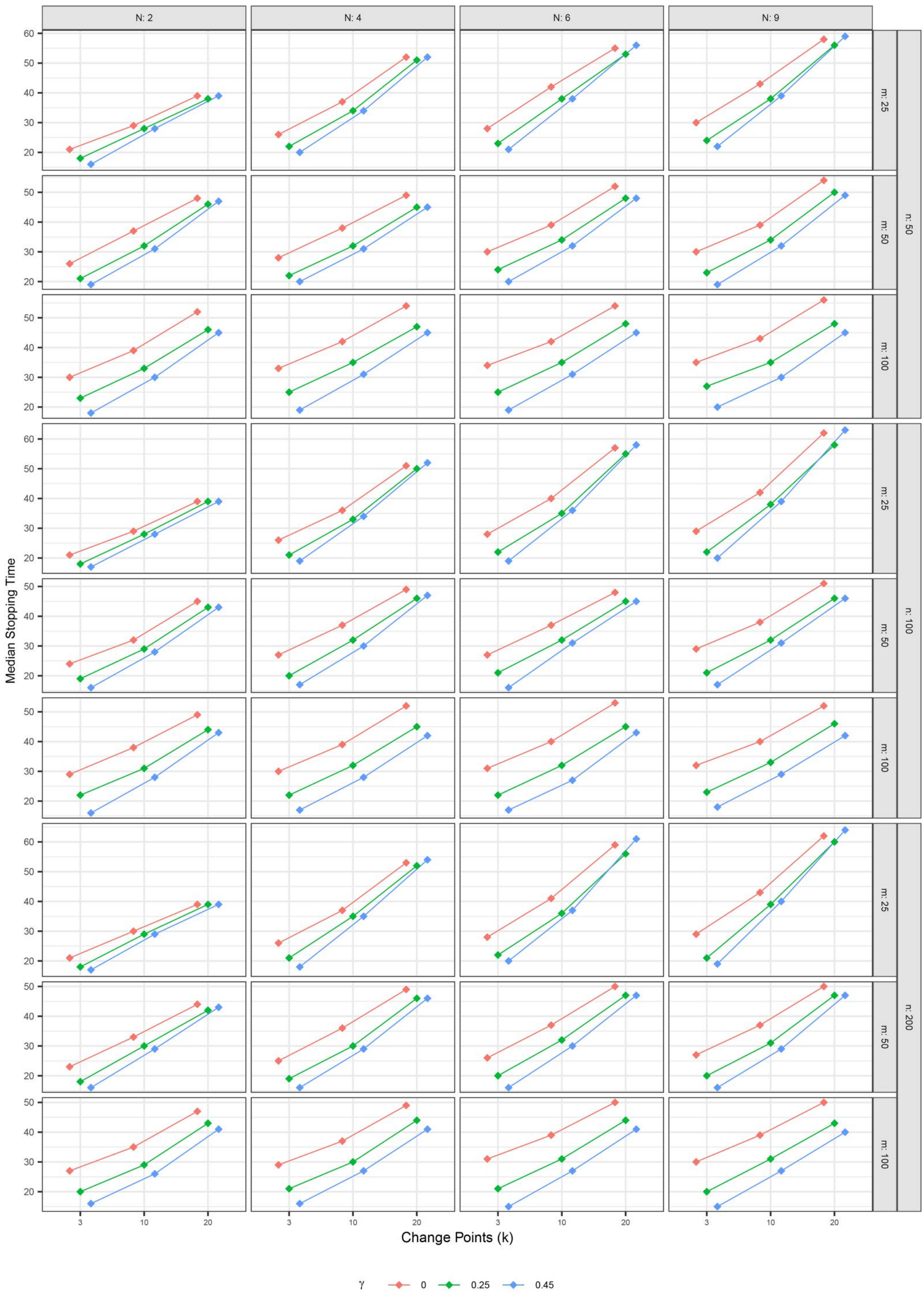

Figure 9. Scenario 2 – Median stopping time for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{3, 10, 20\}$, when $p = 10$.

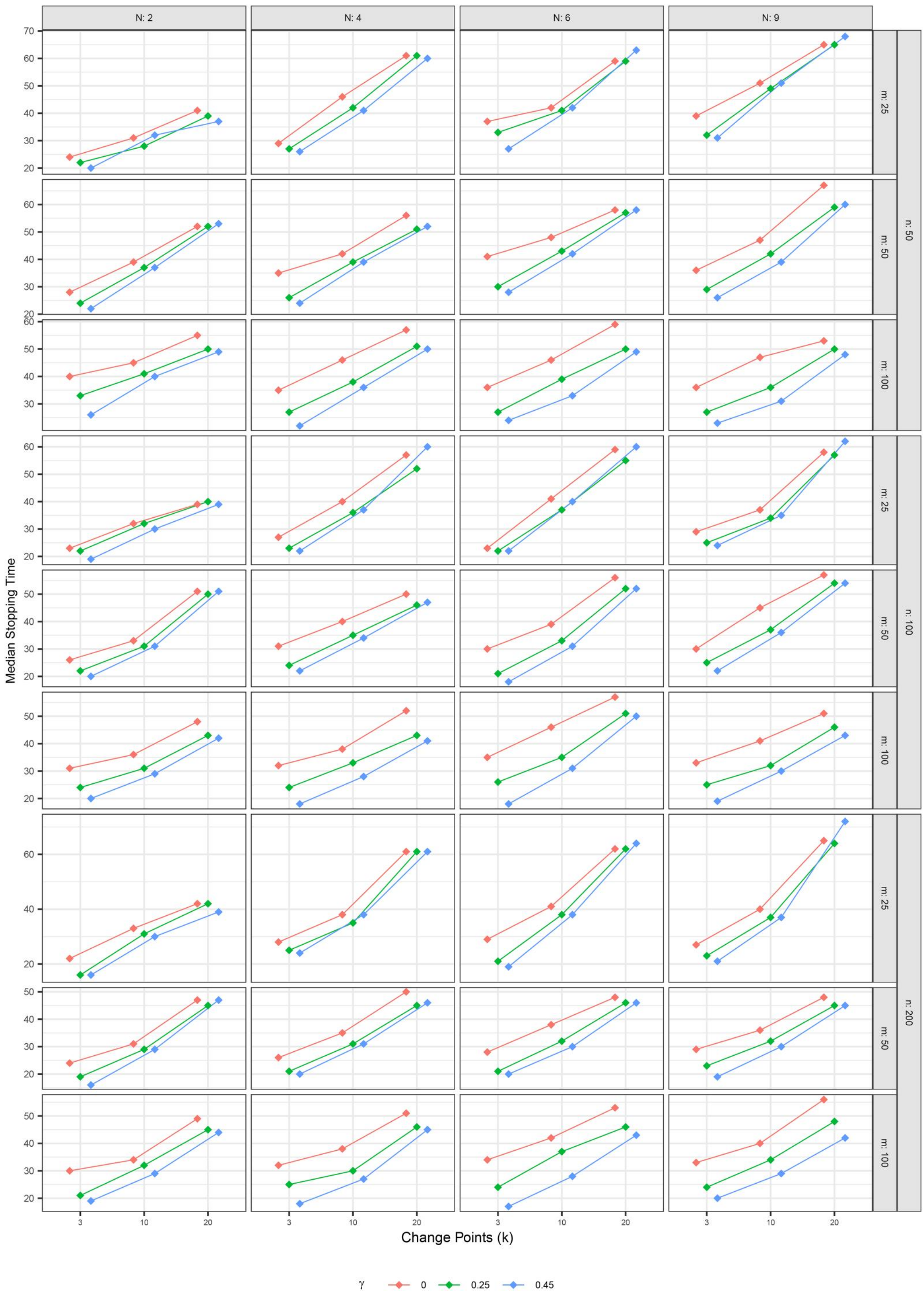

Figure 10. Scenario 2 – Median stopping time for various ratios of monitoring and historical data $N \in \{2, 4, 6, 9\}$, across different sample sizes $n \in \{50, 100, 200\}$, panel sizes $m \in \{25, 50, 100\}$, control parameters $\gamma \in \{0, 0.25, 0.45\}$, and change point locations $\tau \in \{3, 10, 20\}$, when $p = 100$.

$$Y_t = W_t\theta_t + \mathcal{E},$$

where $W_t = (X_t, \mathbf{Y_{t^-}})$ and $\theta_t = (\beta_t, \boldsymbol{\eta}_\mathbf{S})^\top$.

Recall our discussions in Section 3, the historical ADL model is

$$Y_m = W_m\theta_m + \mathcal{E}.$$

The ADL model after historical data is

$$Y_t = W_t\theta_t + \mathcal{E}, \text{ for } t = m+1, m+2, \ldots.$$

Under the null hypothesis, there is no change in the coefficients

$$H_0 : \theta_t = \theta_m, \text{ for } t = m+1, m+2, \ldots.$$

Under the alternative hypothesis, T_m is our monitoring horizon, a change point τ occurs when there exists $\tau \geq 1$ such that

$$H_a : \begin{cases} \theta_t = \theta_m, & \text{for } t = m+1, \ldots, m+\tau, \\ \theta_t \neq \theta_m, & \text{for } t = m+\tau+1, \ldots, m+T_m. \end{cases}$$

Our dataset is accessible on *Kaggle* (Ebrahimi 2024). For this study, we consider "financial distress" as the response variable Y, and $\mathbf{Y_{t^-}}$ represents the lagged financial distress. A company is classified as healthy if its distress value exceeds -0.50. Otherwise, it is considered financially distressed. The matrix X denotes the collection of predictor variables, in which there are 83 features, denoted x_1 through x_{83}, reflecting various financial and non-financial attributes of the companies. Due to privacy concerns, the specific feature variables are not disclosed. We also observe financial distress values over up to 14 time periods for these companies. The chosen data contains 132 companies with complete observations for all 14 time periods.

To address the challenges posed by heteroscedasticity and outliers in this real-world data, it is often necessary to normalize the distribution of Y. This transformation helps stabilize the variance of residuals, thereby improving the model's ability to meet the assumption of homoscedasticity. Additionally, normalizing Y can mitigate the influence of extreme values, making the model more robust and enhancing its predictive accuracy. The histogram (a) and Q–Q plot (b) in Figure 11 reveal that the data do not follow a normal distribution, a result confirmed by the Kolmogorov–Smirnov test. Consequently, we apply the transformation $\text{sign}(Y) \times |Y|^{1/2}$ to the response variable Y. The transformed data are depicted in the histogram (c) and Q-Q plot (d) in Figure 11.

Figure 12 depicts the transformed financial distress across 14 time periods for the selected 132 companies. The first 7 time periods are considered as historical panels, assuming no change points, i.e., $m = 7$ and $T_m = 7$. Intuitively, there is an indication of a potential change point at the last time period ($t = 14$), following the historical panels. By applying the proposed approach, we seek to thoroughly evaluate whether the observed shift is statistically significant, thereby providing more robust evidence for structural changes in the financial distress patterns over time.

We first apply the proposed SCAD-penalized dynamic panel model to simultaneously estimate the effects of predictors and time effects on the response variable. Based on Table 2, the analysis identifies 43 significant predictors out of 83 financial and non-financial factors. It further demonstrates that financial distress in the two preceding

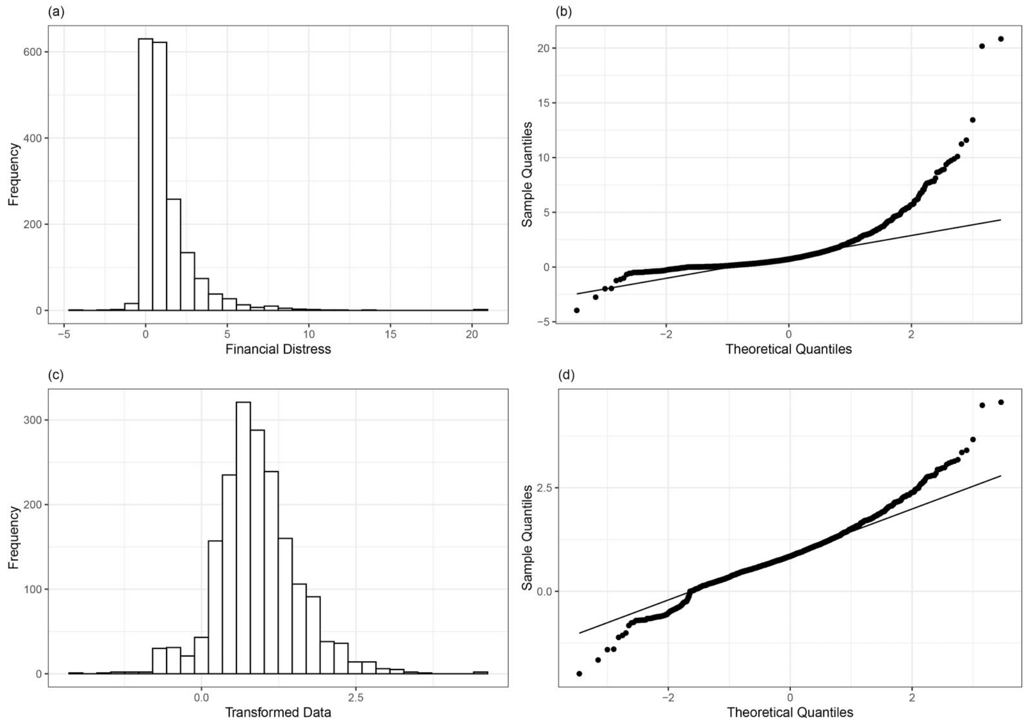

Figure 11. Normality check for financial distress.

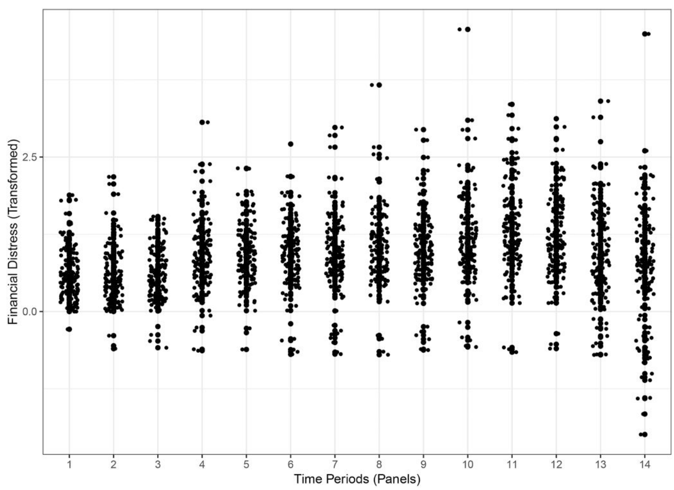

Figure 12. Change point in transformed financial distress across 14 time periods (panels), with the initial 7 periods considered historical.

time periods significantly affects the current financial distress of the companies in the data. Following this, we employ the proposed CUSUM testing procedure to detect structural changes in the distribution of Y over time t, at a confidence level of $\alpha = 0.05$. In this application, two different values of γ are considered: 0.25 and 0.45. After the historical period, we continue to monitor the process sequentially. For $\gamma = 0.25$, no change

Table 2. Significant predictors and lags estimated by the proposed SCAD-penalized dynamic panel model.

Predictor	Estimate	Predictor	Estimate	Predictor	Estimate
x_1	−1.3916	x_{16}	−0.5965	x_{31}	1.9901
x_2	−4.8224	x_{17}	−1.0084	x_{32}	1.2891
x_3	−4.7779	x_{18}	−0.2293	x_{33}	−1.0013
x_4	−0.3789	x_{19}	11.5438	x_{34}	−1.9113
x_5	−2.4284	x_{20}	−1.7763	x_{35}	0.9516
x_6	0.9848	x_{21}	0.9226	x_{36}	1.2376
x_7	−1.0362	x_{22}	3.7456	x_{37}	0.2838
x_8	3.1708	x_{23}	−1.3097	x_{38}	0.7219
x_9	0.9733	x_{24}	−1.8517	x_{39}	0.0632
x_{10}	2.5348	x_{25}	−0.8667	x_{40}	10.0792
x_{11}	−1.7688	x_{26}	−0.0637	x_{41}	−6.3132
x_{12}	−2.9114	x_{27}	−0.3113	x_{42}	−0.0701
x_{13}	−2.4392	x_{28}	−0.2007	x_{43}	0.1479
x_{14}	1.4981	x_{29}	0.0537	y_6	0.1031
x_{15}	0.2699	x_{30}	−0.1165	y_5	0.0525

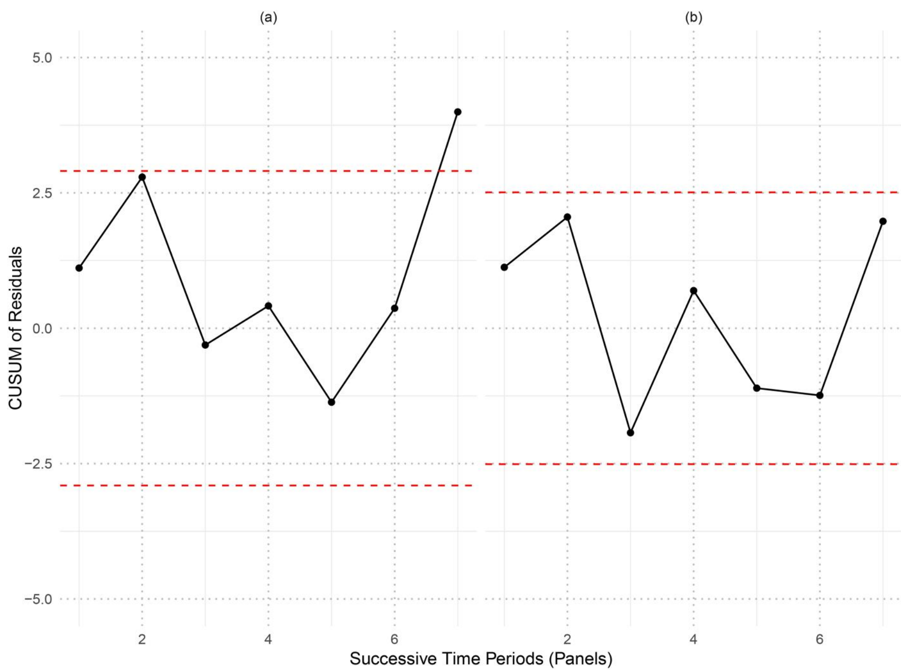

Figure 13. The CUSUM plot of residuals over successive panels, with red dashed lines indicating the detection thresholds at $\alpha = 0.05$: (a) $\gamma = 0.45$ and (b) $\gamma = 0.25$.

points are identified, suggesting stability in the process. However, when $\gamma = 0.45$, a change point is detected at $t = 14$, as shown in Figure 13. This result indicates a significant shift in the behavior of the response variable, highlighting the importance of selecting an appropriate γ value for effective change point detection.

6. CONCLUSION

Analyzing panel data poses significant challenges for researchers aiming to explore intricate relationships between variables and detect change points sequentially. The outstanding models often require sophisticated solutions to handle the complexities of variable interactions over time and across different entities. These models efficiently capture the dynamics of the data, allowing for a nuanced exploration of causal effects and structural changes. They combine information from both cross-sectional units and time periods, providing comprehensive insights into the underlying processes. The desired characteristics include robustness to various data structures, flexibility to accommodate different model specifications, and scalability to handle large datasets.

In this research, we focused on the SCAD-penalized ADL model which simplifies the complexity of GMM and IV models under specific assumptions. We introduced a CUSUM-based testing procedure to sequentially monitor structural changes in the monitoring data, extendable to high-dimensional settings. The consistency of our method and the *oracle* property of the resulting regularized estimators were examined. The asymptotic properties of the test statistics were established under both the null and alternative hypotheses. Our simulation results highlighted the impacts of panel size, historical sample size per panel, monitoring horizon, and control parameters. The proposed approach consistently identified changes in the subsequent monitoring timeframe across both simulated data and real data applications. With limited historical data, careful control parameter selection is essential to control Type I errors and maintain reasonable power. Given sufficient data, a higher control parameter improves alignment with nominal Type I error, enhances power, and increases stopping time accuracy. Furthermore, we will aim to relax certain constraints, requiring the exploration of more complex dynamic models such as GMM and IV. In addition, developing an enhanced testing procedure to distinguish changes caused by lagged effects, predictors, or both can be a valuable extension of this study.

ACKNOWLEDGMENTS

The authors thank three anonymous referees and the Associate Editor for their constructive comments and suggestions, which helped to improve this article significantly.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

ORCID

Suthakaran Ratnasingam http://orcid.org/0000-0002-6802-192X

REFERENCES

Arellano, M., and S. Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence.” *Review of Economic Studies* 58 (2): 277–297. https://doi.org/10.2307/2297968.

Bai, J. 2009. “Panel Data Models with Interactive Fixed Effects.” *Econometrica* 77: 1229–1279. https://www.jstor.org/stable/40263859.

Baltagi, B. 2005. “Econometrics of Panel Data.” *Wiley Series in Probability and Statistics*. 3rd ed. Hoboken, NJ: John Wiley & Sons.

Ebrahimi. 2024. *Financial Distress Prediction*. San Francisco, CA: Kaggle.

Fan, J., and R. Li. 2001. “Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties.” *Journal of the American Statistical Association* 96 (456): 1348–1360. https://doi.org/10.1198/016214501753382273.

Frees, E. W. 2004. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge: Cambridge University Press.

Gu, C., and S. Ratnasingam. 2023. “Real-Time Change Point Detection in Linear Models Using the Ranking Selection Procedure.” *Sequential Analysis* 42 (2): 129–149. https://doi.org/10.1080/07474946.2023.2187416.

Hansen, L. P. 1982. “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica* 50 (4): 1029–1054. https://doi.org/10.2307/1912775.

Horváth, L., and M. Hušková. 2012. “Change-Point Detection in Panel Data.” *Journal of Time Series Analysis* 33 (4): 631–648. https://doi.org/10.1111/j.1467-9892.2012.00796.x.

Horváth, L., M. Hušková, P. Kokoszka, and J. Steinebach. 2004. “Monitoring Changes in Linear Models.” *Journal of Statistical Planning and Inference* 126 (1): 225–251. https://doi.org/10.1016/j.jspi.2003.07.014.

Horváth, L., and G. Rice. 2024. *Change Point Analysis for Time Series*. Berlin, Germany: Springer Series in Statistics. Springer.

Horváth, L., Z. Liu, G. Rice, and Y. Zhao. 2022. “Detecting Common Breaks in the Means of High Dimensional Cross-Dependent Panels.” *The Econometrics Journal* 25 (2): 362–383. https://doi.org/10.1093/ectj/utab028.

Hsiao, C. 2014. *Analysis of Panel Data*. 3rd ed. Cambridge: Cambridge University Press.

Jirák, M. 2015. “Uniform Change Point Tests in High Dimension.” *Annals of Statistics* 43: 2451–2483. http://dx.doi.org/10.1214/15-AOS1347.

Keele, L., and N. J. Kelly. 2006. “Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables.” *Political Analysis* 14 (2): 186–205. https://doi.org/10.1093/pan/mpj006.

Kim, Y., H. Choi, and H.-S. Oh. 2008. “Smoothly Clipped Absolute Deviation on High Dimensions.” *Journal of the American Statistical Association* 103 (484): 1665–1673. https://doi.org/10.1198/016214508000001066.

Knight, K., and W. Fu. 2000. “Asymptotics for Lasso-Type Estimators.” *Annals of Statistics* 28 (5): 1356–1378. https://www.jstor.org/stable/2674097.

Ratnasingam, S., and W. Ning. 2021. “Monitoring Sequential Structural Changes in Penalized High-Dimensional Linear Models.” *Sequential Analysis* 40 (3): 381–404. https://doi.org/10.1080/07474946.2021.1940500.

Stock, J. H., and M. W. Watson. 2003. *Introduction to Econometrics*. New York, NY: Prentice Hall.

Tibshirani, R. 1996. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58 (1): 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Verbeek, M. 2008. “A Guide to Modern Econometrics.” *Wiley Series in Probability and Statistics*. 3rd ed. Hoboken, NJ: John Wiley & Sons.

Wang, H., R. Li, and C. L. Tsai. 2007. “Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method.” *Biometrika* 94 (3): 553–568. https://doi.org/10.1093/biomet/asm053.

Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Zhao, P., and B. Yu. 2006. “On Model Selection Consistency of Lasso.” *Journal of Machine Learning Research* 7: 2541–2563. https://jmlr.org/papers/volume7/zhao06a/zhao06a.pdf.