Received: 8 May 2022 | Revised: 26 August 2023 | Accepted: 11 September 2023

DOI: 10.1111/stan.12323

ORIGINAL ARTICLE

WILEY

Testing conditional independence in casual inference for time series data

Zongwu Cai[1] | **Ying Fang[2,3,4]** | **Ming Lin[2,3,4]** | **Shengfang Tang[5]**

[1]Department of Economics, University of Kansas, Lawrence, Kansas, USA

[2]Wang Yanan Institute for Studies in Economics and Fujian Key Laboratory of Statistical Sciences, Xiamen University, Xiamen, Fujian, China

[3]Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, Fujian, China

[4]Laboratory of Digital Finance, Xiamen University, Xiamen, Fujian, China

[5]School of Finance, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China

Correspondence

Shengfang Tang, School of Finance, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, 330013, China.

Email: tangshengfang103@163.com

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 72301119, 72033008

In this paper, we propose a new procedure to test conditional independence assumption in studying casual inference for time series data. The conditional independence assumption is transformed to a nonparametric conditional moment test with the help of auxiliary variables which are allowed to affect policy choice but the dependence can be fully captured by potential outcomes and observable controls. When the policy choice is binary, a nonparametric statistic test is developed further for testing the conditional independence assumption conditional on policy propensity score. Under some regular conditions, we show that the proposed test statistics are asymptotically normal under the null hypotheses for time series data. In addition, the performances of the proposed methods are illustrated through Monte Carlo simulations and a real example considered in Angrist and Kuersteiner (2011).

KEYWORDS

casual inference, economic policy evaluation, moment test, nonparametric estimation, treatment effect

1 | INTRODUCTION

One of the most important tasks in empirical economics is to quantitatively assess the causal effect of economic or social policies using plausibly exogenous policy variation. Recently, there have been some statistical methods proposed, see, for example, the papers by Angrist and Kuersteiner (2011), Jordà and Taylor (2016), and Angrist, Jordà and Kuersteiner (2018), and the survey paper by Liu, Cai, Fang, and Lin (2020), for studying economic policy evaluation, which is an extension of the Rubin causal model in Rubin (1974) for identically independent distributed (iid) setting to a time series context. Sidestepping the impossible mission of accurately modeling an entire economy, the new framework only needs to estimate a policy propensity score, the probability of how a policy is determined conditional on a set of current information at hand, in a very flexible way. With an assumption of conditional independence (CI), which means that, conditional on a set of controls possibly including past policy choices and lagged outcomes, the current policy choice is independent of potential outcomes, the policy causal effects defined by the so called dynamic treatment effect then can be identified using inverse probability weighting (IPW) estimators.

The CI assumption plays a key role in identifying the causal effect in evaluating economic policies. The conditional independence between current policy choice and potential outcomes implies that, after appropriately considering a set of observable controls, the remaining policy variation can be regarded as exogenous shocks and then can be used to identify policy effects. Exploring policy variation at hand is actually a natural strategy particularly in a situation lack of conducting natural experiments or quasi-experiments, although conditional independence seems to be a strong assumption. Therefore, in this paper, this gap is filled up by proposing a statistical test method for testing the CI assumption in a data-driven approach.

Our test procedure relies on the existence of auxiliary variables, similar to instrumental variables in linear regression, which allow to affect policy choices but the linkage from auxiliary variables to policy choices can be fully captured by potential outcomes and observed controls. With the availability of such auxiliary variables, the conditional independence assumption can be easily tested using conditional moment tests in a nonparametric way. However, such a test procedure encounters a serious implementation problem when the dimension of controls is large due to the curse of dimensionality in nonparametric testing. When the policy choice is only a binary variable, we further develop a test procedure conditional on policy propensity score to avoid the aforementioned dimensionality problem. Actually, the assumption on the existence of auxiliary variables has been widely adopted in statistics literature such as Zhao and Shao (2015) and Breunig (2019) for dealing with missing observation difficulties and Hu and Schennach (2008) for tackling measurement error problems. In other scenarios, Fang, Tang, Cai, and Lin (2020) provided some clues on how to find auxiliary variables in real data among all observable controls.

To the best of our knowledge, our test procedure is the first attempt in the statistics literature to test CI for economic policy evaluation in time series context. Indeed, the previous literature focused on testing CI for policy evaluation only for cross-sectional data. For example, by using a binary instrumental variable, Donald, Hsu, and Lieli (2014) proposed a Durbin–Wu–Hausman type statistic, a popular approach in the econometrics literature, to test the conditional mean independence which is a central identification assumption for economic policy evaluation. With a presumption that the error terms in both the outcome equation and the selection equation are symmetrically distributed, Chen, Ji, Zhou, and Zhu (2018) proposed a Kolmogorov–Smirnov type statistic to test conditional mean independence by comparing two estimators obtained with/without conditional mean independence. Our paper is closely related to the work in Fang et al. (2020) which tested CI by using auxiliary variables for iid data. However, different from Fang et al. (2020), the current paper is on testing CI for economic policy evaluation in a time series context, and furthermore, our paper develops a test for conditional independence conditional on policy propensity score, which is novel in statistics literature.

The rest of the paper is organized as follows. Section 2 first provides a brief introduction of the framework of economic policy evaluation for time series, and then develop statistics based on auxiliary variables to test conditional independence conditional on policy propensity score. Large sample properties of proposed test statistics are also established in the same section. Section 3 reports results from Monte Carlo simulations and Section 4 applies our test procedure to revisiting the famous real example studied by Angrist and Kuersteiner (2011). Section 5 concludes the paper. All technical proofs are gathered in the Appendix.

2 | TESTING PROCEDURES

2.1 | Framework

Suppose that the observed time series process is denoted by (Y_t, M_t, B_t), where Y_t is a vector of outcome variables, $B_t \in \{b_0, \dots, b_J\}$ is a policy choice, and M_t is a vector of observed controls including exogenous and (lagged) endogenous variables. Denote υ as the policy regime, which represents the state of the economy at some time and takes values in a parameter space, denoted by Υ. It is assumed that the policy choice B_t is determined by both observed and unobserved variables according to $B_t = B(M_t, \upsilon, \varepsilon_t)$, where ε_t is the idiosyncratic information or taste variables. In this paper, to describe the potential outcome, we assume that the policy changes while the policy regime remains unchanged. As in Angrist and Kuersteiner (2011) and Angrist et al. (2018), the following definition of potential outcomes in a time series setup is adopted.

Definition 1. Given t, ν and υ, potential outcomes $\{\text{Y}_{t,\nu}^{\upsilon}(b); b \in \mathcal{B}\}$ are defined as the set of values the observed outcome variable $\text{Y}_{t+\nu}$ would take on if the policy choice $B_t = b$, with $b \in \mathcal{B} = \{b_0, \dots, b_j, \dots, b_J\}$.

From the above definition, it is easy to observe that the vector of potential outcomes includes the observed outcome, $\text{Y}_{t+\nu} = \text{Y}_{t,\nu}^{\upsilon}(b)$, and the counterfactual outcome describing the consequences of policy choices not taken. Without loss of generality, it is assumed that the observed outcome variable $\text{Y}_{t+\nu}$ is of dimension-one throughout the remaining paper. Define $\mathcal{Y}_{t,I} = (\text{Y}_{t+1}, \dots, \text{Y}_{t+I})'$ and let $\mathcal{Y}_{t,I}^{\upsilon}(b) = \left(\text{Y}_{t,1}^{\upsilon}(b), \dots, \text{Y}_{t,I}^{\upsilon}(b)\right)'$ be the vector of potential outcomes up to horizon I. Then, the observed outcomes have the following relationship with the potential outcomes

$$\mathcal{Y}_{t,I} = \sum_{b \in \mathcal{B}} \mathcal{Y}_{t,I}^{\upsilon}(b) I\{B_t = b\},$$

where $I\{\cdot\}$ is an indicator function. Under this framework, similar to the case of cross-sectional data setting, one can define a dynamic average response to the policy b_j relative to the benchmark policy b_0 by contrasting two potential outcomes,

$$\lambda_j = E\left[\mathcal{Y}_{t,I}^{\upsilon}(b_j) - \mathcal{Y}_{t,I}^{\upsilon}(b_0)\right].$$

Consequently, a collection of all possible policy effects is given by $\lambda = (\lambda_1, \dots, \lambda_J)'$. It should be noted that potential outcomes for counterfactual policy choices can not be observed, so that the average effects λ_j for $1 \le j \le J$, cannot be estimated directly. To identify the parameters of interest λ_j for $1 \le j \le J$, similar to the Rubin causal model for the cross sectional data, see, for example, Rubin (1974) for details, by following Angrist and Kuersteiner (2011), Angrist et al. (2018), and Liu et al. (2020), we adopt the following conditional independence assumption, which is the so-called *conditional independence* (CI) in statistics literature.

Assumption 1. Conditional independence:

$$\mathrm{Y}_{t,1}^{\upsilon}(b_j), \mathrm{Y}_{t,2}^{\upsilon}(b_j), \dots \perp\!\!\!\perp B_t \mid M_t \text{ for all } b_j \text{ with } \upsilon \text{ fixed for } \upsilon \in \Upsilon,$$

where M_t is a vector of predetermined variables that predict B_t and $\perp\!\!\!\perp$ denotes the statistical independence.

The above assumption is commonly imposed, for example, by Angrist and Kuersteiner (2011) and Angrist et al. (2018) to identify the causal effects of monetary policy shocks and by Jordà and Taylor (2016) to estimate the average treatment effect of fiscal policy, which are under time series framework. It is worth noting that this assumption focuses on variation in policy interventions while holding the policy regime fixed, after conditioning on predetermined variables M_t. In addition, this assumption implies that, given some appropriate predetermined variables M_t, potential outcomes are independent of the policy variables. Based on this assumption, it is easy to see that the parameters of interest λ_j can be identified by

$$\lambda_j = E\left\{E\left[\mathcal{Y}_{t,I}^{\upsilon}(b_j) - \mathcal{Y}_{t,I}^{\upsilon}(b_0)\Big|M_t\right]\right\} = E\left\{E\left[\mathcal{Y}_{t,I}\Big|B_t = b_j, M_t\right] - E\left[\mathcal{Y}_{t,I}\Big|B_t = b_0, M_t\right]\right\},$$

for $1 \leq j \leq J$. Moreover, according to the identification results above, Angrist et al. (2018) proposed an IPW method to estimate λ_j.

2.2 | Testing statistic

It is important to note that the CI assumption given above may not hold in practice, for example, if there are some unobserved confounders that can affect both policy choice B_t and potential outcomes $\mathrm{Y}_{t,\nu}^{\upsilon}(b_j)$. In this case, the IPW estimate considered in Angrist et al. (2018) should be inconsistent in general and then the estimating results would result in misleading inferences for the parameters of interest. Therefore, it is urgent to suggest an approach which can be used to test whether the CI assumption is true or not. To this end, this paper develops a new procedure by using a set of auxiliary variables to test whether Assumption 1 holds under time series setting. A vector of valid auxiliary variables $A_t \in \mathbb{R}^p$ is defined by the following assumption.

Assumption 2. Suppose that there is a vector of variables $A_t \in \mathbb{R}^p$ which are continuously distributed and correlated with the potential outcomes $\mathrm{Y}_{t,\nu}^{\upsilon}(b_j)$. Furthermore, the following condition is satisfied:

$$A_t \perp\!\!\!\perp B_t|(\mathrm{Y}_{t,\nu}^{\upsilon}(b_j), M_t) \text{ for all } \nu \geq 0 \text{ and for all } b_j \text{ with } \upsilon \text{ fixed for } \upsilon \in \Upsilon.$$

Remark 1. Note that this assumption is the time series version of assumption 2.1(i) in Fang et al. (2020), which requires that a policy intervention is mainly driven by potential outcomes and observed predetermined variables M_t. As mentioned in Section 1, choosing auxiliary variables is similar to finding instrumental variables in classical linear regression. But, unfortunately, as pointed out by Fang et al. (2020), there is no general guideline on how to select A_t and it might be done case by case. Of course, it would seem very challenging to find such variables based on a priori reasoning. This is because the potential outcomes $\mathrm{Y}_{t,\nu}^{\upsilon}(b_j)$ presumably depend on latent shocks that hit the economy between time t and $t + \nu$, so the conditioning sigma algebra includes a mix of past and future information. We conjecture that the testing power might depend on the choice of A_t. This issue needs a further investigation and it is warranted as a future research topic.

It is easy to show that under Assumption 2, Assumption 1 implies that $E(B_t|M_t, A_t) = E(B_t|M_t)$, which further implies that Assumption 1 can be tested by investigating whether A_t is significant for the mean of the policy choice B_t conditional on controls M_t. Formally, in order to test Assumption 1 being true or not, the following testing hypothesis is considered:

$$\mathcal{H}_0 : E(B_t|M_t, A_t) = E(B_t|M_t) a.e. \text{versus} \mathcal{H}_1 : P\{E(B_t|M_t, A_t) = E(B_t|M_t)\} < 1. \quad (1)$$

Similar to Fan and Li (1996) and Li (1999), a nonparametric conditional moment statistic is constructed to test the null hypothesis in (1). To this end, first, some additional notations are needed. Denote $G_t = (A_t', M_t')' \in \mathbb{R}^{p+q}$, where $M_t \in \mathbb{R}^q$ and $A_t \in \mathbb{R}^p$, and define $\alpha_t = B_t - E(B_t|A_t)$. Then, it is important to note that $U = E[\alpha_t E(\alpha_t|G_t)] = E\left\{[E(\alpha_t|G_t)]^2\right\} \geq 0$ and the equality holds if and only if $\mathcal{H}_0$ given in (1) is true. Therefore, U can be used to construct test statistic for consistent testing $\mathcal{H}_0$ in (1). Furthermore, to keep away from the random denominator problem, following the standard method we choose using a density weighted version of U as the basis of our test statistic. That is,

$$U^* = E\{\alpha_t f(M_t) E[\alpha_t f(M_t)|G_t] f_G(G_t)\} = E\{\mu_t E[\mu_t|G_t] f_G(G_t)\},$$

where $\mu_t = \alpha_t f(M_t)$, and $f_G(\cdot)$ and $f(\cdot)$ are the density functions of G_t and M_t, respectively. Hence, a kernel-based sample analogue of U^* is defined by

$$U_N^* = \frac{1}{N(N-1)a^{p+q}} \sum_{t=1}^{N} \sum_{s \neq t}^{N} \alpha_t f(M_t) \alpha_s f(M_s) L_{st} = \frac{1}{N(N-1)a^{p+q}} \sum_{t=1}^{N} \sum_{s \neq t}^{n} \mu_t \mu_s L_{st},$$

where $L_{st} = L((G_t - G_s)/a)$ is a kernel function and $a = a_N$ is the smoothing parameter. However, it should be noted that the residual α_t and the density function $f(M_t)$ can not be observed directly, so that to get a feasible test statistic, we should first estimate them nonparametrically. To be specific, we consider the following estimator for $E(B_t|M_t)$, given by

$$\widehat{E}(B_t|M_t) = \frac{1}{(N-1)a_1^q} \sum_{s \neq t}^{N} \mathrm{L}^{(1)}\left(\frac{M_s - M_t}{a_1}\right) B_s / \widehat{f}_{M_t},$$

where

$$\widehat{f}_{M_t} = \frac{1}{(N-1)a_1^q} \sum_{s \neq t}^{N} \mathrm{L}^{(1)}\left(\frac{M_s - M_t}{a_1}\right),$$

is the estimator of $f(M_t)$, $\mathrm{L}^{(1)}(\cdot)$ is another kernel function and a_1 denotes the bandwidth parameter. Consequently, we can obtain the estimator $\widehat{\alpha}_t = B_t - \widehat{E}(B_t|M_t)$ for the residual α_t. Therefore, a feasible test statistic is obtained by replacing $\mu_t = \alpha_t f(M_t)$ by its kernel estimator $\widehat{\alpha}_t \widehat{f}_{M_t}$:

$$U_N = \frac{1}{N(N-1)a^{p+q}} \sum_{t=1}^{N} \sum_{s \neq t}^{N} \left(\widehat{\alpha}_t \widehat{f}_{M_t}\right) \left(\widehat{\alpha}_s \widehat{f}_{M_s}\right) L_{st}.$$

Now, it is ready to investigate the large sample properties of the proposed test statistic U_N. To derive the asymptotic distribution of the test statistic U_N under $\mathcal{H}_0$, the following assumptions are needed, with the class of kernel functions Υ_λ and the class of functions $\mathfrak{U}_\mu^\alpha$; see, for example, their definitions in Robinson (1988), Fan and Li (1996), and Li (1999), or the Appendix for details.

Assumption 3.

(i) Suppose that the process $\{B_t, M_t, A_t\}_{t=1}^N$ is strictly stationary and absolutely regular process with the mixing coefficient $\beta(t) \le C_\beta \rho^t$ defined by

$$\beta(t) = \sup_{s\in\mathbb{N}} E\left[\sup_{A\in\mathcal{F}_{s+t}^\infty}\left\{\left|P(A|\mathcal{F}_{-\infty}^s) - P(A)\right|\right\}\right],$$

for all $s,\ t \ge 1$, where $0 < C_\beta < \infty$ and $0 < \rho < 1$ are constants, and $\mathcal{F}_i^j$ denotes the σ-field generated by $\{(M_t, B_t, A_t) : i \le t \le j\}$.

(ii) Suppose that the residual $\alpha_t = B_t - E(B_t|M_t)$ satisfies $E[\alpha_t|\ \Omega_{t-1}] = 0$ for all $t \ge 1$, where $\Omega_t = \sigma\{(X_{s+1}, B_s) : s \le t\}$ is the σ-field generated by $\{(M_{s+1}, B_s) : s \le t\}$.

(iii) In addition, it is assumed that $E[|\alpha_t^{4+\epsilon}|] < \infty$ and $E\left[\left|\alpha_{t_1}^{i_1}\alpha_{t_2}^{i_2}\cdots\alpha_{t_\nu}^{i_\nu}\right|^{1+\eta}\right] < \infty$ for some arbitrarily small $\epsilon > 0$ and $\eta > 0$, where $2 \le \nu \le 4$ is an integer, $0 \le i_j \le 4$ and $\sum_{j=1}^\nu i_j \le 8$.

Assumption 4.

(i) Denote $\sigma^2(g) = E[\alpha_t^2|G_t = g]$ and $\mu_4(g) = E[\alpha_t^4|G_t = g]$. It is assumed that $\sigma^2(g)$ and $\mu_4(g)$ satisfy some Lipschitz conditions: $|\sigma^2(u+v) - \sigma^2(u)| \le \Gamma(u)\|v\|$ and $|\mu_4(u+v) - \mu_4(u)| \le \Gamma(u)\|v\|$ with $E\left[|\Gamma(G_t)|^{2+\iota}\right] < \infty$ for some small $\iota > 0$, where $\|\cdot\|$ denotes the Euclidean norm.

(ii) Let $f_{\tau_1,\tau_2,\dots,\tau_\nu}(\cdot)$ be the joint probability density of $(G_{1+\tau_1}, \dots, G_{1+\tau_\nu})$ $(1 \le \nu \le 4)$. Suppose that $f_{\tau_1,\tau_2,\dots,\tau_\nu}(\cdot)$ exists, is bounded and satisfies the following Lipschitz condition: $|f_{\tau_1,\tau_2,\dots,\tau_\nu}(g_1+v_1, \dots, g_\nu+v_\nu) - f_{\tau_1,\tau_2,\dots,\tau_\nu}(g_1, \dots, g_\nu)| \le \Lambda_{\tau_1,\tau_2,\dots,\tau_\nu}(g_1, \dots, g_\nu)\|v\|$, where $\Lambda_{\tau_1,\tau_2,\dots,\tau_\nu}(g_1, \dots, g_\nu)$ is integrable and satisfies the following conditions

$$\int \Lambda_{\tau_1,\tau_2,\dots,\tau_\nu}(g_1, \dots, g_\nu)\|g\|^{2\varsigma} dg < \Xi < \infty,$$

and

$$\int \Lambda_{\tau_1,\tau_2,\dots,\tau_\nu}(g_1, \dots, g_\nu) f_{\tau_1,\tau_2,\dots,\tau_\nu}(g_1, \dots, g_\nu) dg < \Xi < \infty,$$

for some $\varsigma > 1$ and constant $\Xi > 0$.

Assumption 5.

(i) The density functions $f(u)$ and $f_G(g)$ of M_t and G_t satisfy $f(u) \in \mathfrak{U}_\mu^\infty, f_G(g) \in \mathfrak{U}_\mu^\infty$, respectively, and $\gamma(u) = E(B_t|M_t = u) \in \mathfrak{U}_\mu^{4+\varepsilon}$ for some integer $\mu \ge 2$ and small $\varepsilon > 0$, and also $f_G(g)$ is bounded.

(ii) $f(u)$, $f_G(g)$ and $\gamma(u)$ all satisfy some Lipschitz conditions: $|\delta(u+v)-\delta(u)| \leq \Lambda(u)\|v\|$, where $\Lambda(u)$ has finite $(2+\eta^*)$th moment for some small $\eta^* > 2$.

Assumption 6.

(i) The product kernel is used for both $L(\cdot)$ and $\mathtt{L}^{(1)}(\cdot)$. Define $l(\cdot)$ and $\mathtt{l}^{(1)}(\cdot)$ to be their corresponding univariate kernel, then $\mathtt{l}^{(1)}(\cdot) \in \Upsilon_\mu$, $l(\cdot)$ is nonnegative and $l(\cdot) \in \Upsilon_2$.

(ii) As $N \to \infty$, $a_1 \to 0$, $a = O(N^{-\overline{\alpha}})$ for some $0 < \overline{\alpha} < 7/(8(p+q))$. In addition, $a^{p+q}/a_1^{2q} = o(1)$ and $Na^{(p+q)/2}a_1^{2\mu} = o(1)$.

These assumptions are commonly imposed in nonparametric literature; see, for example, Li (1999). Under Assumptions 2–6, the asymptotic properties of the test statistic U_N can be established and formally stated in the following theorem with the detailed proof given in the Appendix.

Theorem 1. *Suppose that Assumptions 2–6 are satisfied. Then,*

(1) Under $\mathcal{H}_0$, $\widetilde{U}_N = Na^{(p+q)/2}U_N/\sqrt{2}\widehat{\sigma}_U \to \mathcal{N}(0,1)$ in distribution as $N \to \infty$, where

$$\widehat{\sigma}_U^2 = \frac{1}{N(N-1)a^{p+q}} \sum_{t=1}^{N} \sum_{s \neq t} \widehat{\alpha}_t^2 \widehat{f}_{M_t}^2 \widehat{\alpha}_s^2 \widehat{f}_{M_s}^2 L_{ts}^2,$$

is a consistent estimator of σ_U^2 given by

$$\sigma_U^2 = E\left[f^2(M_t) f_G(G_t) \sigma^4(G_t)\right] \left(\int L^2(u)du\right).$$

(2) Under $\mathcal{H}_1$, $P(\widetilde{U}_N > O_N) \to 1$ for any nonstochastic sequence $O_N = o(Na^{(p+q)/2})$.

Theorem 1 implies that the test statistic $\widetilde{U}_N$ has the asymptotic standard normal distribution under the null hypothesis. Based on Theorem 1(1), $\mathcal{H}_0$ is rejected at significance level α_0 if $\widetilde{U}_N > Z_{\alpha_0}$, where Z_{α_0} is the upper α_0-percentile of the standard normal distribution. It is important to note that in order to use the proposed test $\widetilde{U}_N$ in practice, we need to choose the auxiliary variables A_t that satisfy Assumption 2. However, if the auxiliary variables A_t chosen by practitioners do not satisfy Assumption 2, the proposed test tends to reject CI even though CI is true.[1]

Based on Monte Carlo simulation studies by Li (1999) and Lavergne and Vuong (2000), it concludes that there exists substantial finite sample bias for the normal approximation. Moreover, the test statistic U_N depends on two sets of smoothing parameters a_1 and a, and might be sensitive to the choice of the smoothing parameters. To overcome these difficulties, similar to Lavergne and Vuong (2000), a modified version of the test U_N is adopted, which should have a better finite sample performance than the test U_N, as argued in Lavergne and Vuong (2000). Indeed, Li (1999) adopted the same idea to suggesting a new modified test and further showed that the modified test has the same asymptotic properties as U_N. Alternatively, one also can use Bootstrapping method to better approximate the null distribution of U_N; see Li and Racine (2007) for further details for Bootstrapping method. In this paper, we adopt simply the idea of Lavergne and Vuong (2000) and

Li (1999) to suggest a modified test statistic. To this end, by substituting $\widehat{E}(B_t|M_t)$ and $\widehat{f}_{M_t}$ into the expression of U_N and doing a simplification, we can obtain

$$U'_N = \frac{1}{N(N-1)(N-2)(N-3)} \left[N(N-1)^3 U_N - N(N-1)(N-2)\Gamma_{N,1} - 2N(N-1)(N-2)\Gamma_{N,2}\right],$$

where

$$\Gamma_{N,1} = \frac{1}{N(N-1)(N-2)a^{p+q}a_1^{2q}} \sum_{t=1}^{N}\sum_{s\neq t}\sum_{l\neq t, l\neq s} (B_t - B_l)(B_s - B_l) \times \mathrm{L}^{(1)}\left(\frac{M_t - M_l}{a_1}\right)\mathrm{L}^{(1)}\left(\frac{M_s - M_l}{a_1}\right) L\left(\frac{G_t - G_s}{a}\right),$$

and

$$\Gamma_{N,2} = \frac{1}{N(N-1)(N-2)a^{p+q}a_1^{2q}} \sum_{t=1}^{N}\sum_{s\neq t}\sum_{l\neq t, l\neq s} (B_t - B_s)(B_s - B_l) \times \mathrm{L}^{(1)}\left(\frac{M_t - M_s}{a_1}\right)\mathrm{L}^{(1)}\left(\frac{M_s - M_l}{a_1}\right) L\left(\frac{G_t - G_s}{a}\right).$$

The following theorem shows that both U'_N and U_N share the exactly same asymptotic distribution with the detailed proof relegated to the Appendix.

Theorem 2. *Under Assumptions 2–6, one has*

(1) *Under* $\mathcal{H}_0$, $\widetilde{U}'_N = Na^{(p+q)/2}U'_N/\sqrt{2}\widehat{\sigma}_U \to \mathcal{N}(0,1)$ *in distribution as* $N \to \infty$, *where* $\widehat{\sigma}_U^2$ *is the same as defined in Theorem 1.*

(2) *Under* $\mathcal{H}_1$, $P(\widetilde{U}'_N > O_N) \to 1$ *for any non-stochastic sequence* $O_N = o(Na^{(p+q)/2})$.

2.3 | An extension

In many applications, the policy intervention variable B_t may be a binary variable but the dimension of the predetermined variables M_t may be moderate or high relative to the sample size. Under these cases, the proposed test statistics $\widetilde{U}_N$ in Theorem 1 and $\widetilde{U}'_N$ in Theorem 2 can not be used in practice because of the curse of dimensionality. Therefore, it is urgent to suggest a test statistic which applies to the case where the dimension of the predetermined variables M_t is in moderate or high dimension relative to the sample size.

Note that when B_t is a binary variable, one can show that the CI assumption displayed in Assumption 1 implies that

$$\mathrm{Y}^{\upsilon}_{t,\nu}(b_j) \perp\!\!\!\perp B_t|p(M_t) \text{ for all } \nu \geq 0 \text{ and for all } b_j \text{ with } \upsilon \text{ fixed for } \upsilon \in \Upsilon, \tag{2}$$

where $p(M_t) = P(B_t = 1|M_t)$ is the policy propensity score function. Hence, based on (2), it is clear that all biases due to observable controls can be removed by conditioning solely on the policy propensity score function. Moreover, we have the following lemma with the detailed proof given in the Appendix.

Lemma 1. *Under Assumption 2, (2) implies that*

$$E\left(B_t \,|\, p(M_t), A_t\right) = E\left(B_t \,|\, p(M_t)\right) = p(M_t). \tag{3}$$

From Lemma 1, to test whether the CI assumption holds or not, one just needs to consider testing whether A_t is significant for the mean of the policy choice B_t given the policy propensity score function $p(M_t)$ instead of the full covariates M_t. Therefore, in this section, the following testing hypothesis is investigated

$$\mathcal{H}_0 : E(B_t|p(M_t), A_t) = E(B_t|p(M_t)) \ \ a.\ e.\ \text{ versus } \ \mathcal{H}_1 : P\left\{E(B_t|p(M_t), A_t) = E(B_t|p(M_t))\right\} < 1. \tag{4}$$

Denote $P_t = p(M_t)$ and $R_t = (P_t, A_t')' \in \mathbb{R}^{\kappa}$, where $\kappa = p + 1$. Define $\varepsilon_t = B_t - P_t$. Then, the hypothesis testing problem formulated in (4) can be rewritten as

$$\mathcal{H}_0 : E(\varepsilon_t|R_t) = 0 \ a.e. \text{ versus } \ \mathcal{H}_1 : P\left\{E(\varepsilon_t|R_t) = 0\right\} < 1. \tag{5}$$

Similar to the discussion aforementioned, to test $\mathcal{H}_0$ in (5) holds or not, the following test statistic is suggested

$$V_N^{**} = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t E(\varepsilon_t|R_t) f_R(R_t),$$

where $f_R(\cdot)$ is the density function of R_t. Because ε_t and R_t cannot be observed directly, one first needs to estimate them to obtain a feasible test statistic. To overcome the problem of the curse of dimensionality, a flexible parametric model for the unknown policy propensity score function $p(M_t)$ is adopted, say assuming that

$$P(B_t = 1|M_t) = p(M_t; \theta),$$

where θ is an unknown parameter with dimension r. In addition, θ_0 is used to denote the true value of θ; that is, $p(M_t) = p(M_t; \theta_0)$. Since B_t is a binary variable, a logit or probit model is appropriate for the policy propensity score function $p(M_t)$ so that θ can be estimated by using the maximum likelihood method. Denote $\widehat{\theta}_N$ as the estimator of θ and $\widehat{P}_t = p(M_t; \widehat{\theta}_N)$ as the estimator of $p(M_t)$. Define $\widehat{\varepsilon}_t = B_t - \widehat{P}_t$ and

$$\begin{aligned}
\widehat{f}_{R_t} &= \frac{1}{(N-1)\ell^{\kappa}} \sum_{s \neq t}^{N} L_1\left(\frac{\widehat{P}_s - \widehat{P}_t}{\ell}\right) L_2\left(\frac{A_s - A_t}{\ell}\right) \\
&= \frac{1}{(N-1)\ell^{\kappa}} \sum_{s \neq t}^{N} L_1\left(\frac{p(M_s; \widehat{\theta}_N) - p(M_t; \widehat{\theta}_N)}{\ell}\right) L_2\left(\frac{A_s - A_t}{\ell}\right),
\end{aligned}$$

where $L_1(\cdot)$ and $L_2(\cdot)$ are the kernel functions, and ℓ is the smoothing parameter. Therefore, an estimator of V_N^{**} is given by

$$V_N = \frac{1}{N(N-1)\ell^{\kappa}} \sum_{t=1}^{N} \sum_{s \neq t}^{N} \widehat{\varepsilon}_t \widehat{\varepsilon}_s L_1\left(\frac{\widehat{P}_s - \widehat{P}_t}{\ell}\right) L_2\left(\frac{A_s - A_t}{\ell}\right).$$

Next, to consider the large sample properties of the proposed test statistic V_N, the following assumptions are needed.

Assumption 7.

(i) Assume that the process $\{B_t, P_t, A_t\}_{t=1}^N$ is strictly stationary and absolutely regular process with the mixing coefficient $\beta(\tau) = O(\rho^\tau)$ $(0 < \rho < 1)$.
(ii) The residual $\varepsilon_t = B_t - P_t$ satisfies $E\left(\varepsilon_t | \, \Omega_{t-1}^*\right) = 0$ for all $t \geq 1$, where $\Omega_{t-1}^* = \sigma\{(P_{s+1}, B_s) : s \leq t\}$ is the σ-field generated by $\{(P_{s+1}, B_s) : s \leq t\}$.

Assumption 8. (i) Let $\widetilde{\sigma}^2(r) = E(\varepsilon_t^2 | R_t = r)$ and $\widetilde{\mu}_4(r) = E(\varepsilon_t^4 | R_t = r)$. Then $\widetilde{\sigma}^2(r)$ and $\widetilde{\mu}_4(r)$ satisfy some Lipschitz conditions: $|\widetilde{\sigma}^2(r+u) - \widetilde{\sigma}^2(r)| \leq D(r)\|u\|$ and $|\widetilde{\mu}_4(r+u) - \widetilde{\mu}_4(r)| \leq D(r)\|u\|$ with $E[|D(R)|^{\eta^{**}}] < \infty$ for some $\eta^{**} > 2$.

Assumption 9. Let $\varphi_{\tau_1,\tau_2,\ldots,\tau_v}(\cdot)$ be the joint probability density function of $(R_{1+\tau_1}, \ldots, R_{1+\tau_v})$ $(1 \leq l \leq 4)$. Then for all $(\tau_1, \ldots, \tau_v)$, $\varphi_{\tau_1,\tau_2,\ldots,\tau_v}(\cdot)$ exists and satisfies a Lipschitz condition: $|\varphi_{\tau_1,\tau_2,\ldots,\tau_v}(r_1 + v_1, \ldots, r_v + v_v) - \varphi_{\tau_1,\tau_2,\ldots,\tau_v}(r_1, \ldots, r_v)| \leq D_{\tau_1,\tau_2,\ldots,\tau_v}(r_1, \ldots, r_v)\|v\|$, where $D_{\tau_1,\tau_2,\ldots,\tau_v}(\cdot)$ is integrable and satisfies the conditions that $\int D_{\tau_1,\tau_2,\ldots,\tau_v}(r_1, \ldots, r_v)\|r\|^{2\iota} dr < \Xi < \infty$ and $\int D_{\tau_1,\tau_2,\ldots,\tau_v}(r_1, \ldots, r_v) \varphi_{\tau_1,\tau_2,\ldots,\tau_v}(r_1, \ldots, r_v) dr < \Xi < \infty$ for some $\iota > 1$ and constant $\Xi > 0$.

Assumption 10. (i) $\nabla p(M; \cdot)$ and $\nabla^2 p(M; \cdot)$ are continuous in M and dominated by a function (say $G_p(M)$) with finite second moments, where $\nabla p(M; \cdot)$ and $\nabla^2 p(M; \cdot)$ are $r \times 1$ vector of first-order partial derivatives and $r \times r$ matrix of second partial derivatives of $p(M; \theta)$ with respect to θ, respectively. (ii) $E[\nabla p(M; \theta)\nabla' p(M; \theta)]$ is nonsingular for θ in a neighborhood of the true value θ_0.

Assumption 11. $L_1(\cdot)$ and $L_2(\cdot)$ are nonnegative second order and symmetric kernel functions. Furthermore, $L_1(\cdot)$ has bounded first-order derivative.

Assumption 12. The smoothing parameter ℓ satisfies $\ell = O(N^{-\overline{\delta}})$ for some $0 < \overline{\delta} < 7/(8\kappa)$.

Assumption 13. The estimator $\widehat{\theta}_N$ satisfies $\|\widehat{\theta}_N - \theta_0\| = O_p(1/\sqrt{N})$. Also, it is assumed that $\sup_{m \in \mathcal{M}} \left| p(m; \widehat{\theta}_N) - p(m; \theta_0) \right| = O_p(1/\sqrt{N})$ holds, where $\mathcal{M}$ is the support of M_t.

These assumptions are standard in both parametric and nonparametric literatures. Assumption 7(i) requires that the underlying process $\{B_t, P_t, A_t\}_{t=1}^N$ is absolutely regular with a geometric decay rate and (ii) states that the residual ε_t is a martingale difference. Assumption 8 includes some smoothness conditions on the second and fourth conditional moment functions of ε_t and Assumption 9 contains some Lipschitz type conditions and moment conditions. Assumption 10(i) and (ii) are standard assumptions adopted in nonlinear regression models. Assumption 11 is a standard assumption on the kernel functions $L_1(\cdot)$ and $L_2(\cdot)$ and Assumption 12 is the condition on the smoothing parameter ℓ which is slightly stronger than the usual conditions of $\ell \to 0$ and $N\ell^\kappa \to \infty$. Assumption 13 is known to hold for standard parametric estimation methods under reasonably mild regularity conditions. Under Assumptions 7–13, the asymptotic properties of the test statistic V_N above can be derived, formally stated in the following theorem with the detailed proof given in the Appendix.

Theorem 3. *Under Assumptions 7–13, one has*

(1) Under $\mathcal{H}_0$, $\widetilde{V}_N = N\ell^{\kappa/2}V_N/\sqrt{2}\widehat{\sigma}_V \to \mathcal{N}(0,1)$ in distribution as $N \to \infty$, where

$$\widehat{\sigma}_V^2 = \frac{1}{N(N-1)\ell^{\kappa}}\sum_{t=1}^{N}\sum_{s\neq t}\widehat{\varepsilon}_t^2\widehat{\varepsilon}_s^2 L_1^2\left(\frac{\widehat{P}_s-\widehat{P}_t}{\ell}\right)L_2^2\left(\frac{A_s-A_t}{\ell}\right),$$

is a consistent estimator of σ_V^2 given by

$$\sigma_V^2 = E\left(\widetilde{\sigma}^4(R_t)f_R(R_t)\right)\left(\int L_1^2(u)du\cdot\int L_2^2(v)dv\right),$$

with $\widetilde{\sigma}^2(r) = E(\varepsilon_t^2|R_t = r)$.

(2) Under $\mathcal{H}_1$, $P(\widetilde{V}_N > O_N) \to 1$ for any non-stochastic sequence $O_N = o(N\ell^{\kappa/2})$.

3 | MONTE CARLO STUDIES

This section is devoted to examining the finite-sample performance of the proposed tests U_N and V_N through two Monte Carlo experiments.

Example 1. In this example, we investigate the finite-sample performance of the test statistic U_N using the following data generating processes (DGP):

$$\mathcal{Y}_t = B_t\mathcal{Y}_t(1) + (1-B_t)\mathcal{Y}_t(0),$$

where $\mathcal{Y}_t(B_t) = \alpha A_t + 0.5\ \mathcal{Y}_{t-1} + 0.5\ B_t + \varphi_t$, $B_t = I\left\{\mu\left(\mathcal{Y}_t(1)+\mathcal{Y}_t(0)\right)/2 - \sqrt{1-\mu^2/2}\ \mathcal{Y}_{t-1} > \eta_t\right\}$, $A_t = \rho_A A_{t-1} + u_t$, $\varphi_t = \rho_\varphi\varphi_{t-1} + \upsilon_t$, u_t, υ_t and η_t are mutually independent processes with $u_t \overset{iid}{\sim} \mathcal{N}(0, 0.5^2)$, $\upsilon_t \overset{iid}{\sim} \mathcal{N}(0, 0.3^2)$ and $\eta_t \overset{iid}{\sim} unif(0,1)$. We consider $\rho_A \in \{0.2, 0.8\}$, $\rho_\varphi \in \{0.2, 0.8\}$ and $\alpha \in \{0.3, 0.9\}$. However, the constant $\mu \in [0,1]$ are allowed to vary in various setups. It can be checked that the auxiliary variable A_t always satisfies Assumption 2 in Section 2 regardless of any values of μ taking. The CI assumption is satisfied only when $\mu = 0$. Here the constant α is used to capture the correlation between the auxiliary variable A_t and the potential outcomes $\mathcal{Y}_t(1)$ and $\mathcal{Y}_t(0)$. In this example, the standard normal kernel functions are used for both $\mathrm{L}^{(1)}(\cdot)$ and $L(\cdot)$ and the bandwidths are set to be $a_1 = \widehat{\sigma}_M N^{-1/5}$, $a_M = c_1\cdot\widehat{\sigma}_M N^{-1/4}$ and $a_A = c_1\cdot\widehat{\sigma}_A N^{-1/4}$, where $\widehat{\sigma}_M$ and $\widehat{\sigma}_A$ are the sample *SD* s of $\{M_t\}_{t=1}^N$ with $M_t = \mathcal{Y}_{t-1}$ and $\{A_t\}_{t=1}^N$, respectively. To investigate the effects of different values of the bandwidths on the test U_N, we consider $c_1 = 0.5$, 1.0 and 2.0, respectively. To obtain the empirical test size, the simulation is repeated 1,000 times for each setting and then, the empirical size is estimated by just the empirical rejection rate of U_N based on the 1,000 replications. Finally, Table 1 reports the estimated sizes for all settings considered.

From Table 1, it can be observed that the finite sample performance of the test U_N is pretty good in different situations considered. In particular, the test U_N performs well in most cases considered as the sample size increases to 400. In addition, one also can observe from Table 1 that in all cases, both the autocorrelation coefficients (ρ_A and ρ_φ) and the correlation of the potential outcomes and the auxiliary variable captured by α have little effects on the size performance.

TABLE 1 Estimated sizes of U_N (nominal size $\alpha_0 = 5\%$).

Model		Empirical rejection probability of U_N with					
		$c_1 = 0.5, \rho_\varphi = 0.2$			$c_1 = 0.5, \rho_\varphi = 0.8$		
ρ_A	α	$N = 100$	$N = 200$	$N = 400$	$N = 100$	$N = 200$	$N = 400$
0.2	0.3	0.032	0.036	0.050	0.034	0.050	0.054
	0.9	0.036	0.063	0.051	0.041	0.047	0.058
0.8	0.3	0.038	0.046	0.050	0.054	0.043	0.052
	0.9	0.035	0.046	0.055	0.064	0.057	0.053
		$c_1 = 1.0, \rho_\varphi = 0.2$			$c_1 = 1.0, \rho_\varphi = 0.8$		
ρ_A	α	$N = 100$	$N = 200$	$N = 400$	$N = 100$	$N = 200$	$N = 400$
0.2	0.3	0.021	0.027	0.036	0.026	0.039	0.048
	0.9	0.022	0.034	0.040	0.030	0.036	0.044
0.8	0.3	0.022	0.032	0.039	0.030	0.034	0.045
	0.9	0.028	0.035	0.044	0.033	0.038	0.059
		$c_1 = 2.0, \rho_\varphi = 0.2$			$c_1 = 2.0, \rho_\varphi = 0.8$		
ρ_A	α	$N = 100$	$N = 200$	$N = 400$	$N = 100$	$N = 200$	$N = 400$
0.2	0.3	0.016	0.028	0.038	0.012	0.022	0.033
	0.9	0.018	0.022	0.036	0.014	0.028	0.037
0.8	0.3	0.017	0.024	0.035	0.018	0.030	0.035
	0.9	0.015	0.028	0.032	0.015	0.021	0.034

When the sample size is reasonably large, say $N = 400$, and the bandwidth is not too large, for example $c_1 = 0.5$ or $c_1 = 1.0$, the size performances are very well. However, the test is conservative when the bandwidth becomes too large.

Next, Figure 1 presents the estimated power curves of the U_N test for different bandwidths (the top panel for $c_1 = 0.5$, the middle panel for $c_1 = 1.0$ and the bottom panel for $c_1 = 2.0$). It can be observed from Figure 1 that the test U_N is quite powerful in detecting alternatives in all cases considered in general. Also as expected we observe that the powers of the test U_N increase sharply when both the value of μ and the sample size increase. It is also noticed from these figures that the larger is the value of bandwidth a, the larger is the power of the U_N test. This result can be explained by the fact that the test U_N diverges to infinite at the rate of $Na^{(p+q)/2}$ under $\mathcal{H}_1$. Therefore, a larger a value (in certain range) should lead to a more powerful test against fixed alternatives. However, this does not mean that one should always prefer a large value of a in practice, since there is a tradeoff between powers and sizes for different values of the bandwidth, as one can see in Table 1 for size performance with different choices of the bandwidth a.

Example 2. In this example, we examine the finite sample performance of the test statistic V_N. Similar to Example 1, we consider the following time series data generating process:

$$\mathcal{Y}_t = B_t \mathcal{Y}_t(1) + (1 - B_t)\mathcal{Y}_t(0),$$

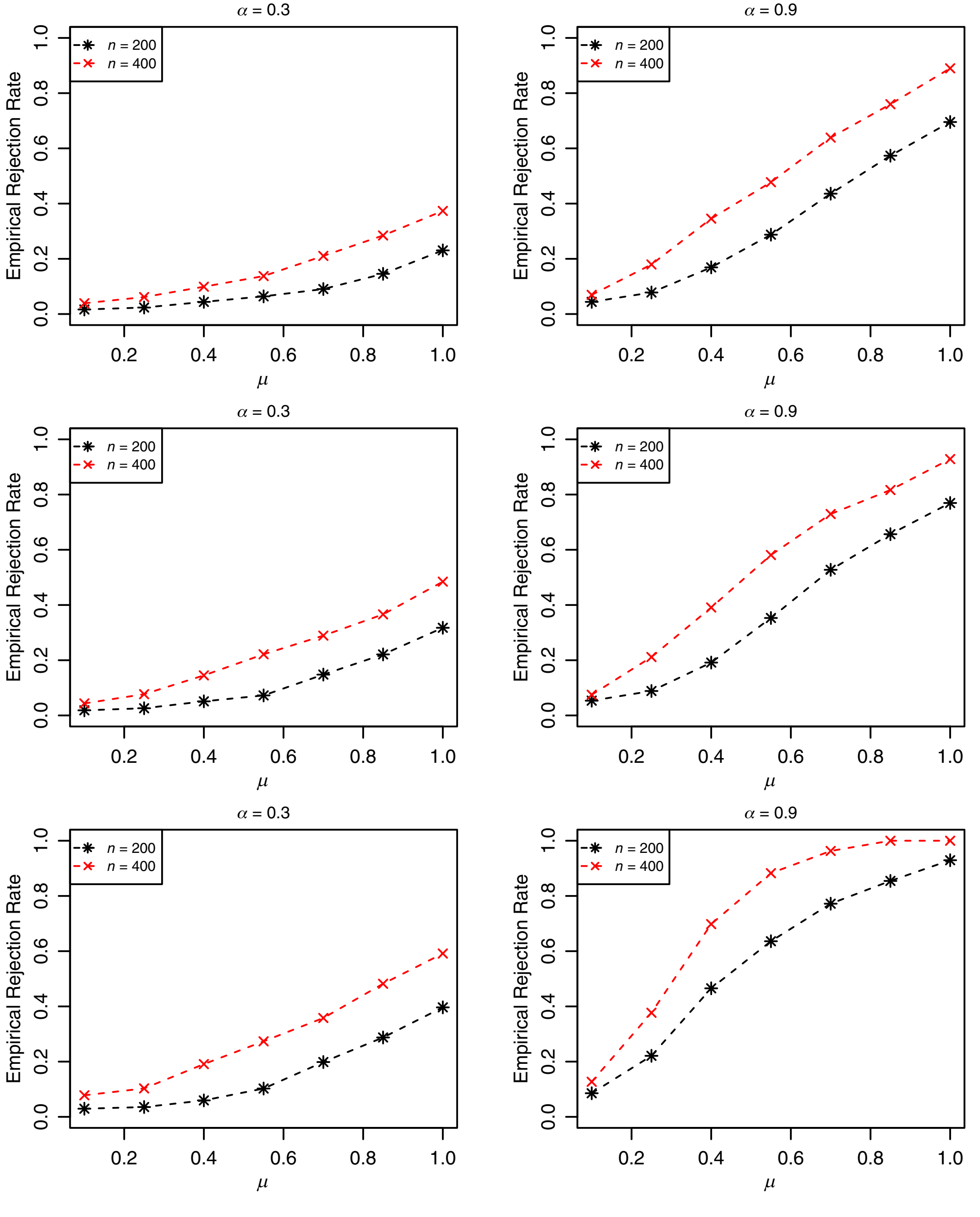

FIGURE 1 Estimated power curves for test statistic U_N with nominal size $\alpha_0 = 5\%$, $\rho_A = 0.5$, $\rho_\varphi = 0.5$ and different bandwidths (the top panel for $c_1 = 0.5$, the middle panel for $c_1 = 1.0$ and the bottom panel for $c_1 = 2.0$).

where $\mathcal{Y}_t(B_t) = \phi A_t + 0.5\ \mathcal{Y}_{t-1} + 0.5\ M_{1,t} + 1.6\ M_{2,t} + 2.5\ B_t + \epsilon_t$, $B_t = I\{\xi_t \leq Q_t\}$,

$$Q_t = \frac{\exp\left\{\nu\left(\mathcal{Y}_t(1) + \mathcal{Y}_t(0)\right)/2 - \sqrt{1-\nu^2/2}\mathcal{Y}_{t-1} + M_{1,t} + 2M_{2,t}\right\}}{1 + \exp\left\{\nu\left(\mathcal{Y}_t(1) + \mathcal{Y}_t(0)\right)/2 - \sqrt{1-\nu^2/2}\mathcal{Y}_{t-1} + M_{1,t} + 2M_{2,t}\right\}},$$

TABLE 2 Estimated sizes of V_N (nominal size $\alpha_0 = 5\%$).

Model		**Empirical rejection probability of V_N with**					
		$c_2 = 0.5, \varpi_\epsilon = 0.2$			$c_2 = 0.5, \varpi_\epsilon = 0.8$		
ϖ_A	ϕ	$N = 100$	$N = 200$	$N = 400$	$N = 100$	$N = 200$	$N = 400$
0.2	0.3	0.030	0.035	0.046	0.033	0.043	0.046
	0.9	0.030	0.040	0.045	0.030	0.036	0.044
0.8	0.3	0.042	0.047	0.053	0.035	0.039	0.045
	0.9	0.030	0.036	0.047	0.036	0.040	0.049
		$c_2 = 1.0, \varpi_\epsilon = 0.2$			$c_2 = 1.0, \varpi_\epsilon = 0.8$		
ϖ_A	ϕ	$N = 100$	$N = 200$	$N = 400$	$N = 100$	$N = 200$	$N = 400$
0.2	0.3	0.025	0.031	0.036	0.027	0.030	0.042
	0.9	0.024	0.027	0.033	0.031	0.034	0.040
0.8	0.3	0.021	0.032	0.041	0.031	0.034	0.042
	0.9	0.021	0.034	0.043	0.031	0.041	0.044
		$c_2 = 2.0, \varpi_\epsilon = 0.2$			$c_2 = 2.0, \varpi_\epsilon = 0.8$		
ϖ_A	ϕ	$N = 100$	$N = 200$	$N = 400$	$N = 100$	$N = 200$	$N = 400$
0.2	0.3	0.020	0.024	0.030	0.022	0.031	0.035
	0.9	0.024	0.029	0.032	0.023	0.027	0.038
0.8	0.3	0.024	0.031	0.037	0.021	0.025	0.031
	0.9	0.025	0.036	0.040	0.025	0.031	0.035

$$A_t = \varpi_A A_{t-1} + \psi_t,\ \epsilon_t = \varpi_\epsilon \epsilon_{t-1} + \vartheta_t,\ M_{1,t} = 0.5M_{1,t-1} + \omega_{1,t},\ M_{2,t} = 0.6M_{2,t-1} + \omega_{2,t},$$

and ψ_t, ϑ_t, ξ_t, $\omega_{1,t}$ and $\omega_{2,t}$ are mutually independent processes with $\psi_t \overset{iid}{\sim} \mathcal{N}(0, 0.5^2)$, $\vartheta_t \overset{iid}{\sim} \mathcal{N}(0, 0.3^2)$, $\omega_{1,t} \overset{iid}{\sim} \mathcal{N}(0, 0.4^2)$, $\omega_{2,t} \overset{iid}{\sim} \mathcal{N}(0, 0.4^2)$ and $\xi_t \overset{iid}{\sim}$ unif(0, 1). Again, we consider $\varpi_A \in \{0.2, 0.8\}$, $\varpi_\varepsilon \in \{0.2, 0.8\}$, $\phi \in \{0.3, 0.9\}$ and allow the constant $\nu \in [0, 1]$ to change in various simulation experiments. Also, it is easy to check that the auxiliary variable A_t always satisfies Assumption 2 in Section 2 regardless of any values of ν taking and the CI assumption is satisfied only when $\nu = 0$. In this example, the correlation between the auxiliary variable A_t and the potential outcomes $\mathcal{Y}_t(1)$ and $\mathcal{Y}_t(0)$ is captured by the constant ϕ. Similarly, to examine the effects of different values of the bandwidth ℓ on the test V_N, we consider $\ell = c_2 \cdot N^{-1/5}$ with $c_2 = 0.5, 1.0$ and 2.0, respectively. Finally, the simulation is repeated 1,000 times for each setting to estimate the empirical test size.

Table 2 reports the estimated sizes for V_N test for various cases, from which it can be seen that both the autocorrelation coefficients (ϖ_A and ϖ_ϵ) and the correlation of the potential outcomes and the auxiliary variable have little effects on the size performance of the test V_N. Again, when the sample size increases to $N = 400$ and the bandwidth is not too large, for example, $c_2 = 0.5$ or $c_2 = 1.0$, the size performances are reasonably well. Similar to Table 1, the test is conservative when the bandwidth becomes too large. Next, Figure 2 presents the estimated power curves for V_N test for different bandwidths (the top panel for $c_2 = 0.5$, the middle panel

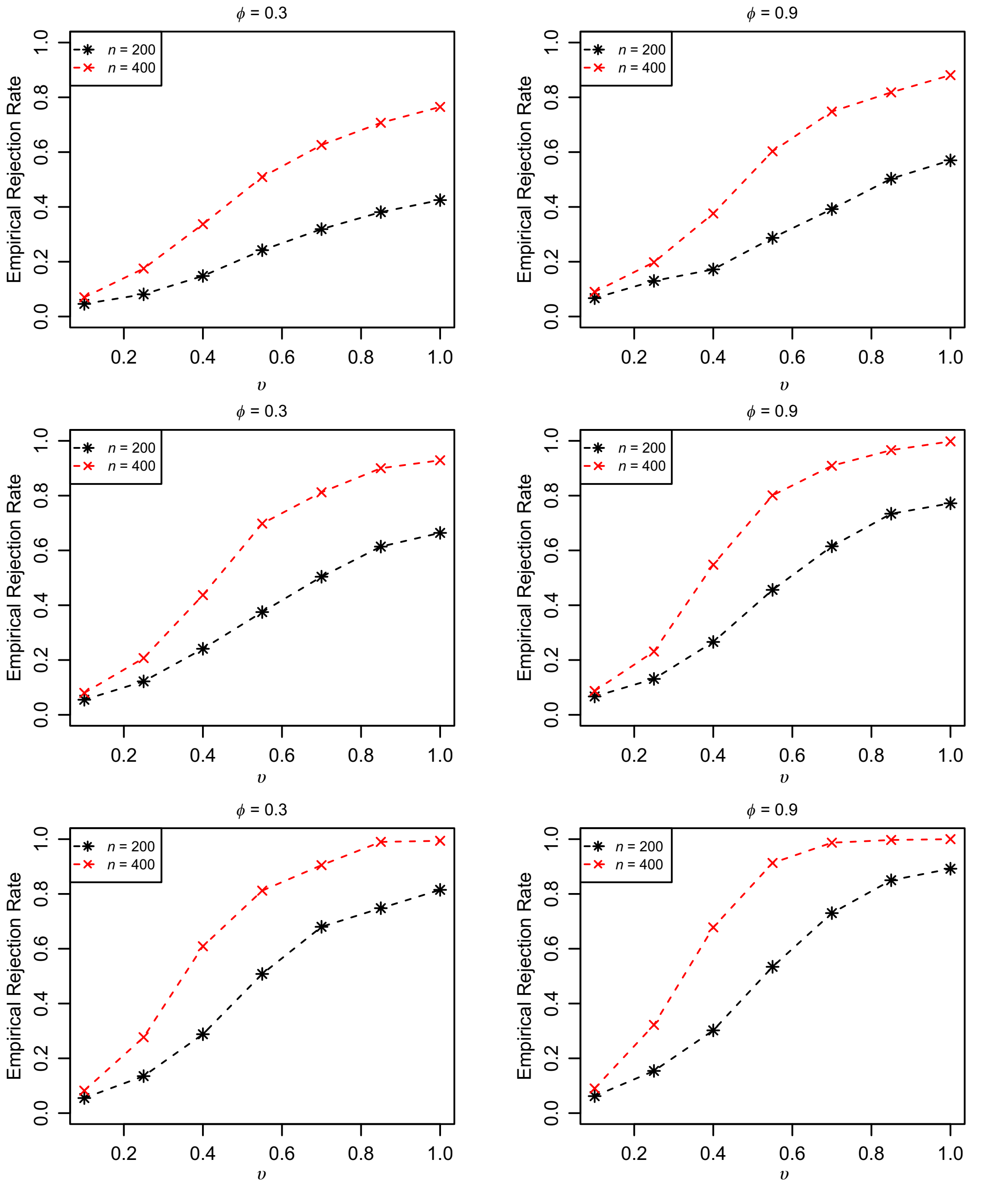

FIGURE 2 Estimated power curves for test statistic V_N with nominal size $\alpha_0 = 5\%$, $\varpi_A = 0.5$, $\varpi_\epsilon = 0.5$ and different bandwidths (the top panel for $c_2 = 0.5$, the middle panel for $c_2 = 1.0$ and the bottom panel for $c_2 = 2.0$).

for $c_2 = 1.0$ and the bottom panel for $c_2 = 2.0$). Clearly, the same pattern as seen in the first example can be observed. Specifically, when the value of ν and the sample size increase, the estimated powers also increase sharply. Moreover, one also can observe that the power performance relies on the correlation between the potential outcomes and the auxiliary variable A_t measured by ϕ.

4 | A REAL EXAMPLE

Identifying the causal connection between monetary policy and real economic variables is one of the most significant and extensively investigated questions in macroeconomics. To answer this question, researchers usually regress an outcome variable of interest on measures of monetary policy, at the same time, controlling for lagged outcomes and contemporaneous and lagged covariates. Consequently, the statistical significance of policy variables provides the evidence on the existence of the causal connection between monetary policy and outcome variable of interest. Two of the most leading empirical studies in this spirit are the papers by Sims (1972, 1980), in which the author discussed the conceptual as well as empirical problems in the money-income nexus.

To apply the procedure of the regression-based causality tests, researchers need a simple conditional independence assumption. That is, in the language of cross-sectional policy evaluation, researchers need to assume that given lagged outcomes and a suitable set of control variables, policy variables are "as good as randomly assigned," so that conditional effects have a causal interpretation. While the conditional independence assumption is a strong assumption, it is the foundation of beginning empirical work, at least in the absence of a true randomized trial or a convincing exclusion restriction. Although the approach of regression-based causality tests provides a flexible tool to analyze the causal relationships between monetary policy and real economic variables, an important drawback of this approach is that it typically needs an array of additional assumptions which are hard to appraise and explain, especially in the time series setting. Another drawback of the regression tests is that in addition to the linearity implicit in any regression test, researchers must choose conditioning variables, lag lengths, and impose assumptions that imply some sort of stationarity. To overcome these drawbacks, recently, Angrist and Kuersteiner (2011) suggested an alternative way to time series causality testing under the potential-outcomes framework. Different from the previous approaches, a major advantage of the procedure proposed in Angrist and Kuersteiner (2011) is that it shifts the focus away from modeling the process determining outcomes towards modeling the process determining policy decisions. More specifically, the procedure developed in Angrist and Kuersteiner (2011) only requires researchers to assume a model for the conditional probability of a policy shift, while leaving the model for outcome variables unspecified, so that this approach reduces the modeling burden to the specification, identification, and estimation of the structural policy innovation and thus increases robustness.

Motivated by the analysis of the Federal Open Market Committee decisions regarding the intended federal funds rate, Romer and Romer (2004) and Angrist and Kuersteiner (2011) applied some procedures to consider the causal effect of changes in the federal funds target rate, which tends to move up or down in quarter-point jumps. It is important to note that the application of the procedure in Angrist and Kuersteiner (2011) relies heavily on the conditional independence assumption which assumes that given a vector of variables which are derived from Federal Reserve forecasts of the growth rate of real GDP, the GDP deflator, and the unemployment rate, as well as a few contemporaneous variables and lags, changes in the intended federal funds target are independent of potential outcomes (in this case, the monthly percent change in industrial production). However, as discussed before, the conditional independence assumption may be violated in practice and if this occurs, the testing results from applying the approach in Angrist and Kuersteiner (2011) may lead to misleading conclusion. Hence, it is important to test whether the conditional independence assumption is true or not before applying the method in Angrist and Kuersteiner (2011). In this paper, we revisit this example by focusing on testing whether the

TABLE 3 Variable description.

Variable names	Description
graym_t	Greenbook forecast of the percentage change in real GDP/GNP (at an annual rate) for the previous quarter
gray0_t	Same as above, for current quarter
gray1_t	Same as above, for one quarter ahead
gray2_t	Same as above, for two quarter ahead
igrym_t	The innovation in the Greenbook forecast for the percentage change in GDP (at an annual rate) for the previous quarter from the meeting before. The horizon of the forecast for the meeting before is adjusted so that the forecasts for the two meetings always refer to the same quarter
igry0_t	Same as above, for current quarter
igry1_t	Same as above, for one quarter ahead
igry2_t	Same as above, for two quarters ahead
gradm_t	Greenbook forecast of the percentage change in the GDP deflator (at an annual rate) for the previous quarter
grad0_t	Same as above, for current quarter
grad1_t	Same as above, for one quarter ahead
grad2_t	Same as above, for two quarters ahead
igrdm_t	The innovation in the Greenbook forecast for the percentage change in the GDP deflator (at an annual rate) for the previous quarter from the meeting before. The horizon of the forecast for the meeting before is adjusted so that the forecasts for the two meetings always refer to the same quarter
igrd0_t	Same as above, for current quarter
igrd1_t	Same as above, for one quarter ahead
igrd2_t	Same as above, for two quarters ahead
innovation_t	Unemployment innovation
dff_t	Change in the intended federal funds rate

changes in the intended federal funds target are independent of the monthly percent change in industrial production conditional on a vector of variables.

Similar to the analysis in Angrist and Kuersteiner (2011), our analysis of the real data used in Romer and Romer (2004)[2] also focuses on a discretized version of changes in the intended federal funds rate. Different from the treatment for the policy variable in Angrist and Kuersteiner (2011), here we treat policy variable as having two values. Specifically, if the present intended federal funds rate is up relative to the last period, the policy variable B_t takes a value of 1 and 0 otherwise. In addition, we also use the same set of conditioning covariates as in the specification (a) in Angrist and Kuersteiner (2011). Specifically, the complete conditioning list includes the lagged change in the intended federal funds rate, plus the covariates graym, gray0, gray1, gray2, igrym, igry0, igry1, igry2, gradm, grad0, grad1, grad2, igrdm, igrd0, igrd1, igrd2, and the constructed unemployment innovation in Angrist and Kuersteiner (2011). Variable definitions are displayed in Table 3 and further details are referred to appendix E in Angrist and Kuersteiner (2011). Finally,

TABLE 4 Estimation results.

		Model								
Covariates	**Intercept**	**graym**	**gray0**	**gray1**	**gray2**	**igrym**	**igry0**	**igry1**	**igry2**	**gradm**
Coefficients	−1.414	−0.142	−0.033	−0.126	0.205	−0.002	0.154	0.871	−0.077	−0.518
p-value	.050	.073	.808	.615	.390	.991	.494	.015	.849	.007
Covariate	grad0	grad1	grad2	igrdm	igrd0	igrd1	igrd2	innovation	dff	
Coefficients	.582	−0.902	0.826	−0.284	−0.642	0.071	−0.872	−0.323	0.408	
p-value	.011	.014	.030	.405	.087	.893	.162	.472	.170	

TABLE 5 Testing results.

Auxiliary variable	**gray0**	**gray1**	**igrym**	**igry0**	**igry2**	**igrdm**	**igrd1**	**igry1**	**innovation**
p-value	.816	.891	.860	.911	.868	.620	.910	.876	.846

we fit a logistic regression model with B_t as the dependent variable and the estimation results are reported in Table 4. It can be seen from Table 4 that the covariates gray0, gray1, igrym, igry0, igry2, igrdm, igrd1 and the constructed unemployment innovation are highly insignificant with large p-values, which motivates us to consider using these variables as the proper candidates for the auxiliary variable A_t. Therefore, we apply our test V_N using these variables as the auxiliary variable A_t to test whether the conditional independence assumption holds or not.

Table 5 reports the testing results using gray0, gray1, igrym, igry0, igry2, igrdm, igrd1 or the constructed unemployment innovation as the auxiliary variable conditional on the remaining covariates. One can observe from Table 5 that for all cases considered, our test cannot reject the null hypothesis for different choices of auxiliary variables. Our results do not conclude that the conditional independence assumption is violated for this example so that the procedure in Angrist and Kuersteiner (2011) can be used to test the causal effect of changes in the federal funds target rate.

5 | CONCLUSION

Conditional independence is a key identification assumption in economic policy evaluation. By using auxiliary variables, we adopt a nonparametric conditional moment test for testing conditional independence in a time series context. When the policy choice is binary, we further develop an unconfoundedness test conditional on policy propensity score while the nonparametric test becomes implementable when the dimension of the observed controls is high. The asymptotic properties of the proposed statistics are provided and Monte Carlo simulations demonstrate that both test statistics have reasonable performance in finite samples. We finally apply our test procedure to revisit the empirical example of Angrist and Kuersteiner (2011). Our testing results support the identification assumption adopted in their paper. For future research, an open but challenging question is how to deal with the dimensionality problem in a general case that the policy choice is not binary.

It is of interest to investigate other type tests such as the testing procedures proposed in Donald et al. (2014) and Chen et al. (2018) for time series data and our results may be applicable to the aforementioned testing procedures. This extension is left as a future research topic.

FUNDING INFORMATION

The authors acknowledge the financial supports, in part, from the National Natural Science Foundation of China with grant numbers 72301119 and 72033008.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ENDNOTES

[1]To support this statement, which was pointed by an anonymous referee, we indeed conducted a simulation and the simulation result supports this statement. To save the space, the simulation result is not reported in the paper, available upon request.

[2]We use the same data set as in Angrist and Kuersteiner (2011), available via the Romer and Romer (2004) posting, downloadable at http://economics.mit.edu/faculty/angrist/data1/data/angrist1. Following Angrist and Kuersteiner (2011), our sample period starts in March 1969 and ends in December 1996. Data for estimation of the policy propensity score are organized by meeting month: only observations during months with Federal Open Market meetings are recorded. In the early part of the sample, the committee met twice in a month on occasion. These instances are treated as separate observations.

ORCID

Zongwu Cai https://orcid.org/0000-0001-8988-8779

REFERENCES

Angrist, J. D., Jordà, Ò., & Kuersteiner, G. M. (2018). Semiparametric estimates of monetary policy effects: String theory revisited. *Journal of Business & Economic Statistics*, *36*(3), 371–387.

Angrist, J. D., & Kuersteiner, G. M. (2011). Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *Review of Economics and Statistics*, *93*(3), 725–747.

Breunig, C. (2019). Testing missing at random using instrumental variables. *Journal of Business & Economic Statistics*, *37*(2), 223–234.

Chen, T., Ji, Y., Zhou, Y., & Zhu, P. (2018). Testing conditional mean independence under symmetry. *Journal of Business & Economic Statistics*, *36*(4), 615–627.

Dette, H., & Spreckelsen, I. (2004). Some comments on specification tests in nonparametric absolutely regular processes. *Journal of Time Series Analysis*, *25*(2), 159–172.

Donald, S. G., Hsu, Y.-C., & Lieli, R. P. (2014). Testing the unconfoundedness assumption via inverse probability weighted estimators of (L)ATT. *Journal of Business & Economic Statistics*, *32*(3), 395–415.

Fan, Y., & Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, *64*(4), 865–890.

Fan, Y., & Li, Q. (1999). Central limit theorem for degenerate u-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics*, *10*(3), 245–271.

Fang, Y., Tang, S., Cai, Z., & Lin, M. (2020). An alternative test for conditional unconfoundedness using auxiliary variables. *Economics Letters*, *194*, 109320.

Guo, X., Wang, T., & Zhu, L. (2016). Model checking for parametric single-index models: A dimension reduction model-adaptive approach. *Journal of the Royal Statistical Society: Series B*, *78*(5), 1013–1035.

Hjellvik, V., Yao, Q., & Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference*, *68*(2), 295–321.

Hu, Y., & Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, *76*(1), 195–216.

Jordà, Ò., & Taylor, A. M. (2016). The time for austerity: Estimating the average treatment effect of fiscal policy. *Economic Journal*, *126*(590), 219–255.

Lavergne, P., & Vuong, Q. (2000). Nonparametric significance testing. *Econometric Theory*, *16*(4), 576–601.

Li, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics*, *92*(1), 101–147.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton, NJ: Princeton University Press.

Liu, Z., Cai, Z., Fang, Y., & Lin, M. (2020). Statistical analysis and evaluation of macroeconomic policies: A selective review. *Applied Mathematics-A Journal of Chinese Universities*, *35*(1), 57–83.

Robinson, P. M. (1988). Root-*n*-consistent semiparametric regression. *Econometrica*, *56*(4), 931–954.

Romer, C. D., & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Phycology*, *66*(5), 688–701.

Sims, C. A. (1972). Money, income, and causality. *American Economic Review*, *62*(4), 540–552.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48.

Zhao, J., & Shao, J. (2015). Semiparametric pseudo-likelihoods in generalized linear models with nonignorable missing data. *Journal of the American Statistical Association*, *110*(512), 1577–1590.

How to cite this article: Cai, Z., Fang, Y., Lin, M., & Tang, S. (2024). Testing conditional independence in casual inference for time series data. *Statistica Neerlandica*, *78*(2), 397–426. https://doi.org/10.1111/stan.12323

APPENDIX . MATHEMATICAL PROOFS

We first give the definitions of the class of kernel functions Υ_λ and the class of functions of $\mathfrak{U}_\mu^\alpha$; see Robinson (1988), Fan and Li (1996) and Li (1999) for details.

Definition 2. Υ_λ, $\lambda \geq 1$, is the class of even functions $k : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_R u^i k(u)du = \delta_{i0} \quad (i = 0, 1, \dots, \lambda - 1),$$

and for some $\delta > 0$,

$$k(u) = O((1 + |u|^{\lambda+1+\delta})^{-1}),$$

where δ_{ij} is the Kronecher's delta.

Definition 3. $\mathfrak{U}_\mu^\alpha$, $\alpha > 0$, $\mu > 0$ is the class of functions $g : \mathbb{R}^p \to \mathbb{R}$ satisfying: g is $(l-1)$-times partially differentiable, for $l - 1 \leq \mu \leq l$; for some $\rho > 0$, $\sup_{y \in \phi_{z\rho}} |g(y) - g(z) - Q_g(y, z)|/|y - z|^\mu \leq D_g(z)$ for all z, where $\phi_{z\rho} = \{y : |y - z| < \rho\}$; $Q_g = 0$ when $l = 1$; Q_g is a $(l-1)$th degree homogeneous polynomial in $y - z$ with coefficients being the partial derivatives of g at z of orders 1 through $l - 1$ when $l > 1$; and $g(z)$, its partial derivatives of order $l - 1$ and less, and $D_g(z)$, have finite αth moments.

The remaining parts of this appendix provide proofs of the results stated in Section 2. Note that the letter C denote a generic positive constant whose value can be different for various contexts.

A.1 Proof of Theorem 1

Proof. Recall that $\alpha_t = B_t - E(B_t|M_t)$ and

$$\widehat{f}_{M_t} = \frac{1}{(N-1)a_1^q}\sum_{s\neq t}^{N} \mathrm{L}^{(1)}\left(\frac{M_s - M_t}{a_1}\right),$$

as well as

$$\widehat{B}_t = \widehat{E}(B_t|M_t) = \frac{1}{(N-1)a_1^q}\sum_{s\neq t}^{N} \mathrm{L}^{(1)}\left(\frac{M_s - M_t}{a_1}\right) B_s \Big/ \widehat{f}_{M_t}.$$

Denote $\gamma_t = \gamma(M_t) = E(B_t|M_t)$ and define

$$\widehat{\gamma}_t = \widehat{\gamma}(M_t) = \frac{1}{(N-1)a_1^q}\sum_{s\neq t}^{N} \mathrm{L}^{(1)}\left(\frac{M_s - M_t}{a_1}\right) \gamma(M_s) \Big/ \widehat{f}_{M_t},$$

and

$$\widetilde{\alpha}_t = \frac{1}{(N-1)a_1^q}\sum_{s\neq t}^{N} \mathrm{L}^{(1)}\left(\frac{M_s - M_t}{a_1}\right) \alpha_s \Big/ \widehat{f}_{M_t}.$$

Then, $\widehat{B}_t = \widehat{\gamma}_t + \widetilde{\alpha}_t$. Using $\widehat{\alpha}_t = B_t - \widehat{B}_t = (\gamma_t - \widehat{\gamma}_t) + (\alpha_t - \widetilde{\alpha}_t)$, the test statistic U_N defined in Section 2 can be rewritten as

$$\begin{aligned} U_N = {} & \frac{1}{N(N-1)a^{p+q}} \sum_{t=1}^{N} \sum_{s\neq t, s=1}^{N} \Big[(\gamma_t - \widehat{\gamma}_t)\widehat{f}_{M_t}(\gamma_s - \widehat{\gamma}_s)\widehat{f}_{M_s} + \alpha_t\alpha_s\widehat{f}_{M_t}\widehat{f}_{M_s} \\ & + \widetilde{\alpha}_t\widetilde{\alpha}_s\widehat{f}_{M_t}\widehat{f}_{M_s} + 2\alpha_t\widehat{f}_{M_t}(\gamma_s - \widehat{\gamma}_s)\widehat{f}_{M_s} \\ & -2\widetilde{\alpha}_t\widehat{f}_{M_t}(\gamma_s - \widehat{\gamma}_s)\widehat{f}_{M_s} - 2\alpha_t\widehat{f}_{M_t}\widetilde{\alpha}_s\widehat{f}_{M_s} \Big] L_{ts} \\ := {} & U_{N1} + U_{N2} + U_{N3} + 2U_{N4} - 2U_{N5} - 2U_{N6}, \end{aligned}$$

where L_{ts} is defined in Section 2. We shall complete the proof of Theorem 1 by investigating U_{Ni} for $i = 1, \ldots, 6$, respectively, in the following Lemmas 2 and 3. Since the proof is similar to that of theorem 3.1 of Li (1999), we only provide some key steps. ■

Lemma 2. *Under Assumptions 2–5, then*,

$$U_{N1} = o_p\left((Na^{(p+q)/2})^{-1}\right), \quad U_{N3} = o_p\left((Na^{(p+q)/2})^{-1}\right),$$

and

$$U_{N4} = o_p\left((Na^{(p+q)/2})^{-1}\right), \quad U_{N5} = o_p\left((Na^{(p+q)/2})^{-1}\right), \quad U_{N6} = o_p\left((Na^{(p+q)/2})^{-1}\right).$$

Proof. See lemmas A.1, A.3, A.4, A.5 and A.6 in Li (1999). ■

Lemma 3. *Suppose that Assumptions 2–5 are satisfied, then,*

(i) $Na^{(p+q)/2}U_{N2} \xrightarrow{D} \mathcal{N}(0, 2\sigma_U^2)$, *where* $\sigma_U^2 = E\left[f^2(M_t)f_G(G_t)\sigma^4(G_t)\right]\left(\int L^2(v)dv\right)$.
(ii) $\hat{\sigma}_U^2 = \sigma_U^2 + o_p(1)$.

Proof. (i) First note that U_{N2} can be rewritten as

$$\begin{aligned} U_{N2} &= \frac{1}{N(N-1)a^{p+q}}\sum_{t=1}^{N}\sum_{s\neq t}^{N}\alpha_t\alpha_s\hat{f}_{M_t}\hat{f}_{M_s}L_{ts} \\ &= \frac{1}{N(N-1)a^{p+q}}\sum_{t=1}^{N}\sum_{s\neq t}^{N}\alpha_t\alpha_s f(M_t)f(M_s)L_{ts} \\ &\quad + \frac{2}{N(N-1)a^{p+q}}\sum_{t=1}^{N}\sum_{s\neq t}^{N}\alpha_t\alpha_s\left(\hat{f}_{M_t} - f(M_s)\right)f(M_s)L_{ts} \\ &\quad + \frac{1}{N(N-1)a^{p+q}}\sum_{t=1}^{N}\sum_{s\neq t}^{N}\alpha_t\alpha_s\left(\hat{f}_{M_t} - f(M_t)\right)\left(\hat{f}_{M_s} - f(M_s)\right)L_{ts} \\ &:= U_{N2}^{(1)} + 2U_{N2}^{(2)} + U_{N2}^{(3)}. \end{aligned}$$

Following lemma A.2 in Li (1999), it can be known that $U_{N2}^{(k)} = o_p\left((Na^{(p+q)/2})^{-1}\right)$ for $k = 2$ and 3. Hence, it remains to check the asymptotic normality of $Na^{(p+q)/2}U_{N2}^{(1)}$. Similar to the proof of theorem 1 in Dette and Spreckelsen (2004), here we also apply lemma 3.2 in Hjellvik, Yao, and Tjøstheim (1998) for the degenerate U-statistic to prove that $Na^{(p+q)/2}U_{N2}^{(1)}$ is normally distributed. To do this, denote $\xi_t = (A_t, \alpha_t)$, $P(\xi_s)$, $P(\xi_s, \xi_t)$, $P(\xi_s, \xi_t, \xi_l)$, and let $P(\xi_s, \xi_t, \xi_l, \xi_k)$ be the probability measures of ξ_s, (ξ_s, ξ_t), (ξ_s, ξ_t, ξ_l) and $(\xi_s, \xi_t, \xi_l, \xi_k)$ for different $s, t, l, k \in \{1, \dots, N\}$, respectively. Define $\varphi_{st} = \varphi_{st}(\xi_t, \xi_s) = \alpha_t\alpha_s f(M_t)f(M_s)L_{ts}/(N(N-1)a^{p+q})$. It is easy to observe that φ_{st} is a symmetric function on its arguments. Thus, $U_{N2}^{(1)} = \sum_{1\le t\neq s\le N}\varphi_{st} = 2\sum_{1\le s<t\le N}\varphi_{st}$ is a degenerate U-statistics. Denote $\sigma_{st}^2 = \mathrm{Var}(\varphi_{st})$ and $\sigma_N^2 = \sum_{1\le s<t\le N}\sigma_{st}^2$. For some constant $\delta > 0$, define

$$M_{N1} = \max_{1<s<t\le N}\max\left\{E|\varphi_{1t}\varphi_{st}|^{1+\delta}, \int|\varphi_{1t}\varphi_{st}|^{1+\delta}dP(\xi_1)dP(\xi_s, \xi_t)\right\},$$

$$\begin{aligned} M_{N2} = \max_{1<s<t\le N}\max\bigg\{&E|\varphi_{1t}\varphi_{st}|^{2(1+\delta)}, \int|\varphi_{1t}\varphi_{st}|^{2(1+\delta)}dP(\xi_1)dP(\xi_s, \xi_t), \\ &\int|\varphi_{1t}\varphi_{st}|^{2(1+\delta)}dP(\xi_1, \xi_s)dP(\xi_t), \int|\varphi_{1t}\varphi_{st}|^{2(1+\delta)}dP(\xi_1)dP(\xi_s)dP(\xi_t)\bigg\}, \end{aligned}$$

$$M_{N3} = \max_{1<s<t\le N}E|\varphi_{1t}\varphi_{st}|^2, \quad M_{N4} = \max_{\substack{1<s,t,l\le n \\ s,t,l\text{different}}}\left\{\max_P\int|\varphi_{1s}\varphi_{tl}|^{2(1+\delta)}dP\right\},$$

$$M_{N5} = \max_{1<s<t\le N}\max\left\{E\left|\int\varphi_{1s}\varphi_{1t}dP(\xi_1)\right|^{2(1+\delta)}, \int\left|\int\varphi_{1s}\varphi_{1t}dP(\xi_1)\right|^{2(1+\delta)}dP(\xi_s)dP(\xi_t)\right\},$$

$$M_{N6} = \max_{1<s<t\le N}E\left|\int\varphi_{1s}\varphi_{1t}dP(\xi_1)\right|^2,$$

where the maximization over P in the equation for M_{N4} is taken over the four probability measures $P(\xi_1, \xi_s, \xi_t, \xi_l)$, $P(\xi_1)P(\xi_s, \xi_t, \xi_l)$, $P(\xi_1)P(\xi_s)P(\xi_t, \xi_l)$ and $P(\xi_1)P(\xi_s)P(\xi_t)P(\xi_l)$

for mutually different s, t, l. It is assumed that all of the above constants are finite.

Based on lemma 3.2 in Hjellvik et al. (1998), $\sigma_N^{-1}\sum_{1\le s<t\le N}\varphi_{st}=\frac{1}{2}\sigma_N^{-1}U_{N2}^{(1)}$ is asymptotically normal with mean zero and variance one if for some $\delta>0$, as $N\to\infty$,

$$\max\sigma_N^{-2}\left\{N^2\left(M_{N1}^{1/(1+\delta)}+M_{N5}^{1/2(1+\delta)}+M_{N6}^{1/2}\right),N^{3/2}\left(M_{N2}^{1/2(1+\delta)}+M_{N3}^{1/2}+M_{N4}^{1/2(1+\delta)}\right)\right\}\to 0.$$

To this end, we only investigate the order of magnitude of $M_{N1}^{1/(1+\delta)}$, because the other terms can be studied in a similar fashion. We first consider M_{N1}. Define $u_t=\alpha_t f(M_t)$ and $p_{st}=L_{st}/(N(N-1)a^{p+q})$, then $\varphi_{st}=u_tu_sp_{st}$. By applying Hölder's inequality, it is easy to obtain that for some $0<\delta<1$ and $1\le s<t<l\le N$

$$\begin{aligned}E\left[\left|\varphi_{1t}\varphi_{st}\right|^{1+\delta}\right]&=E\left[\left|u_1u_su_t^2p_{1t}p_{st}\right|^{1+\delta}\right]\\&\le\left\{E\left[\left|u_1u_su_t^2\right|^{2(1+\delta)(1+\delta_2)}\right]\right\}^{\frac{1}{2(1+\delta_2)}}\left\{E\left[\left|p_{1t}p_{st}\right|^{(1+\delta)(1+\delta_1)}\right]\right\}^{\frac{1}{1+\delta_1}},\end{aligned}$$

where $0<\delta_1<1$ and $0<\delta_2<1$, satisfying $\frac{1}{1+\delta_1}+\frac{1}{2(1+\delta_2)}=1$ and $\frac{1+\delta}{3-\delta}<\delta_1<\frac{1-\delta}{1+\delta}$. Note that

$$1<\zeta=(1+\delta)(1+\delta_2)<2,\quad\text{and}\quad 1<\eta=(1+\delta)(1+\delta_1)<2,$$

so that

$$\left\{E\left[\left|u_1u_su_t^2\right|^{2(1+\delta)(1+\delta_2)}\right]\right\}^{\frac{1}{2(1+\delta_2)}}=\left\{E\left[\left|u_1^2u_s^2u_t^4\right|^{(1+\delta)(1+\delta_2)}\right]\right\}^{\frac{1}{2(1+\delta_2)}}<\infty,$$

by Assumptions 3(iii) and 5(i). By straightforward calculations and by Assumption 4(ii), we can obtain that

$$\begin{aligned}E\left(\left|p_{1t}p_{st}\right|^{\eta}\right)&=\frac{1}{(N(N-1)a^{p+q})^{2\eta}}\int\int\int\left|L\left(\frac{u-r}{h}\right)\right|^{\eta}\left|L\left(\frac{v-r}{h}\right)\right|^{\eta}f(u,v,r)dudvdr\\&=\frac{a^{2(p+q)}}{\left(N(N-1)a^{(p+q)}\right)^{2\eta}}\int\int\int\left|L(x)L(y)\right|^{\eta}f(z+hx,z+hy,z)dxdydz\\&\le C\frac{a^{2(p+q)}}{(N(N-1)a^{p+q})^{2\eta}},\end{aligned}$$

where $f(u,v,r)$ is the joint density function of (G_1,G_s,G_t). Moreover, note that

$$\begin{aligned}\sigma_N^2&=\frac{1}{2}\sum_{t=1}^{N}\sum_{s\ne t}\mathrm{Var}(\varphi_{st})=\frac{1}{2}\sum_{t=1}^{N}\sum_{s\ne t}E(\varphi_{st}^2)\\&=\frac{1}{2(N(N-1)a^{p+q})^2}\sum_{t=1}^{N}\sum_{s\ne t}E\left(\alpha_t^2\alpha_s^2f^2(M_t)f^2(M_s)L_{ts}^2\right)\\&=\frac{1}{2N(N-1)a^{p+q}}\left(\sigma_U^2+o(1)\right),\end{aligned}$$

where the last equality results from lemma A.2(ii) in Li (1999). Consequently, for any $1 < s < t \leq N$, we have

$$N^2 \left[E\left|\varphi_{1t}\varphi_{st}\right|^{1+\delta} \right]^{\frac{1}{1+\delta}} \Big/ \sigma_N^2 \leq \frac{C^{1/\eta} N^2 a^{2(p+q)/\eta} N(N-1)a^{p+q}}{(N(N-1)a^{p+q})^2 \left(\sigma_U^2 + o(1)\right)} = \frac{C^{1/\eta} a^{2(p+q)/\eta-(p+q)}}{\sigma_U^2 + o(1)} \to 0,$$

since $\eta < 2$.

We now consider the second term in M_{N1}. To this end, denote E_i and E_{ij} as the expectations with respect to ξ_i and (ξ_i, ξ_j), respectively. Then, we have

$$\begin{aligned} E_1 E_{st}\left[\left|\varphi_{1t}\varphi_{st}\right|^{1+\delta}\right] &= \int \left|\varphi_{1t}\varphi_{st}\right|^{1+\delta} dP(\xi_1)dP(\xi_s,\xi_t) \\ &\leq CE_1\left\{|\alpha_1|^{1+\delta} E_{st}\left[\left|\alpha_s\alpha_t^2 p_{1t}p_{st}\right|^{1+\delta}\right]\right\} \\ &\leq CE_1\left[|\alpha_1|^{1+\delta}\left\{E_{st}\left[\left|\alpha_s\alpha_t^2\right|^{2(1+\delta)(1+\delta_2)}\right]\right\}^{\frac{1}{2(1+\delta_2)}}\left\{E_{st}\left[\left|p_{1t}p_{st}\right|^{(1+\delta)(1+\delta_1)}\right]\right\}^{\frac{1}{1+\delta_1}}\right] \\ &\leq \left\{\frac{Ca^{2(p+q)}}{(N(N-1)a^{p+q})^{2\eta}}\right\}^{\frac{1}{1+\delta_1}}. \end{aligned}$$

Therefore,

$$\sigma_N^{-2} N^2 (\left\{\int \left|\varphi_{1t}\varphi_{st}\right|^{1+\delta} dP(\xi_1)dP(\xi_s,\xi_t)\right\}^{\frac{1}{1+\delta}} \leq \sigma_N^{-2} N^2 \left\{\frac{Ca^{2(p+q)}}{(N(N-1)a^{p+q})^{2\eta}}\right\}^{\frac{1}{\eta}} \to 0,$$

as $N \to \infty$. Hence, the proof of $\sigma_N^{-2} N^2 M_{N1}^{1/(1+\delta)} \to 0$ is completed.

For M_{N2}, we only consider $E|\varphi_{1t}\varphi_{st}|^{2(1+\delta)}$ and the other terms can be investigated similarly. It is noted that

$$\left[E\left|\varphi_{1t}\varphi_{st}\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}} \leq \frac{Ca^{2(p+q)/\zeta}}{(N(N-1)a^{p+q})^2},$$

where $2 < \zeta = 2(1+\delta)(1+\delta_1) < 4$, which implies that

$$\sigma_N^{-2} N^{3/2} \left[E\left|\varphi_{1t}\varphi_{st}\right|^{2(1+\delta)}\right] \leq \frac{1}{N^{1/2} a^{(p+q)-2(p+q)/\zeta}} \frac{1}{\sigma_U^2 + o(1)} \to 0, \quad \text{as} \quad N \to \infty,$$

by the assumption that $Na^{p+q} \to \infty$. Similarly, it is not difficult to verify the results of other terms and details thus are omitted. Finally, by combining Lemmas 2 and 3, the proof of Theorem 1 is finished. ∎

A.2 Proof of Theorem 2

The proof is similar to that of corollary 3.2 in Li (1999) and the details thus are omitted.

A.3 Proof of Lemma 1

Proof. Under Assumptions 1 and 2, we first show that

$$A_t \perp\!\!\!\perp B_t | \left(Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right) \text{ for all } \nu \geq 0 \text{ and for all } b_j \text{ with } \upsilon \text{ fixed for } \upsilon \in \Upsilon, \tag{A1}$$

holds. Noting that B_t is a binary variable, then for all $\nu \geq 0$ and for all b_j with $\upsilon \in \Upsilon$ fixed, we have

$$\begin{aligned} P\left(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t), A_t\right) &= E\left(B_t | Y_{t,\nu}^{\upsilon}(b_j), p(M_t), A_t\right) \\ &= E\left\{E\left[B_t | Y_{t,\nu}^{\upsilon}(b_j), M_t, A_t\right] \middle| y_{t,\nu}^{\upsilon}(b_j), p(M_t), A_t\right\} \\ &= E\left\{E\left[B_t | Y_{t,\nu}^{\upsilon}(b_j), M_t\right] \middle| Y_{t,\nu}^{\upsilon}(b_j), p(M_t), A_t\right\} \\ &= E\left\{E\left[B_t | M_t\right] \middle| Y_{t,\nu}^{\upsilon}(b_j), p(M_t), A_t\right\} = p(M_t). \end{aligned}$$

Similarly,

$$\begin{aligned} P\left(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right) &= E\left[B_t | Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right] \\ &= E\left\{E\left[B_t | Y_{t,\nu}^{\upsilon}(b_j), M_t\right] \middle| Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right\} \\ &= E\left\{E\left[B_t | M_t\right] \middle| Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right\} = p(M_t). \end{aligned}$$

Therefore, $P(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t), A_t) = P(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t))$, which implies (A1) holds. Next, we show that (3) in Lemma 1 is true. To this end, Assumption 1 implies that

$$E\left(P\left(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right) - P\left(B_t = 1 | p(M_t)\right) | p(M_t), A_t\right) = 0. \tag{A2}$$

By (A1),

$$P\left(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right) = P\left(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t), A_t\right),$$

and thus, the law of iterated expectations yields

$$E\left(P\left(B_t = 1 | Y_{t,\nu}^{\upsilon}(b_j), p(M_t)\right) | p(M_t), A_t\right) = E\left(B_t | p(M_t), A_t\right). \tag{A3}$$

Therefore, a combination of (A2) and (A3) leads to

$$E\left(B_t | p(M_t), A_t\right) = E\left(B_t | p(M_t)\right) = p(M_t),$$

which implies B_t is independent of A_t conditional on $p(M_t)$, and thus the proof of Lemma 1 is completed. ■

A.4 Proof of Theorem 3

Proof. Define $\eta_t = p(M_t;\theta_0) - p(M_t;\widehat{\theta}_N)$. We then have $\widehat{\varepsilon}_t = B_t - p(M_t;\theta_0) + p(M_t;\theta_0) - p(M_t;\widehat{\theta}_N) = \varepsilon_t + \eta_t$. Consequently, V_N can be rewritten as

$$
\begin{aligned}
V_N &= \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}(\varepsilon_t+\eta_t)(\varepsilon_s+\eta_s)L_1\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)L_2\left(\frac{A_s-A_t}{\ell}\right)\\
&= \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}\varepsilon_t\varepsilon_s L_1\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)L_2\left(\frac{A_s-A_t}{\ell}\right)\\
&\quad+\frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}\varepsilon_t\eta_s L_1\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)L_2\left(\frac{A_s-A_t}{\ell}\right)\\
&\quad+\frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}\eta_t\varepsilon_s L_1\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)L_2\left(\frac{A_s-A_t}{\ell}\right)\\
&\quad+\frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}\eta_t\eta_s L_1\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)L_2\left(\frac{A_s-A_t}{\ell}\right)\\
&:= V_{N1}+V_{N2}+V_{N3}+V_{N4}.
\end{aligned}
$$

We first consider V_{N1}. To this end, rewrite $V_{N1} = V_{N1}^{(1)} + V_{N1}^{(2)}$, where

$$
V_{N1}^{(1)} = \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}\varepsilon_t\varepsilon_s L_1\left(\frac{p(M_s;\theta_0)-p(M_t;\theta_0)}{\ell}\right)L_2\left(\frac{A_s-A_t}{\ell}\right),
$$

and

$$
\begin{aligned}
V_{N1}^{(2)} = \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}\varepsilon_t\varepsilon_s\Bigg\{&L_1\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)\\
&-L_1\left(\frac{p(M_s;\theta_0)-p(M_t;\theta_0)}{\ell}\right)\Bigg\}L_2\left(\frac{A_s-A_t}{\ell}\right).
\end{aligned}
$$

For $V_{N1}^{(1)}$, similar to the arguments for showing Theorem 1 above, we can show that

$$
\frac{N\ell^{\kappa/2}V_{N1}^{(1)}}{\sqrt{2}\widehat{\sigma}_V}\xrightarrow{D}\mathcal{N}(0,1),
$$

where

$$
\widehat{\sigma}_V^2 = \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^{N}\sum_{s\neq t}\widehat{\varepsilon}_t^2\widehat{\varepsilon}_s^2 L_1^2\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)L_2^2\left(\frac{A_s-A_t}{\ell}\right),
$$

is a consistent estimator of σ_V^2 given by

$$
\sigma_V^2 = E\left(\sigma^4(R_t)f_R(R_t)\right)\left(\int L_1^2(u)du\cdot\int L_2^2(v)dv\right),
$$

with $\sigma^2(r) = E(\varepsilon_t^2|R_t = r)$ and $f_R(\cdot)$ being the density function of $R_t = (p(M_t), A_t)$. For $V_{N1}^{(2)}$, following the proof of theorem 1 in Guo, Wang, and Zhu (2016), the Taylor expansion yields that

$$V_{N1}^{(2)} = \frac{(\hat{\theta}_N - \theta_0)^T}{\ell} \frac{1}{N(N-1)\ell^\kappa} \sum_{t=1}^{N} \sum_{s \neq t} \varepsilon_t \varepsilon_s L_2\left(\frac{A_s - A_t}{\ell}\right) \times L_1'\left(\frac{p(M_s;\theta_0) - p(M_t;\theta_0)}{\ell}\right) (\nabla p(M_s;\theta_0) - \nabla p(M_t;\theta_0)) \left(1 + o_p(1)\right).$$

Let

$$V_{N1}^{*} = \frac{1}{N(N-1)\ell^\kappa} \sum_{t=1}^{N} \sum_{s \neq t} \varepsilon_t \varepsilon_s L_2\left(\frac{A_s - A_t}{\ell}\right) L_1'\left(\frac{p(M_s;\theta_0) - p(M_t;\theta_0)}{\ell}\right) \times (\nabla p(M_s;\theta_0) - \nabla p(M_t;\theta_0)).$$

It is easy to see that V_{N1}^{*} is a degenerate U-statistic with kernel

$$H_N\left((M_t, A_t, \varepsilon_t), (M_s, A_s, \varepsilon_s)\right) = L_2\left(\frac{A_s - A_t}{\ell}\right) L_1'\left(\frac{p(M_s;\theta_0) - p(M_t;\theta_0)}{\ell}\right) (\nabla p(M_s;\theta_0) - \nabla p(M_t;\theta_0))\, \varepsilon_t \varepsilon_s.$$

Hence, $N\ell^{\kappa/2} V_{N1}^{*} = O_p(1)$. By combining $\|\hat{\theta}_N - \theta_0\| = O_p(1/\sqrt{N})$ and $N\ell^2 \to \infty$, one can obtain $N\ell^{\kappa/2} V_{N1}^{(2)} = o_p(1)$.

Now we deal with the term V_{N2}. To do so, rewrite $V_{N2} = V_{N2}^{(1)} + V_{N2}^{(2)}$, where

$$V_{N2}^{(1)} = \frac{1}{N(N-1)\ell^\kappa} \sum_{t=1}^{N} \sum_{s \neq t} \varepsilon_t L_2\left(\frac{A_s - A_t}{\ell}\right) L_1\left(\frac{p(M_s;\theta_0) - p(M_t;\theta_0)}{\ell}\right) \left(p(M_s;\theta_0) - p(M_s;\hat{\theta}_N)\right),$$

and

$$V_{N2}^{(2)} = \frac{1}{N(N-1)\ell^\kappa} \sum_{t=1}^{N} \sum_{s \neq t} \varepsilon_t L_2\left(\frac{A_s - A_t}{\ell}\right) \left(p(M_s;\theta_0) - p(M_s;\hat{\theta}_N)\right) \times \left[L_1\left(\frac{p(M_s;\hat{\theta}_N) - p(M_t;\hat{\theta}_N)}{\ell}\right) - L_1\left(\frac{p(M_s;\theta_0) - p(M_t;\theta_0)}{\ell}\right)\right].$$

We first show that $V_{N2}^{(1)} = o_p((N\ell^{\kappa/2})^{-1})$, which can be established by the similar arguments to the proof of theorem 3.1 in Fan and Li (1999), so here we only provide some key steps. For simplicity of notation, denote $\Lambda_{ts} = L_2\left(\frac{A_s - A_t}{\ell}\right) L_1\left(\frac{p(M_s;\theta_0) - p(M_t;\theta_0)}{\ell}\right)$. By using $p(M_s;\hat{\theta}_N) - p(M_s;\theta_0) = \nabla' p(M_s;\theta_0)(\hat{\theta}_N - \theta_0) + \frac{1}{2}(\hat{\theta}_N - \theta_0)' \nabla^2 p(M_s;\tilde{\theta}_N)(\hat{\theta}_N - \theta_0)$, where $\tilde{\theta}_N$ is between $\hat{\theta}_N$ and θ_0, we obtain the following

$$V_{N2}^{(1)} = -\frac{1}{N(N-1)\ell^\kappa} \sum_{t=1}^{N} \sum_{s \neq t} \left\{ (\hat{\theta}_N - \theta_0)' \varepsilon_t \nabla p(M_s;\theta_0) \right.$$

$$+(\widehat{\theta}_N - \theta_0)'\varepsilon_t \nabla^2 p(M_s; \widetilde{\theta}_N)(\widehat{\theta}_N - \theta_0)/2 \Big\} \Lambda_{ts}$$
$$:= -(\widehat{\theta}_n - \theta_0)'\Xi_{1N} - (\widehat{\theta}_N - \theta_0)'\Xi_{2N}(\widehat{\theta}_N - \theta_0),$$

where

$$\Xi_{1N} = \frac{1}{N(N-1)\ell^{\kappa}} \sum_{t=1}^{N} \sum_{s \neq t} \varepsilon_t \nabla p(M_s; \theta_0)\Lambda_{ts},$$

and

$$\Xi_{2N} = \frac{1}{2N(N-1)\ell^{\kappa}} \sum_{t=1}^{N} \sum_{s \neq t} \varepsilon_t \nabla^2 p(M_s; \widetilde{\theta}_N)\Lambda_{ts}.$$

For Ξ_{1N}, let (a) denotes the case of $\min\{|s-s'|, |s-t|, |s-t'|\} > e$, (b) denotes the case of $\min\{|s-s'|, |s-t|, |s-t'|\} \leq e$, where $e = [C\log(N)]$, and $\Delta_t = \nabla p(M_s; \theta_0)$. Note that

$$\begin{aligned}
\max_{t\neq s, t' \neq s'} E\left(\Delta_t \Delta'_{t'} \varepsilon_s \varepsilon_{s'} \Lambda_{ts}\Lambda_{t's'}\right) &\leq C \max_{t\neq s, t'\neq s'} E\left(G_p(M_t)G_p(M_{t'})\varepsilon_s\varepsilon_{s'}\Lambda_{ts}\Lambda_{t's'}\right) \\
&\leq C \max_{t\neq s\neq t'} \{E|\varepsilon_s\varepsilon_{s'}|^{\xi}\}^{1/\xi} \left\{E\left[\left(G_p(M_t)G_p(M_{t'})\Lambda_{ts}\Lambda_{t's'}\right)^{\eta}\right]\right\}^{1/\eta} \\
&= O(\ell^{2\kappa/\eta}),
\end{aligned}$$

by Assumption 9, where $\eta = (1-\xi^{-1})^{-1} (\xi > 2, 1 < \eta < 2)$. Then,

$$\begin{aligned}
E(\|\Xi_{1N}\|^2) &= (N(N-1)\ell^{\kappa})^{-2} \sum_t \sum_{s\neq t} \sum_{t'} \sum_{s'\neq t'} E\left(\Delta'_t \varepsilon_s \Delta_{t'}\varepsilon_{s'}\Lambda_{ts}\Lambda_{t's'}\right) \\
&= (N(N-1)\ell^{\kappa})^{-2} \left\{ \sum_{(a)} + \sum_{(b)} \right\} E\left(\Delta'_t \varepsilon_s \Delta_{t'}\varepsilon_{s'}\Lambda_{ts}\Lambda_{t's'}\right) \\
&\leq (N(N-1)\ell^{\kappa})^{-2} \left(CN^4 \beta_e^{\delta/(1+\delta)} + eN^3\right) \max_{t\neq s, t'\neq s'} E\left(\Delta'_t \Delta_{t'}\varepsilon_s\varepsilon_{s'}\Lambda_{ts}\Lambda_{t's'}\right) \\
&= (N(N-1)\ell^{\kappa})^{-2} \left(o(1) + N^3 O(\ell^{2\kappa/\eta})\right),
\end{aligned}$$

for some $1 < \eta < 2$, where $C > 4(1+\delta)/(\eta\delta)$. Therefore, $E(\|\Xi_{1N}\|^2) = o\left((N(N-1)\ell^{\kappa})^{-2}\right) + O(e(N\ell^{2\kappa(\eta-1)/\eta})^{-1})$ and this implies $\Xi_{1N} = o_p((N(N-1)\ell^{\kappa})^{-1}) + O_p\left(e^{1/2}(N^{-1/2}\ell^{-\kappa(\eta-1)/\eta})\right)$, which leads to $(\widehat{\theta}_N - \theta_0)'\Xi_{1N} = N^{-1/2} o_p((N(N-1)\ell^{\kappa})^{-1}) + O_p\left(e^{1/2}(N^{-1}\ell^{-\kappa(\eta-1)/\eta})\right) = o_p((N\ell^{\kappa/2})^{-1})$ because $1 < \eta < 2$, $e = [C\log(N)]$ and $\ell \sim N^{-\overline{\delta}}\left(\frac{7}{8}\kappa > \overline{\delta} > 0\right)$. It remains to evaluate the order of A_{2N}. According to Assumption 10(i), it is easy to obtain that

$$\begin{aligned}
E\|\Xi_{2N}\| &\leq ((N(N-1)\ell^{\kappa})^{-1})N^2 \max_{t\neq s} E\|\varepsilon_s G_p(M_t)\Lambda_{ts}\| \\
&\leq \ell^{-\kappa} \max_{t\neq s} \left\{[E(\varepsilon_s^{\xi})]^{1/\xi}\right\} \left\{E[(G_p(M_t)\Lambda_{ts})^{\eta}]\right\}^{1/\eta} = \ell^{-\kappa} O(\ell^{\kappa/\eta}) = O(\ell^{-(\eta-1)\kappa/\eta}),
\end{aligned}$$

for some $1 < \eta < 2$. Hence, $(\widehat{\theta}_N - \theta_0)'\Xi_{2N}(\widehat{\theta}_N - \theta_0) = O_p((N\ell^{(\eta-1)\kappa/\eta})^{-1}) = o_p((N\ell^{\kappa/2})^{-1})$. By summarizing the above, we have shown that $V_{N2}^{(1)} = o_p((N\ell^{\kappa/2})^{-1})$.

Denote

$$\Xi_N^* = \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^N\sum_{s\neq t}\varepsilon_t L_2\left(\frac{A_s-A_t}{\ell}\right)\left(p(M_s;\theta_0)-p(M_s;\widehat{\theta}_N)\right)L_1'\left(\frac{p(M_s;\theta_0)-p(M_t;\theta_0)}{\ell}\right)$$
$$\times\frac{\left((p(M_s;\widehat{\theta}_N)-p(M_s;\theta_0))-(p(M_t;\widehat{\theta}_N)-p(M_t;\theta_0))\right)}{\ell}.$$

By following the same arguments for proving theorem 1 in Guo et al. (2016), we have

$$V_{N2}^{(2)} = A_N^* + o_p(\Xi_N^*).$$

Denote,

$$\Xi_N^{**} = \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^N\sum_{s\neq t}\varepsilon_t L_2\left(\frac{A_s-A_t}{\ell}\right)L_1'\left(\frac{p(M_s;\theta_0)-p(M_t;\theta_0)}{\ell}\right)\left(p(M_s;\theta_0)-p(M_s;\widehat{\theta}_N)\right).$$

Then, by the similar arguments for proving $V_{N2}^{(1)} = o_p((N\ell^{\kappa/2})^{-1})$, one can show that $A_N^{**} = o_p((N\ell^{\kappa/2})^{-1})$. By combining $\sup_{m\in\mathcal{M}}\left|p(m;\widehat{\theta}_N)-p(m;\theta_0)\right| = O_p(1/\sqrt{N})$ and $N\ell^2\to\infty$, it concludes that $V_{N2}^{(2)} = o_p((N\ell^{\kappa/2})^{-1})$. As a result, $V_{N2} = o_p((N\ell^{\kappa/2})^{-1})$. Using a similar argument for the term V_{N2} above, we also can show that $V_{N3} = o_p((N\ell^{\kappa/2})^{-1})$. Finally, we deal with the term V_{N4}. Rewrite it as

$$\begin{aligned}
V_{N4} &= \frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^N\sum_{s\neq t}\eta_t\eta_s L_1\left(\frac{p(M_s;\theta_0)-p(M_t;\theta_0)}{\ell}\right)L_2\left(\frac{A_s-A_t}{\ell}\right)\\
&\quad+\frac{1}{N(N-1)\ell^\kappa}\sum_{t=1}^N\sum_{s\neq t}\eta_t\eta_s L_2\left(\frac{A_s-A_t}{\ell}\right)\left\{L_1\left(\frac{p(M_s;\widehat{\theta}_N)-p(M_t;\widehat{\theta}_N)}{\ell}\right)\right.\\
&\quad\left.-L_1\left(\frac{p(M_s;\theta_0)-p(M_t;\theta_0)}{\ell}\right)\right\}\\
&:= V_{N4}^{(1)} + V_{N4}^{(2)}.
\end{aligned}$$

By the mean value theorem and Assumption 10(i), $\left|p(m;\widehat{\theta}_N)-p(m;\theta_0)\right| \leq CG_p(m)\|\widehat{\theta}_N-\theta_0\|$. Hence,

$$\begin{aligned}
|V_{N4}^{(1)}| &\leq (C^2/(N(N-1)\ell^\kappa))\sum_{t=1}^N\sum_{s\neq t}G_p(M_t)G_p(M_s)L_2\left(\frac{A_s-A_t}{\ell}\right)\\
&\quad\times L_1\left(\frac{p(M_s;\theta_0)-p(M_t;\theta_0)}{\ell}\right)\left\|\widehat{\theta}_n-\theta_0\right\|^2\\
&:= \Delta_N\left\|\widehat{\theta}_N-\theta_0\right\|^2 = O_p(N^{-1}),
\end{aligned}$$

because $E|\Delta_N| = O(1)$ and $\|\hat{\theta}_N - \theta_0\| = O_p(1/\sqrt{N})$. For the term $V_{N4}^{(2)}$, it is easy to see that

$$V_{N4}^{(2)} = \frac{1}{N(N-1)\ell^{\kappa}} \sum_{t=1}^{N} \sum_{s \neq t} \eta_t \eta_s L_2 \left(\frac{A_s - A_t}{\ell} \right) L_1' \left(\frac{p(M_s;\theta_0) - p(M_t;\theta_0)}{\ell} \right)$$
$$\times \frac{\left((p(M_s;\hat{\theta}_N) - p(M_s;\theta_0)) - (p(M_t;\hat{\theta}_N) - p(M_t;\theta_0)) \right)}{\ell} \left(1 + o_p(1) \right).$$

A similar argument for proving $V_{N4}^{(1)}$ and Assumption 13 can be used to derive that $V_{N4}^{(2)} = o_p((N\ell^{\kappa/2})^{-1})$. Consequently, $V_{N4} = ((N\ell^{\kappa/2})^{-1})$. Therefore, the proof of Theorem 3 is completed. ■