Uniform Nonparametric Inference for Time Series

Jia Li† and Zhipeng Liao‡

May 14, 2019

Abstract
This paper provides the first result for the uniform inference based on nonparametric series estimators in a general time-series setting. We develop a strong approximation theory for sample averages of mixingales with dimensions growing with the sample size. We use this result to justify the asymptotic validity of a uniform confidence band for series estimators and show that it can also be used to conduct nonparametric specification test for conditional moment restrictions. New results on the validity of high-dimensional heteroskedasticity and autocorrelation consistent (HAC) estimators are established for making feasible inference. Further extensions include time-series inference theories for intersection bounds and convex sieve M-estimators, which permit applications in partially identified models and nonparametric conditional quantile estimation, respectively. An empirical application on the unemployment volatility puzzle for the search and matching model is provided as an illustration.

Keywords: martingale difference, mixingale, series estimation, specification test, strong approximation, uniform inference.

---

†Department of Economics, Duke University, Durham, NC 27708; e-mail: jl410@duke.edu.
‡Department of Economics, UCLA, Los Angeles, CA 90095; e-mail: zhipeng.liao@econ.ucla.edu.
1 Introduction

Series estimators play a central role in econometric analysis that involves nonparametric components. Such problems arise routinely from applied work because the economic intuition of the guiding economic theory often does not depend on stylized parametric model assumptions. The simple, but powerful, idea of series estimation is to approximate the unknown function using a large (asymptotically diverging) number of basis functions. This method is intuitively appealing and easy to use in various nonparametric and semiparametric settings. In fact, an empirical researcher’s “flexible” parametric specification can often be given a nonparametric interpretation by invoking properly the series estimation theory.

The inference theory of series estimation is well understood in two broad settings; see, for example, Andrews (1991a), Newey (1997) and Chen (2007). The first is the semiparametric setting in which a researcher makes inference about a finite-dimensional parameter and/or a “regular” finite-dimensional functional of the nonparametric component. In this case, the finite-dimensional estimator has the parametric $n^{1/2}$ rate of convergence. The second setting pertains to the inference of “irregular” functionals of the nonparametric component, with the leading example being the pointwise inference for the unknown function, where the irregular functional evaluates the function at a given point. The resulting estimator has a slower nonparametric rate of convergence.

The uniform series inference for the unknown function, on the other hand, is a relatively open question. Unlike pointwise inference, a uniform inference procedure speaks to the global, instead of local, properties of the function. It is useful for examining functional features like monotonicity, convexity, symmetry and, more generally, function-form specifications, which are evidently of great empirical interest. In spite of its clear relevance, the uniform inference theory for series estimation appears to be “underdeveloped” in the current literature mainly due to the lack of asymptotic tools available to the econometrician, particularly in time-series analysis. Technically speaking, the asymptotic problem at hand involves a functional convergence that is non-Donsker, which is very different from Donsker-type functional central limit theorems commonly used in various areas of modern econometrics (Davidson (1994), van der Vaart and Wellner (1996), White (2001), Jacod and Shiryaev (2003), Jacod and Protter (2012)).

Recently, Chernozhukov, Lee, and Rosen (2013) and Belloni, Chernozhukov, Chetverikov, and
Kato (2015) have made important contributions on uniform series inference. The innovative idea underlying this line of research is to construct a strong Gaussian approximation for the functional series estimator, which elegantly circumvents the deficiency of the conventional “asymptotic normality” concept (formalized in terms of weak convergence) in this non-Donsker context. With independent data, the strong approximation for the functional estimator can be constructed using Yurinskii’s coupling which, roughly speaking, establishes the asymptotic normality for the sample mean of a “high-dimensional” data vector.\(^1\) The uniform series inference theory of Chernozhukov, Lee, and Rosen (2013) and Belloni, Chernozhukov, Chetverikov, and Kato (2015) relies on this type of coupling and, hence, are restricted to cross-sectional applications with independent data.\(^2\)

Set against this background, our initial contribution (see Section 3) is to develop a uniform inference theory for series estimators in time-series applications. To do so, we establish a novel strong approximation (i.e., coupling) result for general heterogeneous martingale difference arrays, which arise routinely from empirical applications on dynamic stochastic equilibrium models. Compared with the classical Yurinskii coupling for independent data, the key complication in the martingale-difference setting stems from the stochastic volatility of time series data, and we address this complication using a novel martingale technique. Armed with this coupling result, we then establish a uniform inference theory for series estimators in the time-series setting. We also show that the uniform inference can be conveniently used for nonparametrically testing conditional moment equalities implied by Euler or Bellman equations in dynamic stochastic equilibrium models.

In Section 4, we extend the aforementioned baseline theory in various directions. All these additional results are new to the literature, which further set our analysis apart from existing work. First, we extend our martingale-difference coupling result to general mixingales by developing a high-dimensional martingale approximation technique. It is well known that mixingales form a far more general class of processes than those characterized by various mixing concepts. For example, Andrews (1984, 1985) showed constructively that even nearly independent triangular arrays are not strongly mixing; as a result of the ranking of mixing coefficients, they are not \(\rho\)-mixing or \(\beta\)-mixing, either.\(^3\) By contrast, mixingales include martingale differences, ARMA processes, linear processes, linear processes,

\(^1\)In the present paper, we refer to a random vector as high-dimensional if its dimension grows to infinity with the sample size.

\(^2\)Yurinskii’s coupling concerns the strong approximation of a high-dimensional vector under the Euclidean distance. Chernozhukov, Chetverikov, and Kato (2014) establish a strong approximation for the largest entry of a high-dimensional vector under a more general setting.

\(^3\)On the other hand, Andrews’s examples are linear processes, which are special cases of mixingales. In probability theory, counterexamples of this kind can be traced back to Ibragimov and Linnik (1971) and Chernick (1981). See also Davidson (1994) for a comprehensive review on various dependence concepts used in econometrics.
various mixing and near-epoch dependent series as special cases. Therefore, our contribution is not limited to merely allowing for “some” specific form of dependence, but rather to provide a very general theory for essentially all types of dependence used in time-series econometrics. The general limit theorems developed here should be broadly useful for future research in nonparametric/high-dimensional time-series settings.

Second, in order to conduct feasible inference, we prove the validity of classical HAC estimators for long-run covariance matrices but in the current nonstandard setting with growing dimensions. This result is of independent econometric interest and may be useful in other inference problems as well.

Third, by using our general coupling and HAC estimation results, we develop an inference procedure based on intersection bounds (Chernozhukov, Lee, and Rosen (2013)) in the time-series setting. This method is broadly relevant for partially identified models and is particularly useful for testing conditional moment inequalities. To our knowledge, our result is the first for establishing the validity of intersection-bound-based inference in a general time-series setting.

Last but not least, we extend our theory on the least-square series estimation to a more general class of convex sieve M-estimations, with the main goal being the incorporation of the important nonparametric quantile series estimation. Perhaps not surprisingly, the aforementioned coupling and HAC estimation results also play a key role in establishing the uniform inference for quantile series estimation. The resulting uniform inference theory is broadly useful for the nonparametric estimation and specification test of conditional quantiles, such as value-at-risk and expected shortfall models used in financial risk management.

As a concrete empirical illustration of the proposed method, we study the unemployment volatility puzzle within the standard search and matching model (Pissarides (1985), Mortensen and Pissarides (1994), Pissarides (2000)). In an influential paper, Shimer (2005) shows that the standard Mortensen–Pissarides model, when calibrated in the conventional way, generates unemployment volatility that is far lower than the empirical estimate. Various modifications to the standard model have been proposed to address this puzzle; see Shimer (2004), Hall (2005), Mortensen and Nagypál (2007), Hall and Milgrom (2008), Pissarides (2009), and references therein. Hagedorn and Manovskii (2008), on the other hand, take a different route and show that the standard model actually can generate high levels of unemployment volatility using their alternative calibration strategy. The plausibility of their alternative calibration remains a contentious issue in the literature (see Hornstein, Krusell, and Violante (2005)). To shed some light on this debate from an econometric perspective, we derive a conditional moment restriction from the equilibrium Bellman equations. We then test whether this restriction holds or not at the parameter values
calibrated by Hagedorn and Manovskii (2008) using the proposed uniform inference method. The nonparametric specification test strongly rejects the hypothesis that these calibrated values are compatible with the equilibrium conditional moment restriction, and hence suggests that modifications to the standard Mortensen–Pissarides model are necessary for a better understanding of the cyclicality of unemployment. Constructively, we compute the Anderson–Rubin confidence set of parameter values which the test does not reject, which is informative about the “admissible” range of parameters for future research.


For conducting feasible inference, we extend the classical HAC estimation result in econometrics (see, e.g., Newey and West (1987), Andrews (1991b), Hansen (1992), de Jong and Davidson (2000)) to the setting with “large” long-run covariance matrices with growing dimensions. This result is of independent interest more generally for high-dimensional time-series inference.

Finally, on the statistical side, our strong approximation results for heterogeneous martingale difference arrays and (more generally) mixingales are related to the literature on high-dimensional coupling; in particular, we extend Yurinskii’s coupling (Yurinskii (1978)) from the independent data setting to a general time-series setting. The recent work of Chernozhukov, Lee, and Rosen (2013) relies on Yurinskii’s coupling and our (much) more general coupling theory helps extending their intersection-bound-based inference quite straightforwardly to a general time series setting.

4In a recent paper, Chen and Christensen (2018) establish the minimax sup-norm rate and strong approximation in nonparametric instrumental variables (NPIV) problems in the i.i.d. setting. We do not consider time-series NPIV problems in this paper, which may be interesting for future research.
Chernozhukov, Chetverikov, and Kato (2013a) constructed a strong approximation for the largest entry of a high-dimensional vector and Belloni, Chernozhukov, Chetverikov, and Kato (2015) applied this coupling to conduct uniform series inference. This alternative form of coupling is implied by Yurinskii’s coupling but can be obtained under weaker restrictions on the growth rate of the dimensionality. We focus purposefully on Yurinskii-type coupling so as to cover a wide range of empirical applications (such as the intersection-bound-based inference in Section 4.3), leaving the technical pursuit of weaker growth conditions on the dimensionality to future research. There has been limited research on high-dimensional coupling in the time-series setting. Chernozhukov, Chetverikov, and Kato (2013b) establish the strong approximation for the largest entry of a high-dimensional $\beta$-mixing sequence.\(^5\) As mentioned before, our coupling result is valid for general heterogeneous mixingales, which are far more general than mixing processes, and is thus free of the well-known critique of Andrews (1984, 1985).\(^6\) Regarding future research, our martingale approach is of further importance because it provides a necessary theoretical foundation for a more general theory involving discretized semimartingales that are widely used in the burgeoning literature of high-frequency econometrics (Aït-Sahalia and Jacod (2014), Jacod and Protter (2012)).\(^7\)

The paper is organized as follows. Section 2 provides a heuristic guidance of our econometric method in the context of several classical empirical examples. Section 3 represents the baseline econometric theory, which is further extended in Section 4 in various directions. The empirical application is given in Section 5. Section 6 concludes. All proofs for our theoretical results are in the supplemental appendix of this paper.

**Notations.** For any real matrix $A$, we use $\|A\|$ and $\|A\|_S$ to denote its Frobenius norm and spectral norm, respectively. We use $a^{(j)}$ to denote the $j$th component of a vector $a$; $A^{(i,j)}$ is defined similarly for a matrix $A$. For a random matrix $X$, $\|X\|_p$ denotes its $L_p$-norm, that is, $\|X\|_p = (\mathbb{E} \|X\|^p)^{1/p}$.

---

\(^5\) Zhang and Wu (2017) establish a similar coupling based on a notion of dependence obtained from stationary nonlinear systems (Wu (2005)).

\(^6\) Technically speaking, the martingale-based technique developed here is very different from the “large-block-small-block” technique employed in Chernozhukov, Chetverikov, and Kato (2013b), and it is necessitated by the distinct dependence structure studied in the present paper.

\(^7\) High-frequency asymptotic theory is mainly based on a version (see, e.g., Theorem IX.7.28 in Jacod and Shiryaev (2003)) of the martingale difference central limit theorem. The key difficulty for extending our coupling results further to the high-frequency setting is to accommodate non-ergodicity, which by itself is a very challenging open question.
2 Theoretical heuristics and motivating examples

In this section, we provide a heuristic discussion for our econometric method in the context of several “prototype” empirical examples. These examples consist of a broad range of macroeconomic and financial applications, including nonparametric estimation in empirical microstructure and specification tests based on Euler and Bellman equations in dynamic stochastic equilibrium models. Section 2.1 provides some background about strong approximation. Sections 2.2 and 2.3 discuss a battery of potential applications of our econometric method.

2.1 High-dimensional strong approximation

As discussed in the introduction, the main econometric contribution of the current paper concerns the uniform inference for series estimators in the time-series setting, for which the key (probabilistic) ingredient is a novel result for high-dimensional strong approximation. The issue of high dimensionality arises because series estimation involves “many” regressors. In this subsection, we introduce the notion of strong approximation and position it in the broad econometrics literature.

Consider a sequence $S_n$ of $m_n$-dimensional statistics defined on some probability space. We stress that the dimension $m_n$ is allowed to grow to infinity as $n \to \infty$. A sequence $\tilde{S}_n$ of random vectors, defined on the same probability space, is called a strong approximation of $S_n$ if

$$\|S_n - \tilde{S}_n\| = o_p(\delta_n) \quad (2.1)$$

for some real sequence $\delta_n \to 0$; we reserve the symbol $\delta_n$ for this role throughout.\(^8\) A useful special case is when the approximating variable $\tilde{S}_n$ has a Gaussian $\mathcal{N}(0, \Sigma_n)$ distribution with some $m_n \times m_n$ covariance matrix $\Sigma_n$, so that (2.1) formalizes a notion of “asymptotic normality” for the random vector $S_n$; we refer to $\Sigma_n$ as the pre-asymptotic covariance matrix of $S_n$. By contrast, in a conventional “textbook” setting with fixed dimension, the asymptotic normality is stated in terms of convergence in distribution (i.e., weak convergence), which in turn can be deduced by using a proper central limit theorem (Davidson (1994), White (2001), Jacod and Shiryaev (2003)). The conventional notion is evidently not applicable when the dimension of $S_n$ also grows asymptotically; indeed, the limiting variable would have a growing dimension and become a “moving target.”\(^9\)

\(^8\)We note that without specifying $\delta_n$, condition (2.1) is equivalent to $\|S_n - \tilde{S}_n\| = o_p(1)$. Indeed, a random real sequence $X_n = o_p(1)$ if and only if $X_n = o_p(\delta_n)$ for some real sequence $\delta_n \to 0$, although the convergence of the latter could be arbitrarily slow. The rate $\delta_n$ is needed explicitly for justifying feasible inference (by relying on anti-concentration inequalities) in the high-dimensional case.

\(^9\)Technically speaking, the limit theorem of interest here is non-Donsker. It is therefore fundamentally different from the strong invariance principle used by Mikusheva (2007), who considers the approximation for a partial sum.
An immediate nontrivial theoretical question is whether a strong approximation like (2.1) actually exists for general data generating processes. In the cross-sectional setting with independent data, Yurinskii’s coupling (Yurinskii (1978)) provides the strong approximation for sample moments. Establishing this result requires calculations that are more refined than those used for obtaining a “usual” central limit theorem for independent data; we refer the reader to Chapter 10 of Pollard (2001) for technical details. In principle, this limit theorem for sample moments can be extended to more general moment-based inference problems using the insight of Hansen (1982). As a first contribution in this direction, Chernozhukov, Lee, and Rosen (2013) and Belloni, Chernozhukov, Chetverikov, and Kato (2015) develop a uniform inference theory for the series estimator in the cross-sectional setting using Yurinskii’s coupling and a related extension by Chernozhukov, Chetverikov, and Kato (2014). In the present paper, we extend Yurinskii’s coupling to a general setting with dependent data so as to advance this line of econometric research towards time-series applications.

Before diving into the formal theory (see Section 3 and Section 4), we now proceed to illustrate the proposed econometric method in some classical empirical examples that emerge from various areas of empirical economics. Our goal is to provide some intuition underlying the theoretical construct in concrete empirical contexts so as to guide practical application.

### 2.2 Uniform inference for series estimators

The main focus of our econometric analysis is on the uniform inference for nonparametric series estimators constructed using dependent data. Uniform inference is useful in many cross-sectional problems. In this subsection, we provide examples for time-series applications so as to motivate directly our new theory.

Consider a nonparametric time-series regression model:

\[ Y_t = h(X_t) + u_t, \quad \mathbb{E}[u_t|X_t] = 0, \quad (2.2) \]

where the unknown function \( h(\cdot) \) is the quantity of econometric interest and the data series \((X_t, Y_t)_{1 \leq t \leq n}\) is generally serially dependent. We aim to make inference about the entire function \( h(\cdot) \) without relying on specific parametric assumptions. More precisely, the goal is to construct a confidence band \([\hat{L}_n(x), \hat{U}_n(x)]\) such that the uniform coverage probability

\[ P\left( \hat{L}_n(x) \leq h(x) \leq \hat{U}_n(x) \text{ for all } x \in \mathcal{X} \right) \quad (2.3) \]

process using a Brownian motion. In her case, the limiting law (induced by the Brownian motion) is fixed and the limit theorem is of Donsker type.
converges to a desired nominal level (say, 95%) in large samples.

A case in point is the relationship between volume \( Y \) and volatility \( X \) for financial assets. Since the seminal work of Clark (1973), a large literature has emerged for documenting and explaining the positive relationship between volume and volatility in asset markets; see, for example, Tauchen and Pitts (1983), Karpoff (1987), Gallant, Rossi, and Tauchen (1992), Andersen (1996), Bollerslev, Li, and Xue (2018) and references therein. Stylized microstructure models imply various specific functional relation between the expected volume and volatility (see, e.g., Kyle (1985), Kim and Verrecchia (1991), Kandel and Pearson (1995)). Gallant, Rossi, and Tauchen (1993) propose a nonparametric method for computing nonlinear impulse responses, which is adopted by Tauchen, Zhang, and Liu (1996) for studying the nonparametric volume–volatility relationship.

The series estimator \( \hat{h}_n(\cdot) \) of \( h(\cdot) \) is formed simply as the best linear prediction of \( Y_t \) given a growing number \( m_n \) of basis functions of \( X_t \), collected by \( P(X_t) \equiv (p_1(X_t), \ldots, p_{m_n}(X_t))^\top \). More precisely, we set \( \hat{h}_n(x) = P(x)^\top \hat{b}_n \), where \( \hat{b}_n \) is the least-square coefficient obtained from regressing \( Y_t \) on \( P(X_t) \), that is,

\[
\hat{b}_n = \left( \sum_{t=1}^{n} P(X_t) P(X_t)^\top \right)^{-1} \left( \sum_{t=1}^{n} P(X_t) Y_t \right). \tag{2.4}
\]

Unlike the standard least-square problem with fixed dimension, the dimension of \( \hat{b}_n \) grows asymptotically, which poses the key challenge for making uniform inference on the \( h(\cdot) \) function.

This issue can be addressed by using the strong approximation device. The intuition is as follows. Let \( b_n^* \) denote the “population” analogue of \( \hat{b}_n \) such that the approximation error \( r_n(x) \equiv h(x) - P(x)^\top b_n^* \) is close to zero uniformly in \( x \) as \( m_n \to \infty \); see Assumption 2(i) for the formal requirement. The sampling error of \( \hat{b}_n \) is measured by

\[
S_n = n^{1/2}(\hat{b}_n - b_n^*) = \left( n^{-1} \sum_{t=1}^{n} P(X_t) P(X_t)^\top \right)^{-1} \left( n^{-1/2} \sum_{t=1}^{n} P(X_t) u_t + n^{-1/2} \sum_{t=1}^{n} P(X_t) r_n(X_t) \right) \approx \left( n^{-1} \sum_{t=1}^{n} P(X_t) P(X_t)^\top \right)^{-1} n^{-1/2} \sum_{t=1}^{n} P(X_t) u_t. \tag{2.5}
\]

Based on the strong approximation for the sample average of \( P(X_t) u_t \), we can construct a strong Gaussian approximation \( \tilde{S}_n \) for \( S_n \) such that \( \tilde{S}_n \sim N(0, \Sigma_n) \). Since \( \hat{h}_n(x) = P(x)^\top \hat{b}_n \), the standard error of \( \hat{h}_n(x) \) is \( \sigma_n(x) = (P(x)^\top \Sigma_n P(x))^{1/2} \). We can further show that the standard error

\[10\] The price volatility is not directly observed. A standard approach in the recent literature is to use high-frequency realized volatility measures as a proxy.
function \( \sigma_n(\cdot) \) can be estimated “sufficiently well” by a sample-analogue estimator \( \tilde{\sigma}_n(\cdot) \), which generally involves a high-dimensional HAC estimator (see Section 4.2).

Taken together, these results eventually permit a strong approximation for the t-statistic process indexed by \( x \):

\[
\frac{n^{1/2} (\hat{h}_n(x) - h(x))}{\tilde{\sigma}_n(x)} = \frac{P(x)\top \tilde{S}_n}{\sigma_n(x)} + O_p(\delta_n),
\]

which is directly useful for feasible inference. The above coupling result shows clearly that the sampling variability of the t-statistics at various \( x \)’s is driven by the high-dimensional Gaussian vector \( \tilde{S}_n \) but with different loadings (i.e., \( P(x)/\sigma_n(x) \)). Importantly, (2.6) depicts the asymptotic behavior of \( \hat{h}(x) \) jointly across all \( x \)’s and, hence, provides the theoretical foundation for conducting uniform inference. The resulting econometric procedure is very easy to implement. It differs from a textbook linear regression only in the computation of critical values, which is detailed in Algorithm 1 in Section 3.2.

The nonparametric regression (2.2) can be easily modified to accommodate partially parameterized models, which is a notable advantage of series estimators compared with kernel-based alternatives; see Andrews (1991a) for a comprehensive discussion. We briefly discuss an important empirical example as a further motivation. Engle and Ng (1993) study the estimation of the news impact curve, which depicts the relation between volatility and lagged price shocks. Classical GARCH-type models (e.g., Engle (1982), Bollerslev (1986), Nelson (1991), etc.) typically imply specific parametric forms for the news impact curve. In order to “allow the data to reveal the curve directly (p. 1763),” Engle and Ng (1993) estimate a partially linear model of the form

\[
Y_t = aY_{t-1} + h(X_{t-1}) + u_t,
\]

where \( Y_t \) is the volatility, \( X_{t-1} \) is the price shock and the function \( h(\cdot) \) is the news impact curve. While the curve \( h(\cdot) \) is left fully nonparametric, this regression is partially parameterized in lagged volatility (via the term \( aY_{t-1} \)) as a parsimonious control for self-driven volatility dynamics. To estimate \( h(\cdot) \), we regress \( Y_t \) on \( Y_{t-1} \) and \( P(X_{t-1}) \) and obtain their least-square estimates \( \hat{a}_n \) and \( \hat{b}_n \), respectively. The nonparametric estimator for \( h(\cdot) \) is then \( \hat{h}_n(\cdot) = P(\cdot)\top \hat{b}_n \). The uniform inference for \( h(\cdot) \) can be done in the same way as in the fully nonparametric case. In the same vein, we can conduct uniform inference for nonparametric impulse responses using the local projection method proposed by Jordà (2005), which allows for a more flexible functional form in the conditioning state variable.\(^\text{11}\)

\(^{11}\)See Section II of Jordà (2005) for details.
2.3 Nonparametric specification tests for conditional moment restrictions

The uniform confidence band (recall (2.3)) can also be used conveniently for testing conditional moment restrictions against nonparametric alternatives. To fix idea, consider a test for the following conditional moment restriction

\[ E \left[ g \left( Y^*_t, \gamma_0 \right) \mid X_t \right] = 0, \quad (2.7) \]

where \( g(\cdot, \cdot) \) is a known function, \( Y^*_t \) is a vector of observed endogenous variables, \( X_t \) is a vector of observed state variables and \( \gamma_0 \) is a finite-dimensional parameter. To simplify the discussion, we assume for the moment that the test is performed with respect to a known parameter \( \gamma_0 \), and will return to the case with unknown \( \gamma_0 \) at the end of this subsection.

To implement the test, we cast (2.7) as a nonparametric regression in the form of (2.2) by setting \( Y_t = g(Y^*_t, \gamma_0) \), \( h(x) = \mathbb{E}[Y_t \mid X_t = x] \) and \( u_t = Y_t - h(X_t) \). Testing the conditional moment restriction (2.7) is then equivalent to testing whether the regression function \( h(\cdot) \) is identically zero. The formal test can be carried out by checking whether the “zero function” is covered by the uniform confidence band, that is,

\[ \hat{L}_n(x) \leq 0 \leq \hat{U}_n(x) \quad \text{for all} \quad x \in \mathcal{X}. \quad (2.8) \]

This procedure is in spirit analogous to the t-test used most commonly in applied work and it can reveal directly where (in terms of \( x \)) the conditional moment restriction is violated. We also note that, by the duality between tests and confidence sets, we can construct Anderson–Rubin type confidence sets for \( \gamma_0 \) by inverting the specification test.

Conditional moment restrictions are prevalent in dynamic stochastic equilibrium models. A leading example is from consumption-based asset pricing (see, e.g., Section 13.3 of Ljungqvist and Sargent (2012)), for which we set (with \( Y^*_t = (C_t, C_{t+1}, R_{t+1}) \))

\[ g \left( Y^*_t, \gamma_0 \right) = \frac{\delta u'(C_{t+1})}{u'(C_t)} R_{t+1}, \]

where \( R_{t+1} \) is the excess return of an asset, \( C_t \) is the consumption, \( \delta \) is the discount rate and \( u'(\cdot) \) is the marginal utility function parameterized by \( \gamma \). The variable \( X_t \) includes \( (R_t, C_t) \) and possibly other observed state variables.

The conditional moment restriction in the asset pricing example above, like in many other cases, is derived as the Euler equation in a dynamic program. More generally, it is also possible to derive conditional moment restrictions from a system of Bellman equations. Our empirical application on the search and matching model in Section 5 provides a concrete example of this type.
Finally, we return to the issue with unknown $\gamma_0$. In this case, $\gamma_0$ should be replaced by an estimated or, perhaps more commonly in macroeconomic applications, a calibrated value $\hat{\gamma}_n$. The feasible version of the test is then carried out using $Y_t = g(Y_t^*, \hat{\gamma}_n)$. As we shall show theoretically in Section 3.3, the estimation/calibration error in $\hat{\gamma}_n$ is asymptotically negligible under empirically plausible conditions. The intuition is straightforward: since $\gamma_0$ is finite-dimensional, its estimation/calibration error vanishes at a (fast) parametric rate, which is dominated by the sampling variability in the nonparametric inference with a (slow) nonparametric rate of convergence. Simply put, when implementing the nonparametric test, which is relatively noisy, one can treat $\hat{\gamma}_n$ effectively as $\gamma_0$ with negligible asymptotic consequences. This type of negligibility is not only practically convenient, but often necessary in macroeconomic applications for justifying formally the “post-calibration” inference. Indeed, the calibration may be done by following “consensus estimates” or is based on summary statistics provided in other papers (which themselves may rely on data sources that are not publicly available); in such cases, the limited statistical information from the calibration is insufficient for the econometrician to formally account for its sampling variability via standard sequential inference technique (e.g., Section 6 of Newey and McFadden (1994)). Our nonparametric test is, at least asymptotically, immune to this issue and, hence, provides a convenient but econometrically formal inference tool in this important type of empirical applications.

3 Baseline results on uniform inference

In order to streamline the discussion, we start in this section with our baseline theoretical results for uniform nonparametric time-series inference. Further extensions, which are substantially more general, are gathered in Section 4. Section 3.1 presents the strong approximation theorem for heterogeneous martingale differences. The uniform inference theory for series estimators in the time-series setting is presented in Section 3.2. Section 3.3 provides further results on how to use this uniform inference theory for testing conditional moment restrictions.

3.1 Strong approximation for martingale difference arrays

In this subsection, we present the strong approximation result for heterogeneous martingale difference arrays. This result serves as our first step for extending Yurinskii’s coupling, which is applicable for independent data, towards a general setting with serial dependency and heterogene-

---

12 We refer the reader to the comprehensive review of Dawkins, Srinivasan, and Whalley (2001) for discussions about estimation and calibration.
ity; a further extension to mixingales is in Section 4.1. We single out the result for martingale differences mainly because they arise routinely from dynamic stochastic equilibrium models that are equipped with information filtrations. Hence, this result is directly applicable in many economic applications. In addition, the inference for martingale differences does not involve the high-dimensional HAC estimation and, hence, permits a relatively simple implementation.

Fix a probability space \((\Omega, \mathcal{F}, \mathbb{P})\). We consider an \(m_n\)-dimensional square-integrable martingale difference array \((X_{n,t})_{1 \leq t \leq k_n, n \geq 1}\) with respect to a filtration \((\mathcal{F}_{n,t})_{1 \leq t \leq k_n, n \geq 1}\), where \(k_n \to \infty\) as \(n \to \infty\). That is, \(X_{n,t}\) is \(\mathcal{F}_{n,t}\)-measurable with finite second moment and \(\mathbb{E}[X_{n,t} | \mathcal{F}_{n,t-1}] = 0\). Let \(V_{n,t} \equiv \mathbb{E}[X_{n,t} X_{n,t}^\top | \mathcal{F}_{n,t-1}]\) denote the conditional covariance matrix of \(X_{n,t}\) and set

\[
\Sigma_{n,t} \equiv \sum_{s=1}^t \mathbb{E}[V_{n,s}].
\]

For example, in the setting of (2.5), \(k_n = n\) is the sample size and \(X_{n,t} = k_n^{-1/2} P(X_t) u_t\) represents a normalized version of the series \(P(X_t) u_t\); the order of magnitude of \(V_{n,t}\) is then \(k_n^{-1}\). For simplicity, we denote \(\Sigma_n \equiv \Sigma_{n,k_n}\) in the sequel.

Our goal is to construct a strong Gaussian approximation for the statistic

\[
S_n \equiv \sum_{t=1}^{k_n} X_{n,t}.
\]

In the conventional setting with fixed dimension, the classical martingale difference central limit theorem (see, e.g., Theorem 3.2 in Hall and Heyde (1980)) implies that

\[
S_n \overset{d}{\to} \mathcal{N}(0, \Sigma),
\]

where \(\Sigma = \lim_{n \to \infty} \Sigma_n\). In the present paper, however, we are mainly interested in the case with \(m_n \to \infty\). We aim to construct a coupling sequence \(\tilde{S}_n \sim \mathcal{N}(0, \Sigma_n)\) such that \(\|S_n - \tilde{S}_n\| = O_p(\delta_n)\) for some \(\delta_n \to 0\). The following assumption is needed.

**Assumption 1.** Suppose (i) the eigenvalues of \(k_n \mathbb{E}[V_{n,t}]\) are uniformly bounded from below and from above by some fixed positive constants; (ii) uniformly for any sequence \(h_n\) of integers that satisfies \(h_n \leq k_n\) and \(h_n/k_n \to 1\),

\[
\left\| \sum_{t=1}^{h_n} V_{n,t} - \Sigma_n h_n \right\|_S = O_p(r_n),
\]

where \(r_n\) is a real sequence such that \(r_n = o(1)\).
Condition (i) of Assumption 1 states that the random vector $X_{n,t}$ is non-degenerate. Condition (ii) requires the conditional covariance of the martingale $S_n$ (i.e., $\sum_{t=1}^{h_n} V_{n,t}$) to be close to its unconditional mean. This condition can be verified easily under mild weak-dependence type assumptions on the conditional covariance $V_{n,t}$.

We are now ready to state the strong approximation result for martingale difference arrays.

**Theorem 1.** Under Assumption 1, there exists a sequence $\tilde{S}_n$ of $m_n$-dimensional random vectors with distribution $\mathcal{N}(0, \Sigma_n)$ such that

$$\|S_n - \tilde{S}_n\| = O_p(m_n^{1/2} r_n^{1/2} + (B_n m_n)^{1/3}),$$

where $B_n \equiv \sum_{t=1}^{k_n} E[\|X_{n,t}\|^3]$.

Theorem 1 extends Yurinskii’s coupling towards general heterogeneous martingale difference arrays. In order to highlight the difference between these results, we describe briefly the construction underlying Theorem 1. Our proof consists of two steps. The first step is to construct another martingale $S_n^*$ whose conditional covariance matrix is exactly $\Sigma_n$ such that $\|S_n - S_n^*\| = O_p(m_n^{1/2} r_n^{1/2})$. This approximation step is not needed in the conventional setting with independent data, because in the latter case the conditional covariance process $V_{n,t}$ is nonrandom. Hence, the $O_p(m_n^{1/2} r_n^{1/2})$ error term is the “cost” for accommodating stochastic volatility.\(^{13}\) In the second step, we establish a strong approximation for $S_n^*$. Since the conditional covariance matrix of $S_n^*$ is engineered to be exactly $\Sigma_n$ (which is nonrandom), we can use a version of Lindeberg’s method and Strassen’s theorem for establishing the strong approximation. The resulting approximation error is $O_p((B_n m_n)^{1/3})$, which is essentially the rate of the classical Yurinskii’s coupling for independent data. The typical magnitude of $B_n$ is $O(k_n^{-1/2} m_n^{3/2})$ and, correspondingly, $(B_n m_n)^{1/3} = O(k_n^{-1/6} m_n^{5/6})$.

Finally, we note that an alternative form of strong approximation could be constructed, that is, a coupling for the largest entry of the vector $S_n$ instead of the vector itself; this alternative type of coupling is weaker than Yurinskii’s coupling (i.e., the former is implied by the latter), and may require less restrictions on the growth rate of the dimension $m_n$. That being said, we focus intentionally on the Yurinskii-type coupling in this paper for two related reasons. One is that establishing this type of coupling in the general time-series setting is evidently of its own independent theoretical importance and sets a relevant benchmark for future work. Indeed, some inference problems may actually require the strong approximation for the entire vector instead of

\(^{13}\)In order to construct $S_n^*$, we introduce a stopping time defined as the “hitting time (under the matrix partial order)” of the predictable covariation process $\sum_{s=1}^{t} V_{n,s}$ at the covariance matrix $\Sigma_n$. Condition (3.2) is used to establish an asymptotic lower bound for this stopping time, which in turn is needed for bounding the approximation error between $S_n^*$ and $S_n$. 

14
just its largest entry. The other reason is more “practical”: our results allow us to extend existing inference method that are based on Yurinskii-type coupling to a general time-series setting with minimum additional cost. For example, the theory of Chernozhukov, Lee, and Rosen (2013) relies on Yurinskii’s coupling, and we provide a time-series version of this method in Section 4.3, below. The pursuit of weaker conditions on the growth rate of $m_n$ is secondary to our main focus in the current paper, and is left for future research.

3.2 Uniform inference for nonparametric series regressions

In this subsection, we apply the coupling theorem above to develop an asymptotic theory for conducting uniform inference based on series estimation. We describe the implementation details for the procedure outlined in Section 2.2 and show its asymptotic validity.

Consider the following nonparametric regression model: for $1 \leq t \leq n$,

$$Y_t = h(X_t) + u_t \quad (3.4)$$

where $h(\cdot)$ is the unknown function to be estimated, $X_t$ is a random vector that may include lagged $Y_t$'s, and $u_t$ is an error term that satisfies

$$
\mathbb{E}[u_t|\mathcal{F}_{t-1}] = 0, \quad (3.5)
$$

where the information flow $\mathcal{F}_{t-1}$ is a $\sigma$-field generated by $\{X_s, u_{s-1}\}_{s \leq t}$ and possibly other variables. As described in Section 2.2, the series estimator of $h(x)$ is given by $\hat{h}_n(x) \equiv P(x)\top \hat{b}_n$, where $P(\cdot)$ collects the basis functions and $\hat{b}_n$ is the least-square coefficient obtained by regressing $Y_t$ on $P(X_t)$; recall (2.4).

We need some notations for characterizing the sampling variability of the functional estimator $\hat{h}_n(\cdot)$. The pre-asymptotic covariance matrix for $\hat{b}_n$ is given by $\Sigma_n \equiv Q_n^{-1}A_nQ_n^{-1}$, where

$$Q_n \equiv n^{-1} \sum_{t=1}^{n} \mathbb{E}\left[P(X_t)P(X_t)\top\right], \quad A_n \equiv \text{Var}\left[n^{-1/2} \sum_{t=1}^{n} u_t P(X_t)\right].$$

The pre-asymptotic standard error of $n^{1/2}(\hat{h}_n(x) - h(x))$ is thus

$$\sigma_n(x) \equiv \left(P(x)\top \Sigma_n P(x)\right)^{1/2}.$$

To conduct feasible inference, we need to estimate $\sigma_n(x)$, which amounts to estimating $Q_n$ and $A_n$. The $Q_n$ matrix can be estimated by

$$\hat{Q}_n \equiv n^{-1} \sum_{t=1}^{n} P(X_t)P(X_t)\top.$$
Since \( u_t \) forms a martingale difference sequence, \( A_n = n^{-1} \sum_{t=1}^{n} \mathbb{E}[u_t^2 P(X_t)P(X_t)'] \) and it can be estimated by
\[
\hat{A}_n = n^{-1} \sum_{t=1}^{n} \hat{u}_t^2 P(X_t)P(X_t)', \quad \text{where} \quad \hat{u}_t = Y_t - \hat{h}_n(X_t).
\] (3.6)

More generally, if we suppose only the mean independence assumption like in (2.2), then \( A_n \) is generally a (pre-asymptotic) long-run covariance matrix, and \( \hat{A}_n \) is the corresponding HAC estimator. This extension is discussed in details in Section 4.2 below.

With \( \hat{\Sigma}_n \equiv \hat{Q}^{-1}_n \hat{A}_n \hat{Q}^{-1}_n \), the estimator of \( \sigma_n(x) \) is given by
\[
\hat{\sigma}_n(x) \equiv \left( P(x)' \hat{\Sigma}_n P(x) \right)^{1/2}.
\]

Under some regularity conditions, we shall show (see Theorem 2) that the “sup-t” statistic
\[
\hat{T}_n \equiv \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} \left( \hat{h}_n(x) - h(x) \right)}{\hat{\sigma}_n(x)} \right|
\] (3.7)
can be (strongly) approximated by
\[
\tilde{T}_n \equiv \sup_{x \in \mathcal{X}} \left| \frac{P(x)' \tilde{\Sigma}_n^{1/2}}{\hat{\sigma}_n(x)} \right|, \quad \tilde{\Sigma}_n \sim \mathcal{N}(0, \Sigma_n).
\]

For \( \alpha \in (0, 1) \), the \( 1 - \alpha \) quantile of \( \tilde{T}_n \) can be used to approximate that of \( \hat{T}_n \). We can use Monte Carlo simulation to estimate the quantiles of \( \tilde{T}_n \), and then use them as critical values to construct uniform confidence bands for the function \( h(\cdot) \). Algorithm 1, below, summarizes the implementation details.

**Algorithm 1 (Uniform confidence band construction)**

Step 1. Draw \( m_n \)-dimensional standard normal vectors \( \xi_n \) repeatedly and compute
\[
\tilde{T}_n^* \equiv \sup_{x \in \mathcal{X}} \left| \frac{P(x)' \tilde{\Sigma}_n^{1/2} \xi_n}{\hat{\sigma}_n(x)} \right|.
\]

Step 2. Set \( cv_{n,\alpha} \) as the \( 1 - \alpha \) quantile of \( \tilde{T}_n^* \) in the simulated sample.

Step 3. Report \( \tilde{L}_n(x) = \hat{h}_n(x) - cv_{n,\alpha} \hat{\sigma}_n(x) \) and \( \tilde{U}_n(x) = \hat{h}_n(x) + cv_{n,\alpha} \hat{\sigma}_n(x) \) as the \( (1 - \alpha) \)-level uniform confidence band for \( h(\cdot) \).

We are now ready to present the asymptotic theory that justifies the validity of the confidence band described in the algorithm above. Although we consider the case with martingale-difference \( u_t \) error term here, the theory is valid in much more general settings as we show in Section 4.
To facilitate our later extensions, we collect the key ingredients of the theorem in the following high-level assumption. These conditions are either standard in the series estimation literature or can be verified using the limit theorems developed in the current paper.\footnote{In Supplemental Appendix S.B.4, we verify these high-level conditions under more primitive ones.} Below, we denote $\zeta_n^L \equiv \sup_{x_1, x_2 \in X} \|P(x_1) - P(x_2)\| / \|x_1 - x_2\|$. 

**Assumption 2.** For each $j = 1, \ldots, 4$, let $\delta_{j,n} = o(1)$ be a positive sequence. Suppose: (i) $\log(\zeta_n^L) = O(\log(m_n))$ and there exists a sequence $(b^*_n)_{n \geq 1}$ of $m_n$-dimensional constant vectors such that
\[
\sup_{x \in X} \left(1 + \|P(x)\|^{-1}\right) n^{1/2} \left|h(x) - P(x) b^*_n\right| = O(\delta_{1,n});
\]
(ii) the eigenvalues of $Q_n$ and $A_n$ are bounded from above and away from zero; (iii) the sequence $n^{-1/2} \sum_{t=1}^n P(X_t) u_t$ admits a strong approximation $\tilde{N}_n \sim \mathcal{N}(0, A_n)$ such that
\[
\left\|n^{-1/2} \sum_{t=1}^n P(X_t) u_t - \tilde{N}_n\right\| = O_p(\delta_{2,n});
\]
(iv) $\|\hat{Q}_n - Q_n\|_S = O_p(\delta_{3,n})$; (v) $\|\hat{A}_n - A_n\|_S = O_p(\delta_{4,n})$.

A few remarks on Assumption 2 are in order. Conditions (i) and (ii) are fairly standard in series estimation; see, for example, Andrews (1991a), Newey (1997), Chen (2007) and Belloni, Chernozhukov, Chetverikov, and Kato (2015). In particular, condition (i) specifies the precision for approximating the unknown function $h(\cdot)$ via basis functions, for which comprehensive results are available from numerical approximation theory. When $u_t$ is a martingale difference sequence, $X_{n,t} = n^{-1/2} P(X_t) u_t$ forms a martingale difference array, so the strong approximation in condition (iii) can be verified by using Theorem 1. More generally, this condition can be verified by using Theorem 4 below for mixingales. Conditions (iv) and (v) pertain to the convergence rates of $\hat{Q}_n$ and $\hat{A}_n$. In Section 4.2, we will develop convergence-rate results for $\hat{A}_n$ in a more general setting with HAC estimation. The conditions regarding $\hat{Q}_n$ can be verified in a similar, actually simpler, way.\footnote{In the special case when $u_t$ forms a martingale difference sequence, and the data series are strictly stationary and $\beta$-mixing, conditions (iv) and (v) can be directly verified by the theory developed in Chen and Christensen (2015).}

The asymptotic validity of the uniform confidence band $[\hat{L}_n(\cdot), \hat{U}_n(\cdot)]$ is justified by the following theorem.

**Theorem 2.** The following statements hold under Assumption 2:

(a) the sup-$t$ statistic $\hat{T}_n$ admits a strong approximation, that is, $\hat{T}_n = \tilde{T}_n + O_p(\delta_n)$ for
\[
\delta_n = \delta_{1,n} + \delta_{2,n} + m_n^{1/2} (\delta_{3,n} + \delta_{4,n});
\]
(b) if $\delta_n(\log m_n)^{1/2} = o(1)$ holds in addition, the uniform confidence band described in Algorithm 1 has asymptotic level $1 - \alpha$:

$$
\mathbb{P}\left( \tilde{L}_n(x) \leq h(x) \leq \tilde{U}_n(x) \text{ for all } x \in \mathcal{X} \right) \to 1 - \alpha.
$$

### 3.3 Specification test for conditional moment restrictions

In this subsection, we provide a formal discussion on the specification test outlined in Section 2.3. Recall that our econometric interest is to test conditional moment restrictions of the form

$$
\mathbb{E}\left[g(Y_t^*, \gamma_0) | X_t\right] = 0, \tag{3.8}
$$

where $g(\cdot)$ is a known function and $\gamma_0$ is a finite-dimensional parameter from a parameter space $\Upsilon \subseteq \mathbb{R}^d$. As discussed in Section 2.3, when $\gamma_0$ is known, we can cast the testing problem as a nonparametric regression by setting

$$
Y_t = g(Y_t^*, \gamma_0), \quad h(x) = \mathbb{E}[Y_t | X_t = x] \quad \text{and} \quad u_t = Y_t - \mathbb{E}[Y_t | X_t]. \tag{3.9}
$$

The test for (3.8) can then be carried out by examining whether the uniform confidence bound $[\tilde{L}_n(x), \tilde{U}_n(x)]$ covers the zero function (recall (2.8)).

Anderson–Rubin type confidence sets for $\gamma_0$ can also be constructed by inverting the tests.

The situation becomes somewhat more complicated when $\gamma_0$ is unknown, but a “proxy” $\hat{\gamma}_n$ is available; this proxy may be estimated by a conventional econometric procedure (e.g., Hansen (1982)) or calibrated from a computational experiment (Kydland and Prescott (1996)). For flexibility, we intentionally remain agnostic about how $\hat{\gamma}_n$ is constructed; in fact, we do not even assume that $\gamma_0$ is identified from the conditional moment restriction (3.8) which we aim to test. This setup is particularly relevant when $\hat{\gamma}_n$ is calibrated using a different data set (e.g., micro-level data) and/or based on an auxiliary economic model.

Equipped with $\hat{\gamma}_n$, we can implement the econometric procedure described in Section 3.2, except that we take $Y_t$ as the “generated” variable $g(Y_t^*, \hat{\gamma}_n)$. More precisely, we set

$$
\hat{b}_n = \left( n^{-1} \sum_{t=1}^{n} P(X_t) P(X_t)^\top \right)^{-1} \left( n^{-1} \sum_{t=1}^{n} P(X_t) g(Y_t^*; \hat{\gamma}_n) \right),
$$

16 This nonparametric test is similar in spirit to the test of Hardle and Mammen (1993). This method is distinct from Bierens-type tests (see, e.g., Bierens (1982) and Bierens and Ploberger (1997)) that are based on transforming the conditional moment restriction into unconditional ones using a continuum of instruments. These two approaches are complementary with their own merits.

17 It might be possible to refine the finite-sample performance of this “plug-in” procedure if additional structure about $\hat{\gamma}_n$ is available. We aim to establish a general approach for a broad range of applications, leaving specific refinements for future research.
\[ \hat{h}_n(x) = P(x)^\top b_n, \quad \tilde{u}_t = g(Y_t^*, \gamma_n) - \hat{h}_n(X_t) \] and then define \( \sigma_n(x) \) similarly as in Section 3.2. As alluded to previously (see Section 2.3), we aim to provide sufficient conditions such that replacing \( \gamma_0 \) with \( \gamma_n \) leads to negligible errors. The intuition is that the parametric proxy error in \( \gamma_n \) tends to be asymptotically dominated by the “statistical noise” in the nonparametric test.\(^{18}\) We formalize this intuition with a few assumptions.

**Assumption 3.** Conditions (i)-(iv) of Assumption 2 hold with \( h(x) = \mathbb{E}[g(Y_t^*, \gamma_0)|X_t = x] \) and \( u_t = g(Y_t^*, \gamma_0) - h(X_t) \), condition (v) of Assumption 2 holds for \( \hat{A}_n \) defined using \( \tilde{u}_t = g(Y_t^*, \gamma_n) - \hat{h}_n(X_t) \), and \( \delta_n(\log m_n)^{1/2} = o(1) \).

Assumption 3 allows us to cast the testing problem into the nonparametric regression setting of Section 3.2. These conditions can be verified in the same way as discussed above. However, this assumption is not enough for our analysis because condition (iii) pertains only to the strong approximation of the infeasible estimator defined using \( g(Y_t^*, \gamma_0) \) as the dependent variable. For this reason, we need some additional regularity conditions for closing the gap between the infeasible estimator and the feasible one. Below, we use \( g_\gamma(\cdot) \) and \( g_{\gamma\gamma}(\cdot) \) to denote the first and the second partial derivatives of \( g(y, \gamma) \) with respect to \( \gamma \), and we set

\[ G_n \equiv n^{-1} \sum_{t=1}^{n} \mathbb{E} \left[ P(X_t)g_\gamma(Y_t^*, \gamma_0)^\top \right], \quad H(x) \equiv \mathbb{E} \left[ g_{\gamma\gamma}(Y_t^*, \gamma_0) | X_t = x \right]. \]

**Assumption 4.** Suppose (i) for any \( y \), \( g(y, \gamma) \) is twice continuously differentiable with respect to \( \gamma \); (ii) there exists a positive sequence \( \delta_{5,n} \) such that \( \delta_{5,n}(\log m_n)^{1/2} = o(1) \) and

\[ n^{-1} \sum_{t=1}^{n} P(X_t)g_\gamma(Y_t^*, \gamma_0)^\top - G_n = O_p(\delta_{5,n}); \]

(iii) for some constant \( \rho > 0 \) and \( m_n \times d \) matrix-valued sequence \( \phi_n^\rho \), \( \sup_{x \in \mathcal{X}} \| P(x)^\top \phi_n^\rho - H(x) \| = O(m_n^{-\rho}); \)

(iv) \( \sup_{\gamma \in \mathcal{T}} n^{-1} \sum_{t=1}^{n} \| g_{\gamma\gamma}(Y_t^*, \gamma) \|^2 = O_p(1), \sup_{x \in \mathcal{X}} \| H(x) \| < \infty \) and \( \mathbb{E}[\| g_{\gamma\gamma}(Y_t^*, \gamma_0) \|^2] \)

is bounded; (v) \( \max_{1 \leq k \leq m_n} \sup_{x \in \mathcal{X}} | p_k(x) | \leq \zeta_n \) for a non-decreasing positive sequence \( \zeta_n = O(m_n^{n-1/2}); \)

(vi) \( \hat{\gamma}_n - \gamma_0 = O_p(n^{-1/2}); \)

(vii) \( \sup_{x \in \mathcal{X}} \| P(x) \|^{-1} = o((\log m_n)^{-1/2}) \) and \( \zeta_n m_n n^{-1/2} = o(1) \).

Conditions (i)–(v) of Assumption 4 jointly impose a type of (stochastic) smoothness for the moment functions with respect to \( \gamma \). These conditions are useful for controlling the effect of

\(^{18}\)While this “negligibility” intuition may be plausible for our nonparametric test (at least asymptotically), it is not valid for Bierens-type tests for which it is necessary to account for the sampling variability in the preliminary estimator \( \hat{\gamma}_n \). Therefore, when \( \hat{\gamma}_n \) is calibrated with limited statistical information to the econometrician, it is unclear how to formally justify Bierens-type tests.
the estimation error in $\hat{\gamma}_n$ on $\hat{h}_n(\cdot)$. Condition (vi) states that $\hat{\gamma}_n$ is a $n^{1/2}$-consistent estimator for $\gamma_0$, which is natural because the latter is finite-dimensional. Condition (vii) mainly reflects the fact that the standard error $\sigma_n(\cdot)$ of the nonparametric estimator is divergent due to the moderately growing number of series terms. These regularity conditions are imposed to formalize the simple intuition that the preliminary parametric estimation (or calibration) error is dominated by the relatively large sampling variability in the uniform nonparametric inference, and hence asymptotically negligible in the latter inference. The same techniques used in the verification of Assumption 2 can be applied to verify these conditions as well.

As a practical guide, we summarize the implementation details for the specification test in the following algorithm, followed by its theoretical justification.

**Algorithm 2 (Specification Test of Conditional Moment Restrictions)**

Step 1. Implement Algorithm 1 with $Y_t = g(Y^*_t, \hat{\gamma}_n)$ and obtain the sup-t statistic $\hat{T}_n$ and the critical value $cv_{n,\alpha}$.

Step 2. Reject the null hypothesis (3.8) at significance level $\alpha$ if $\hat{T}_n > cv_{n,\alpha}$. □

**Theorem 3.** Suppose that Assumptions 3 and 4 hold. Then under the null hypothesis (3.8), the test described in Algorithm 2 has asymptotic level $\alpha$. Under the alternative hypothesis that $\mathbb{E}[g(Y^*_t, \gamma_0)|X_t = x] \neq 0$ for some $x \in \mathcal{X}$, the test rejects with probability approaching one.

4 Extensions

We extend our baseline results in the previous section in various directions, which further set our analysis apart from the prior literature. Section 4.1 extends the martingale-difference strong approximation result towards the much more general class of mixingales. Section 4.2 establishes the validity of HAC estimators for mixingale data in the high-dimensional setting. Section 4.3 describes how to conduct inference for intersection bounds. Section 4.4 extends the least-square series estimation to a more general form of sieve M-estimation, which includes the nonparametric conditional quantile estimation as a special case.

4.1 Strong approximation for mixingales

Theorem 1 is restrictive for some time-series applications because martingale differences are serially uncorrelated. In this subsection, we extend that baseline coupling result towards mixingale processes by using a high-dimensional martingale approximation. Mixingales form a very general class of models, including martingale differences, linear processes and various types of mixing and
near-epoch dependent processes as special cases, and naturally allow for data heterogeneity. In particular, the mixingale concept is substantially more general than various mixing concepts (see, e.g., Andrews (1984, 1985) for counterexamples and Davidson (1994) for a comprehensive review) and readily accommodate most (if not all) applications in time series econometrics.

Turning to the formal setup, we consider an \( m_n \)-dimensional \( L_q \)-mixingale array \( (X_{n,t}) \) with respect to a filtration \( (\mathcal{F}_{n,t}) \) that satisfies the following conditions: for \( 1 \leq j \leq m_n \) and \( k \geq 0 \),

\[
\left\| \mathbb{E}[X_{n,t}^{(j)}|\mathcal{F}_{n,t-k}] \right\|_q \leq c_{n,t} \psi_k, \quad \left\| X_{n,t}^{(j)} - \mathbb{E}[X_{n,t}^{(j)}|\mathcal{F}_{n,t+k}] \right\|_q \leq c_{n,t} \psi_{k+1},
\]

where the constants \( c_{n,t} \) and \( \psi_k \) control the magnitude and the dependence of the \( X_{n,t} \) variables, respectively. We maintain the following assumption, where \( \bar{c}_n \) depicts the magnitude of \( k_n^{1/2} X_{n,t} \).

**Assumption 5.** The array \( (X_{n,t}) \) satisfies (4.1) for some \( q \geq 3 \). Moreover, for some positive sequence \( \bar{c}_n, \sup_t |c_{n,t}| \leq \bar{c}_n k_n^{-1/2} = O(1) \) and \( \sum_{k \geq 0} \psi_k < \infty \).

Assumption 5 allows us to approximate the partial sum of the mixingale \( X_{n,t} \) using a martingale. More precisely, we can represent

\[
X_{n,t} = X^*_n + \tilde{X}_{n,t} - \tilde{X}_{n,t+1}
\]

where \( X^*_n = \sum_{s=-\infty}^{\infty} \{ \mathbb{E}[X_{n,t+s}|\mathcal{F}_{n,t}] - \mathbb{E}[X_{n,t+s}|\mathcal{F}_{n,t-1}] \} \) forms a martingale difference and the “residual” variable \( \tilde{X}_{n,t} \) satisfies \( \sup_{j,t} \|X_{n,t}^{(j)}\|_2 = O(\bar{c}_n k_n^{-1/2}) \).

This representation further permits an approximation of \( S_n \) via the martingale \( S^*_n = \sum_{t=1}^{k_n} X^*_n, \) that is,

\[
\|S_n - S^*_n\|_2 = \|\tilde{X}_{n,1} - \tilde{X}_{n,k_n+1}\|_2 = O(\bar{c}_n m_n^{1/2} k_n^{-1/2}).
\]

In the typical case with \( \bar{c}_n = O(1) \), the approximation error in (4.3) is negligible as soon as the dimension \( m_n \) grows at a slower rate than \( k_n \). Consequently, a strong approximation for the martingale \( S^*_n \) (as described in Theorem 1) is also a strong approximation for \( S_n \). Theorem 4, below, formalizes this result.

**Theorem 4.** Suppose (i) Assumption 5 holds; (ii) Assumption 1 is satisfied for the martingale difference array \( X^*_n; \) and (iii) the largest eigenvalue of \( \Sigma_n \) is bounded. Then there exists a sequence \( \bar{S}_n \) of \( m_n \)-dimensional random vectors with distribution \( \mathcal{N}(0, \Sigma_n) \) such that

\[
\|S_n - \bar{S}_n\| = O_p(\bar{c}_n m_n^{1/2} k_n^{-1/2}) + O_p(m_n^{1/2} \bar{c}_n^{1/2} + (B_n^* m_n)^{1/3}) + O_p(\bar{c}_n m_n k_n^{-1/2} + \bar{c}_n^2 m_n^{3/2} k_n^{-1}),
\]

where \( \Sigma_n = \text{Var}(S_n) \) and \( B_n^* = \sum_{t=1}^{k_n} \mathbb{E}[\|X^*_{n,t}\|^3] \).

\(^{19}\)See Lemma A4 in the supplemental appendix for technical details about this approximation.
There are three types of approximation errors underlying this strong approximation result. The first term is due to the martingale approximation. The second term arises from the approximation of the martingale $S_n^*$ using a centered Gaussian variable $\tilde{S}_n^*$ with covariance matrix $\Sigma_n^* \equiv \mathbb{E}[S_n^* S_n^* \mathcal{]}$. The magnitude of this error is characterized by Theorem 1 as $O_p(\tilde{c}_n m^{1/2} n^{-1/2} + B_n^* m_n^{1/3})$. The third error component measures the distance between the two coupling variables $\tilde{S}_n^*$ and $\tilde{S}_n$, and is of order $O_p(\tilde{c}_n m^{1/2} m_n^{1/3} n^{-1/2} + \tilde{c}_2 m^{3/2} m_n^{1/3} n^{-1/2})$.

Theorem 4 can be used to verify the high-level condition (iii) in Assumption 2 and, hence, permits the uniform series inference in the general case in which the series $P(X_t) u_t$ is a mixingale. This setting arises when the error term $u_t$ is not a martingale difference, but only satisfies the mean independence condition $\mathbb{E}[u_t | X_t] = 0$. With $X_{n,t} \equiv n^{-1/2} P(X_t) u_t$, Theorem 4 implies that $n^{-1/2} \sum_{t=1}^n P(X_t) u_t$ can be strongly approximated by some Gaussian variable $\tilde{N}_n \sim \mathcal{N}(0, A_n)$, where $A_n \equiv \text{Var}(n^{-1/2} \sum_{t=1}^n P(X_t) u_t)$ is the long-run covariance matrix. Of course, we need a HAC estimator for $A_n$ in order to conduct feasible inference, to which we now turn.

### 4.2 High-dimensional HAC estimation

We have shown in Theorem 4 the strong approximation for the statistic $\sum_{t=1}^n X_{n,t}$ in a general time-series setting. In this subsection, we establish the asymptotic validity of a class of HAC estimators for its (long-run) covariance matrix $\Sigma_n = \text{Var}(\sum_{t=1}^n X_{n,t})$ that is needed for conducting feasible inference. This result can be used to verify Assumption 2(v) for constructing uniform confidence bands. Compared with the conventional setting on HAC estimation (see, e.g., Hannan (1970), Newey and West (1987), Andrews (1991b), Hansen (1992), de Jong and Davidson (2000), etc.), the main difference in our analysis is to allow the dimension $m_n$ to diverge asymptotically. Since the HAC estimation theory in the current high-dimensional setting is clearly of independent interest and may be used in other types of problems, we aim to build the theory in a general setting.

We study standard Newey–West type estimators. For each $s \in \{0, \ldots, k_n - 1\}$, define the sample covariance matrix at lag $s$, denoted $\tilde{\Gamma}_{X,n}(s)$, as

$$
\tilde{\Gamma}_{X,n}(s) \equiv \sum_{t=1}^{k_n-s} X_{n,t} X_{n,t+s}^\top
$$

and further set $\tilde{\Gamma}_{X,n}(-s) = \tilde{\Gamma}_{X,n}(s)^\top$. The HAC estimator for $\Sigma_n$ is then defined as

$$
\tilde{\Sigma}_n \equiv \sum_{s=-k_n+1}^{k_n-1} K(s/M_n) \tilde{\Gamma}_{X,n}(s)
$$

Moreover, in the current setting, feasible inference requires not only the consistency of the HAC estimator, but also a characterization of its rate of convergence (see Theorem 2(b)).
where $\mathcal{K}(\cdot)$ is a kernel smoothing function and $M_n$ is a bandwidth parameter that satisfies $M_n \to \infty$ as $n \to \infty$. The kernel function satisfies the following standard assumption.

**Assumption 6.** (i) $\mathcal{K}(\cdot)$ is bounded, Lebesgue-integrable, symmetric and continuous at zero with $\mathcal{K}(0) = 1$; (ii) for some constants $C \in \mathbb{R}$ and $r_1 \in (0, \infty]$, $\lim_{x \to 0}(1 - \mathcal{K}(x))/|x|^{r_1} = C$.\(^{21}\)

In order to analyze the limit behavior of $\tilde{\Gamma}_{X,n}(s)$ under general forms of serial dependence, we assume that the demeaned components of $X_{n,t}X_{n,t+j}^{\top}$ also behave like mixingales (recall (4.1)). More precisely, we maintain the following assumption.

**Assumption 7.** We have Assumption 5. Moreover, (i) for any $n > 0$, any $t$ and any $j$, $E[X_{n,t} X_{n,t+j}]$ only depends on $n$ and $j$; (ii) for all $j \geq 0$ and $s \geq 0$,

$$
\sup_t \max_{1 \leq l, k \leq m_n} \left\| \mathbb{E} \left[ X_{n,t}^{(l)} X_{n,t+j}^{(k)} | \mathcal{F}_{n,t-s} \right] - \mathbb{E} \left[ X_{n,t}^{(l)} X_{n,t+j}^{(k)} \right] \right\|_2 \leq c_n^2 k^{-1} \psi_s;
$$

(iii) $\sup_t \max_{1 \leq l, k \leq m_n} \left\| X_{n,t}^{(k)} X_{n,t+j}^{(l)} \right\|_2 \leq c_n^2 k^{-1}$ for all $j \geq 0$; (iv) $\sup_{s \geq 0} s \psi_s^2 < \infty$ and $\sum_{s=0}^{\infty} s^2 \psi_s < \infty$ for some $r_2 > 0$.

In this assumption, condition (i) imposes covariance stationarity on the array $X_{n,t}$ mainly for the sake of expositional simplicity. Condition (ii) extends the mixingale property from $X_{n,t}$ to the centered version of $X_{n,t} X_{n,t+j}^{\top}$.\(^{22}\) Conditions (iii) reflects that the scale of $k_n^{1/2} X_{n,t}$ is bounded by $\tilde{c}_n$. Condition (iv) specifies the level of weak dependence. The rate of convergence of the HAC estimator is given by the following theorem.

**Theorem 5.** Under Assumptions 6 and 7, $\| \tilde{\Sigma}_n - \Sigma_n \| = O_p\left( c_n^2 m_n (M_n k_n^{-1/2} + M_n^{-r_1/r_2}) \right)$.

**Comment.** Theorem 5 provides an upper bound for the convergence rate of the HAC estimator. It is interesting to note that, in the conventional setting with fixed $m_n$ and $\tilde{c}_n = O(1)$, the convergence rate is simply $O_p(M_n k_n^{-1/2} + M_n^{-r_1/r_2})$. In this special case, $\tilde{\Sigma}_n$ is a consistent estimator under the conditions $M_n k_n^{-1/2} = o(1)$ and $M_n \to \infty$, which are weaker than the requirement imposed by Newey and West (1987), Hansen (1992) and De Jong (2000). With $m_n$ diverging to infinity, the convergence rate slows down by a factor of $m_n$.

---

\(^{21}\)This condition holds for many commonly used kernel functions. For example, it holds with $(C, r_1) = (0, \infty)$ for the truncated kernel, $(C, r_1) = (1, 1)$ for the Bartlett kernel, $(C, r_1) = (6, 2)$ for the Parzen kernel, $(C, r_1) = (\pi^2/4, 2)$ for the Tukey-Hanning kernel and $(C, r_1) = (1.41, 2)$ for the quadratic spectral kernel. See Andrews (1991b) for more details about these kernel functions.

\(^{22}\)Generally speaking, the mixingale coefficient for $X_{n,t} X_{n,t+j}^{\top}$ may be different from that of $X_{n,t}$. Here, we assume that they share the same coefficient $\psi_s$ so as to simplify the technical exposition.
In many applications, we need to form the HAC estimator using “generated variables” that rely on some (possibly nonparametric) preliminary estimator. For example, specification tests described in Section 2.3 involve estimating/calibrating a finite-dimensional parameter in the structural model. In nonparametric series estimation problems, the HAC estimator is constructed using residuals from the nonparametric regression. We now proceed to extend Theorem 5 to accommodate generated variables.

We formalize the setup as follows. In most applications, the true (latent) variable $X_{n,t}$ has the form

$$X_{n,t} = k_n^{-1/2}g(Z_t, \theta_0),$$

where $Z_t$ is observed and $g(z, \theta)$ is a measurable function known up to a parameter $\theta$. The unknown parameter $\theta_0$ may be finite or infinite dimensional and can be estimated by $\hat{\theta}_n$. We use $\hat{X}_{n,t} = k_n^{-1/2}g(Z_t, \hat{\theta}_n)$ as a proxy for $X_{n,t}$. The feasible versions of (4.5) and (4.6) are then given by

$$\hat{\Gamma}_{X,n}(s) \equiv k_n^{-s} \sum_{t=1}^{k_n^{-1}} \hat{X}_{n,t} \hat{X}_{n,t+s}^\top, \quad \hat{\Gamma}_{X,n}(-s) = \hat{\Gamma}_{X,n}(s)^\top, \quad 0 \leq s \leq k_n - 1,$$

and $\hat{\Sigma}_n \equiv \sum_{s=-k_n+1}^{k_n-1} K(s/M_n) \hat{\Gamma}_{X,n}(s)$, respectively.

Theorem 6, below, characterizes the convergence rate of the feasible HAC estimator $\hat{\Sigma}_n$ when $\hat{\theta}_n$ is “sufficiently close” to the true value $\theta_0$; the latter condition is formalized as follows.

**Assumption 8.** (i) $k_n^{-1} \sum_{t=1}^{k_n} \|g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0)\|^2 = O_p(\delta_{\theta,n}^2)$ where $\delta_{\theta,n} = o(1)$ is a positive sequence; (ii) $\max_t \|g(Z_t, \theta_0)\|_2 = O(m_n^{1/2})$.

Assumption 8(i) is a high-level condition that embodies two types of regularities: the smoothness of $g(\cdot)$ with respect to $\theta$ and the convergence rate of the preliminary estimator $\hat{\theta}_n$. Quite commonly, $g(\cdot)$ is stochastically Lipschitz in $\theta$ and $\delta_{\theta,n}$ equals the convergence rate of $\hat{\theta}_n$. Sharper primitive conditions might be tailored in more specific applications. Assumption 8(ii) states that the $m_n$-dimensional vector is of size $O(m_n^{1/2})$ in $L_2$-norm, which holds trivially in most applications.

**Theorem 6.** Under Assumptions 6, 7 and 8, we have

$$\|\hat{\Sigma}_n - \Sigma_n\| = O_p \left( c_n^2 m_n (M_n k_n^{-1/2} + M_n^{-r_1/r_2}) \right) + O_p \left( M_n m_n^{1/2} \delta_{\theta,n} \right). \quad (4.7)$$

**Comments.** (i) The estimation error shown in (4.7) contains two components. The first term accounts for the estimation error in the infeasible estimator $\hat{\Sigma}_n$ and the second $O_p(M_n m_n^{1/2} \delta_{\theta,n})$ term is due to the difference between the feasible and the infeasible estimators. If the infeasible estimator is consistent, the feasible one is also consistent provided that $M_n m_n^{1/2} \delta_{\theta,n} = o(1)$.
(ii) The error bound in (4.7) can be further simplified when $\theta$ is finite-dimensional. In this case, one usually has $\delta_{\theta,n} = k_n^{-1/2}$. It is then easy to see that the second error component in (4.7) is dominated by the first. Simply put, the “plug-in” error resulted from using a parametric preliminary estimator $\hat{\theta}_n$ is negligible compared to the intrinsic sampling variability that is present even in the infeasible case with known $\theta_0$. When $\theta$ is infinite-dimensional, $\delta_{\theta,n}$ converges to zero at a rate slower than $k_n^{-1/2}$, and both error terms are potentially relevant.

Finally, as an illustration of the use of Theorem 6, we revisit the uniform inference theory formalized by Theorem 2. As discussed in Section 4.1, when $(u_t)$ does not form a martingale difference sequence, $A_n \equiv Var(n^{-1/2} \sum_{t=1}^n P(X_t)u_t)$ is generally a long-run covariance matrix. We can estimate $A_n$ using the HAC estimator described above. More specifically, we set

$$\hat{\Gamma}_n(s) \equiv n^{-1} \sum_{t=1}^{n-s} \hat{u}_t \hat{u}_{t+s} P(X_t)P(X_{t+s})^\top, \quad \hat{\Gamma}_n(-s) = \hat{\Gamma}_n(s)^\top,$$

and

$$\hat{A}_n \equiv n^{-1} \sum_{s=-n+1}^{n-1} K(s/M_n) \hat{\Gamma}_n(s). \quad (4.8)$$

Theorem 6 then provides the rate of convergence for this HAC estimator $\hat{A}_n$, as needed for verifying Assumption 2(v).

### 4.3 Time-series inference on intersection bounds

Our strong approximation results (i.e., Theorems 1 and 4) and the HAC estimation theory (i.e., Theorems 5 and 6) are not only useful for conducting uniform inference in nonparametric series estimation, but also useful in other problems. One important case in point is the estimation and inference concerning intersection bounds (Chernozhukov, Lee, and Rosen (2013)). This type of inference arises commonly from partial identification problems and is particularly relevant for testing conditional moment inequalities. In this subsection, we show how to use our time-series limit theorems to justify tests for conditional moment inequalities. To our knowledge, this provides the first result for intersection-bound-based inference in a general time-series setting.

We now turn to the details. Let $h(x) = \mathbb{E}[Y_t | X_t = x]$ be an unknown function and $\hat{h}_n(\cdot)$ be its nonparametric series estimator. A conditional moment inequality model specifies that $h(x) \geq 0$ for all $x \in \mathcal{X}$. This hypothesis amounts to $\eta^* \geq 0$, where

$$\eta^* = \inf_{x \in \mathcal{X}} h(x)$$

[23] With a mild extension, the method can be adapted to test for the monotonicity of unknown functions, which amounts to testing inequalities concerning derivative functions.
is the quantity of interest in the study of intersection bounds. Our statistical interest is to construct a \((1 - \alpha)\)-level upper confidence bound for \(\eta^*\), which we denote by \(\hat{\eta}_{n,1-\alpha}\). Chernozhukov, Lee, and Rosen (2013) propose an estimator of the form

\[
\hat{\eta}_{n,1-\alpha} \equiv \inf_{x \in \mathcal{X}} \left[ \hat{h}_n(x) + \hat{k}_{n,1-\alpha} \hat{\sigma}_n(x) \right]
\]

where \(\hat{k}_{n,1-\alpha}\) is a well-defined sequence of critical values and \(\hat{\sigma}_n(\cdot)\) is the estimator for the standard error function. Under certain high-level conditions, these authors show that (see Theorem 2 there)

\[
\liminf_{n \to \infty} P(\eta^* \leq \hat{\eta}_{n,1-\alpha}) \geq 1 - \alpha,
\]

that is, \(\hat{\eta}_{n,1-\alpha}\) is an asymptotically valid upper confidence bound at confidence level \(1 - \alpha\).

The key high-level assumption of Chernozhukov, Lee, and Rosen (2013) concerns the uniform strong approximation of the t-statistic process (see their Condition NS(i)(a)), that is,

\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} (\hat{h}_n(x) - h(x))}{\sigma_n(x)} - \frac{P(x) \top \bar{S}_n}{\sigma_n(x)} \right| = o_p \left( \frac{1}{\log(n)} \right),
\]

where \(\bar{S}_n \sim \mathcal{N}(0, \Sigma_n)\) and \(P(x)\) is the vector of series functions used in estimating \(h(x)\). It is worth noting that the condition above concerns the entire t-statistic process instead of only the sup-t statistic. Chernozhukov, Lee, and Rosen (2013) verify this strong approximation condition in the setting with independent data using Yurinskii’s coupling. By using Theorems 1 and 4, we can readily verify this condition for martingale differences and mixingales and, hence, justify the asymptotic validity of the confidence bound \(\hat{\eta}_{n,1-\alpha}\) in general time-series settings. This extension is formalized by Proposition 1 below. By convention, we assume that the first component function \(p_1(x)\) of \(P(x)\) satisfies \(p_1(x) = 1\).

**Proposition 1.** Suppose Assumption 2 holds. If we further have: (i) \(\max_{1 \leq j \leq m_n} \sup_{x \in \mathcal{X}} |p_j(x)| \leq \zeta_n\) where \(\zeta_n\) is a non-decreasing positive sequence; and (ii) \(\zeta_n^2 m_n \log(n)/n = o(1), \delta_{1,n} + \delta_{2,n} + m_n^{1/2} \delta_{3,n} = o(1/\log(n))\) and \(m_n^{1/2} (\delta_{3,n} + \delta_{4,n}) = n^{-b}\) where \(b > 0\) is a constant, then

\[
\liminf_{n \to \infty} P(\eta^* \leq \hat{\eta}_{n,1-\alpha}) \geq 1 - \alpha.
\]

**4.4 Uniform inference for sieve M-estimators**

In this subsection, we study the uniform inference for the unknown function \(h(\cdot)\) estimated by a type of sieve M-estimation, which generalizes the least-square series estimation studied in Section

\[\text{24}\]To avoid repetition, we refer the reader to Algorithm 1 in Chernozhukov, Lee, and Rosen (2013) for the construction of the critical value.
3.2. We suppose that \( h(\cdot) \) is identified as the unique minimizer of \( \mathbb{E}[n^{-1} \sum_{t=1}^{n} \rho(Z_t, h)] \) over a (functional) parameter space \( \mathcal{H} \), where \( Z_t = (Y_t^\top, X_t^\top)^\top \) is a random vector and \( \rho(\cdot, \cdot) \) is a loss function. The unknown function \( h(\cdot) \) is estimated by \( \hat{b}_n(x) = P(x)^\top \hat{b}_n \) where \( \hat{b}_n \) is defined as
\[
\hat{b}_n = \arg\min_{b \in \mathbb{R}^{m_n}} n^{-1} \sum_{t=1}^{n} \rho(Z_t, P(X_t)^\top b).
\]
(4.9)
We assume that the loss function \( \rho(z, h) \) is convex in \( h \) for any \( z \).

This type of sieve M-estimation includes the nonparametric least-square series estimation and the nonparametric quantile regression as special cases. Indeed, the series estimator studied in Section 3.2 corresponds to \( \rho(z, h) = (y - h(x))^2/2 \), while the \( q \)-quantile nonparametric series regression, \( q \in (0, 1) \), is obtained by setting \( \rho(z, h) = (q - 1_{\{y \leq h(x)\}})(y - h(x)) \). In the time-series context, the quantile regression is of great empirical interest, especially for risk management applications concerning the value-at-risk.

Parallel to the nonparametric least-square regression, the key to the uniform inference for the \( h(\cdot) \) function in the current setting is also to derive a strong approximation for the estimator \( \hat{b}_n \). To do so, we need some regularity conditions. We assume that \( \rho(z, P(x)^\top b) \) is differentiable in \( b \) except at finitely many points. We characterize the estimator \( \hat{b}_n \) using the following approximate first-order condition
\[
n^{-1} \sum_{t=1}^{n} w(Z_t, P(X_t)^\top \hat{b}_n) = O_p(\bar{\delta}_n)
\]
(4.10)
for some positive real sequence \( \bar{\delta}_n = o(1) \), where \( w(\cdot, \cdot) \) is a \( \mathbb{R}^{m_n} \)-valued function that defines the directional derivative of the mapping \( b \mapsto \rho(z, P(x)^\top b) \) with respect to \( b \). More specifically, \( w(z, h) = -(y - h(x))P(x) \) for the least-square regression and \( w(z, h) = (1_{\{y \leq h(x)\}} - q)P(x) \) for the nonparametric \( q \)-quantile regression. We also define
\[
\mu_n(b) \equiv n^{-1} \sum_{t=1}^{n} \left( w(Z_t, P(X_t)^\top b) - \mathbb{E} \left[ w(Z_t, P(X_t)^\top b) \right] \right),
\]
which is an (scaled) empirical process indexed by \( b \).

**Assumption 9.** We have (4.10) for some \( \bar{\delta}_n = o(\log(m_n)^{-1/2} n^{-1/2}) \). In addition, we suppose
(i) \( n^{-1/2} \sum_{t=1}^{n} w(Z_t, h) = O_p(m_n^{1/2}) \); (ii) there exists a sequence \( b_n^* \in \mathbb{R}^{m_n} \) such that
\[
n^{-1} \sum_{t=1}^{n} (w(Z_t, h) - w(Z_t, h_n)) = O_p(\bar{\delta}_n)
\]
where \( h_n(x) = P(x)^\top b_n^* \); (iii) \( \sup_{\|b - b_n^*\| \leq (\log(n)m_n/n)^{1/2}} \|\mu_n(b) - \mu_n(b_n^*)\| = O_p(\bar{\delta}_n) \); (iv) there exists a sequence of \( m_n \times m_n \) symmetric matrices \( Q_n \) such that
\[
\sup_{\|b - b_n^*\| \leq (\log(n)m_n/n)^{1/2}} \left\| n^{-1} \sum_{t=1}^{n} \mathbb{E} \left[ w(Z_t, P(X_t)^\top b) - w(Z_t, h_n) \right] - Q_n(b - b_n^*) \right\| = O(\bar{\delta}_n);
\]
the smallest eigenvalue of $Q_n$ is bounded away from zero.

Assumption 9(i) concerns the magnitude of the sample average of the vector score $w(Z_t, h)$. Since $E[w(Z_t, h)] = 0$, this condition is easily verified provided that the series $w(Z_t, h)$ exhibits sufficiently weak dependence. Assumption 9(ii) mainly concerns the convergence rate of series approximation of the unknown function $h$. The function $h_n(x) = P(x)^\top b_n^*$ is the pseudo-true value of $h$ in the finite-dimensional sieve space spanned by the approximating functions in $P(x)$. Assumption 9(iii) is a stochastic equicontinuity condition. Assumption 9(iv) is on the first-order expansion of the function $n^{-1} \sum_{t=1}^n E[w(Z_t, P(X_t)^\top b)]$ with respect to $b$ around $b_n^*$. The matrix $Q_n$ is the Hessian matrix of the population criterion function $n^{-1} \sum_{t=1}^n E[\rho(Z_t, P(X_t)^\top b)]$ in most cases. In the nonparametric regression $Q_n = n^{-1} \sum_{t=1}^n E[P(X_t) P(X_t)^\top]$, while in the nonparametric quantile regression $Q_n = n^{-1} \sum_{t=1}^n E[f_{Y_t|X_t}(h(X_t)) P(X_t) P(X_t)^\top]$, where $f_{Y_t|X_t}$ denotes the conditional density of $Y_t$ given $X_t$. Assumption 9(v) is a regularity condition commonly used in the series estimation literature.

Under Assumption 9, $\hat{b}_n$ admits an asymptotic linear representation that is akin to (2.5) obtained in the least-square case. The formal statement is given by the following proposition.

**Proposition 2.** Under Assumption 9,

$$n^{1/2}(\hat{b}_n - b_n^*) = -Q_n^{-1} n^{-1/2} \sum_{t=1}^n w(Z_t, h) + o_p(\log(m_n)^{-1/2}).$$

Given this linear representation of the estimation error $\hat{b}_n - b_n^*$, we can proceed in essentially the same way as in the least-square case for making uniform inference on the $h(\cdot)$ function. More specifically, we can construct a strong approximation for $n^{-1/2} \sum_{t=1}^n w(Z_t, h)$ with a Gaussian distribution $N(0, A_n)$, where $A_n \equiv \text{Var}(n^{-1/2} \sum_{t=1}^n w(Z_t, h))$. With $\hat{Q}_n$ and $\hat{A}_n$ being the estimators of $Q_n$ and $A_n$, we set $\hat{\Sigma}_n \equiv \hat{Q}_n^{-1} \hat{A}_n \hat{Q}_n^{-1}$ and $\hat{\sigma}_n(x) \equiv (P(x)^\top \hat{Q}_n P(x))^{1/2}$. The uniform confidence band of $h(\cdot)$ can (again) be constructed using the “sup-t” statistic

$$\hat{T}_n \equiv \sup_{x \in X} \left| \frac{n^{1/2}(\hat{h}_n(x) - h(x))}{\hat{\sigma}_n(x)} \right|$$

by following Algorithm 1. We omit the details to avoid repetition.

## 5 Empirical application on a search and matching model

### 5.1 The model and the equilibrium conditional moment restriction

This model has helped economists understand how regulation and economic policies affect unemployment, job vacancies, and wages. However, in an influential work, Shimer (2005) reports that the standard Mortensen–Pissarides model calibrated in the conventional way cannot explain the large volatility in unemployment observed in the data, that is, the unemployment volatility puzzle (Pissarides (2009)). A large literature has emerged to address this puzzle by modifying the standard model; see, for example, Shimer (2004), Hall (2005), Hall and Milgrom (2008), Mortensen and Nagypál (2007), Gertler and Trigari (2009), and Pissarides (2009), among others.

Hagedorn and Manovskii (2008), henceforth HM, take a different route to confront the Shimer critique. They demonstrate that the standard model actually can generate a high level of volatility in unemployment if the parameters are calibrated using their alternative calibration strategy. The key outcome of their calibration is a high value of nonmarket activity (i.e., opportunity cost of employment) that is very close to the level of productivity. Consequently, the fundamental surplus fraction is low (Ljungqvist and Sargent (2017)), resulting in a large elasticity of market tightness with respect to productivity, which in turn greatly improves the standard model’s capacity for generating unemployment volatility. By this logic, the Shimer critique to the standard model is less of a concern.

Whether this alternative calibration is plausible remains to be a contentious issue in the literature. For example, Hall and Milgrom (2008) state that HM’s calibrated nonmarket return would imply too high an elasticity of labor supply. Costain and Reiter (2008), cited by Pissarides (2009), argue that HM’s calibration would imply effects of the unemployment insurance policy much higher than empirical estimates. While these critiques are sound in principle, the actual quantitative statements invariably rely on additional economic or econometric assumptions, bringing in new quantities that can be equally difficult to calibrate or to estimate. Indeed, Chodorow-Reich and Karabarbounis (2016) demonstrate that, depending on the specific auxiliary assumptions used in calibration, the value of nonmarket activity can range quite wildly.

We aim to shed some light on this debate from an econometric point of view. Rather than resorting to some “external” calibration target, we rely on a conditional moment restriction that arises “internally” from the equilibrium Bellman equations. Specifically, we apply the proposed nonparametric test as described in Subsection 3.3, in order to examine whether the calibrated parameters are compatible with the equilibrium conditional moment restriction.

Turning to the details, we first briefly restate HM’s version of the standard Mortensen–Pissarides model with aggregate uncertainty. Time is discrete. There is a unit measure of infinitely lived workers and a continuum of infinitely lived firms. The workers maximize their expected lifetime utility and the firms maximize their expected profit. Workers and firms share the same
discount factor $\delta$. The only source of aggregate shock is the labor productivity $p_t$ (i.e., the output per each unit of labor), which follows a Gaussian AR(1) model in log level.

Workers can either be unemployed or employed. An unemployed worker gets flow utility $z$ from nonmarket activity and searches for a job. As alluded to above, the value of nonmarket activity $z$ is the key parameter of interest, because it determines the fundamental surplus fraction in the standard model (Ljungqvist and Sargent (2017)). Firms attract workers by maintaining an open vacancy at flow cost $c_p$, parameterized as a function of productivity.

The number of new matches is determined by the level of unemployment $u_t$ and the number of vacancies $v_t$ through the matching function $m(u_t, v_t) = u_tv_t/(u_t^l + v_t^l)^{1/l}$ for some matching parameter $l > 0$ (see den Haan, Ramey, and Watson (2000)). The key quantity in the search and matching model is the market tightness $\theta_t \equiv v_t/u_t$. The job finding rate and the vacancy filling rate are given by, respectively, $f(\theta_t) \equiv m(u_t, v_t)/u_t$ and $q(\theta_t) \equiv m(u_t, v_t)/v_t$. Matched firms and workers separate exogenously with probability $s$ per period. There is free entry of firms, which drives the expected present value of an open vacancy to zero. Matched firms and workers split the surplus according to the generalized Nash bargaining solution. The workers’ bargaining power is $\beta \in (0, 1)$.

We now describe the equilibrium of this model and derive from it a conditional moment restriction on observed data. Denote the firm’s value of a job by $J$, the firm’s value of an unfilled vacancy by $V$, the worker’s value of having a job by $W$, the worker’s value of being unemployed by $U$ and the wage by $w$; these quantities are functions of the state variable in equilibrium. Following the convention of macroeconomics, for a generic variable $X$, let $E_p[X_{p'}]$ denote the one-period ahead conditional expectation of $X$ given the current productivity $p$. The equilibrium is characterized by the following Bellman equations:

$$
\begin{align*}
J_p &= p - w_p + \delta (1 - s) E_p [J_{p'}] \\
V_p &= -c_p + \delta q(\theta_p) E_p [J_{p'}] \\
U_p &= z + \delta \left\{ f(\theta_p) E_p [W_{p'}] + (1 - f(\theta_p)) E_p [U_{p'}] \right\} \\
W_p &= w_p + \delta \left\{ (1 - s) E_p [W_{p'}] + s E_p [U_{p'}] \right\}.
\end{align*}
$$

(5.1) \hspace{1cm} (5.2) \hspace{1cm} (5.3) \hspace{1cm} (5.4)

The model is closed by imposing free-entry and Nash bargaining, corresponding to $V_p = 0$ and $J_p = (W_p - U_p)(1 - \beta)/\beta$, respectively.

From these equilibrium conditions, we can solve the functions $J_p$, $V_p$, $U_p$, $W_p$ and $w_p$ in terms of $\theta_p$, and then reduce the system into one functional equation.\footnote{The detailed derivation is given in Supplemental Appendix S.A.8.} Instead of solving the fixed-point
problem, we replace $p$ and $\theta$ with their observed time series, yielding the following the equilibrium conditional moment restriction

$$E[\zeta_{t+1} - z | p_t] = 0,$$

(5.5)

where, for ease of notation, we define (with $c_t$ denoting $c_{pt}$)

$$\zeta_{t+1} \equiv p_{t+1} - \frac{\beta \theta_{t+1} c_{t+1}}{1 - \beta} + \frac{(1 - s) c_{t+1}}{(1 - \beta) q(\theta_{t+1})} - \frac{c_t}{(1 - \beta) \delta q(\theta_t)}. \tag{5.6}$$

Below, we apply the proposed nonparametric test on this conditional moment restriction.

### 5.2 Empirical results

We start with testing whether the equilibrium conditional moment restriction (5.5) holds or not for the parameters calibrated by HM.\(^{26}\) It is instructive to briefly recall their calibration strategy, which involves two stages. All parameters except for ($z, \beta, l$) are calibrated by matching certain empirical quantities in the first stage. The second stage further pins down these three parameters by matching model-implied wage-productivity elasticity, average job finding rate, and average market tightness with their empirical estimates, which is the more contentious part of the calibration (Hornstein, Krusell, and Violante (2005)). For this reason, we focus on these key parameters so as to directly speak to the core of the unemployment volatility puzzle. The value of nonmarket activity $z$ is of particular importance because it is the sole determinant of the fundamental surplus fraction in the standard Mortensen–Pissarides model (Ljungqvist and Sargent (2017)). We use the same data from 1951 to 2004 as in Hagedorn and Manovskii (2008).\(^{27}\)

Figure 1a shows the scatter of the residual $\zeta_{t+1} - z$ in the moment condition (5.5) versus the conditioning variable $p_t$. Under the equilibrium conditional moment restriction, $\zeta_{t+1} - z$ should be centered around zero conditional on each level of $p_t$, and there should be no correlation pattern between these variables. In contrast, we find that the scatter of $\zeta_{t+1} - z$ is centered below zero, suggesting that $z$ is too high given the other calibrated parameters. In addition, there appears to be a mild positive relationship between the residual and productivity. These patterns are more clearly revealed by the nonparametric fit of $E[\zeta_{t+1} - z | p_t]$, displayed as the solid line. The uniform confidence band of the conditional moment function does not cover zero for a wide range of productivity levels, indicating a strong rejection (with the p-value being virtually zero) of the equilibrium conditional moment restriction given the calibrated parameter values.

\(^{26}\)The calibrated parameters play the role of $\hat{\gamma}_n$ in the setting of Section 3.3.

\(^{27}\)The data is obtained from the publisher’s website. The $p_t$ and $\theta_t$ variables are measured using their cyclical component obtained from the Hodrick–Prescott filter with smoothing parameter 1600. The calibrated parameters are adjusted to the quarterly frequency. We refer the reader to Hagedorn and Manovskii (2008) for additional information about their data and calibration.
Figure 1: Nonparametric test for equilibrium conditional moment restriction. Panel (a) shows
the scatter of the residual of the equilibrium conditional moment restriction $\zeta_{t+1} - z$ versus the
productivity $p_t$, the nonparametric fit (solid) and the 95% uniform two-sided confidence band
(dashed). The series estimator (solid) is computed using a cubic polynomial and the standard
error is computed under the martingale difference assumption implied by the conditional moment
restriction. Panel (b) plots the 95% Anderson-Rubin confidence set for value of nonmarket activity
$z$, worker’s bargaining power $\beta$, and matching parameter $l$.

We next ask a more constructive question: Which parameter values, if there are any, are
compatible with the equilibrium conditional moment restriction (5.5)? To answer this question
formally, we construct the Anderson–Rubin confidence set for $(z, \beta, l)$ obtained by inverting the
nonparametric specification test, while fixing the other parameters at their calibrated values.28
Figure 1b shows the 3-dimensional 95%-level confidence set. The confidence set is far away from
empty, suggesting that the equilibrium conditional moment restriction is compatible with the
data for a wide range of parameter values and, to this extent, is not overly restrictive. But we
also see that “admissible” values of $z$ is generally notably lower than the calibrated value 0.955.
By the theory of Ljungqvist and Sargent (2017), a mild decrease in $z$ can significantly reduce
the fundamental surplus fraction. For example, changing $z$ from 0.955 to 0.9 will reduce the
fundamental surplus fraction by more than half, and hence causes the unemployment volatility to
drop by a similarly amount. Our finding thus suggests that the unemployment volatility puzzle
remains a puzzle for the standard model once we insist that the parameters—particularly the value

28The inversion is implemented by using a (standard) grid search: We consider $z \in [0.01, 0.99]$, $\beta \in [0.01, 0.2]$, $l \in [0.3, 0.5]$, and discretize these intervals with mesh size 0.001.
of nonmarket activity—are econometrically compatible with the equilibrium conditional moment restriction.

6 Conclusion

We develop a uniform inference theory for nonparametric series estimators in time-series settings. While the pointwise inference problem has been addressed in the literature, uniform series inference in the time-series setting remains an open question to date. The uniform inference theory relies crucially on our novel strong approximation theory for heterogeneous dependent data with growing dimensions. To conduct feasible inference, we also develop a HAC estimation theory in the high-dimensional setting. Further results concerning the inference on intersection bounds and convex sieve M-estimation are also developed. The proposed inference procedure is easy to implement and is broadly applicable in a wide range of empirical problems in economics and finance. The technical results on strong approximation and HAC estimation also provide theoretical tools for other econometric problems involving high-dimensional data vectors.

References


Supplemental Appendix to
Uniform Nonparametric Inference for Time Series

Jia Li* and Zhipeng Liao†

May 14, 2019

Abstract
Supplemental Appendix S.A contains the proofs for all results in the main text. Supplemental Appendix S.B contains additional technical results on the verification of high-level conditions using more primitive ones.

*Department of Economics, Duke University, Durham, NC 27708; e-mail: jl410@duke.edu.
†Department of Economics, UCLA, Los Angeles, CA 90095; e-mail: zhipeng.liao@econ.ucla.edu.
S.A Appendix: Proofs

For any real matrix $A$, we use $\|A\|$ and $\|A\|_S$ to denote its Frobenius norm and spectral norm, respectively. If $A$ is a real square matrix, we denote its trace, the smallest and the largest eigenvalues by $\text{Tr}(A)$, $\lambda_{\text{min}}(A)$ and $\lambda_{\text{max}}(A)$, respectively. We use $a^{(j)}$ to denote the $j$th component of a vector $a$; $A^{(i,j)}$ is defined similarly for a matrix $A$. For a random matrix $X$, $\|X\|_p$ denotes its $L_p$-norm, that is, $\|X\|_p = (\mathbb{E} \|X\|^p)^{1/p}$. For any two positive sequences $a_n$ and $b_n$, $a_n \ll b_n$ means that $a_n = o(b_n)$. For any two real constants $a$ and $b$, $a \wedge b = \min\{a, b\}$. Throughout the proofs, we use $K$ to denote a generic constant that may change from line to line.

S.A.1 Proof of Theorem 1

The proof of Theorem 1 consists of two steps. The first step is to approximate $S_n$ with another martingale $S_n^*$ whose conditional covariance matrix is exactly $\Sigma_n$; see Lemma A1. We then establish the coupling between $S_n^*$ and $\tilde{S}_n$ by using Lindeberg’s method and Strassen’s theorem; see Lemma A2.

Turning to the details, we start with describing the approximating martingale $S_n^*$. Consider the following stopping time:

$$
\tau_n \equiv \max \left\{ t \in \{1, \ldots, k_n\} : \Sigma_n - \sum_{s=1}^{t} V_{n,s} \text{ is positive semi-definite} \right\},
$$

with the convention that $\max \emptyset = 0$. We note that $\tau_n$ is a stopping time because $V_{n,t}$ is $\mathcal{F}_{n,t-1}$-measurable for each $t$ and $\Sigma_n$ is nonrandom. The matrix

$$
\xi_n \equiv \begin{cases} 
\Sigma_n & \text{when } \tau_n = 0, \\
\Sigma_n - \sum_{t=1}^{\tau_n} V_{n,t} & \text{when } \tau_n \geq 1,
\end{cases}
$$

is positive semi-definite by construction.

Let $K_n$ be a sequence of integers such that $K_n \to \infty$ and let $(\eta_{n,t})_{k_n+1 \leq t \leq k_n+K_n}$ be independent $m_n$-dimensional standard normal vectors. We construct another martingale difference array $(Z_{n,t}, \mathcal{H}_{n,t})_{1 \leq t \leq k_n + K_n}$ as follows:

$$
Z_{n,t} \equiv \begin{cases} 
X_{n,t} 1\{t \leq \tau_n\} & \text{when } 1 \leq t \leq k_n, \\
K_n^{-1/2} \xi_n^{1/2} \eta_{n,t} & \text{when } k_n + 1 \leq t \leq k_n + K_n,
\end{cases}
$$

and the filtration is given by

$$
\mathcal{H}_{n,t} \equiv \begin{cases} 
\mathcal{F}_{n,t} & \text{when } 1 \leq t \leq k_n, \\
\mathcal{F}_{n,k_n} \vee \sigma(\eta_{n,s} : s \leq t) & \text{when } k_n + 1 \leq t \leq k_n + K_n.
\end{cases}
$$

Since $\tau_n$ is a stopping time, it is easy to verify that $(Z_{n,t}, \mathcal{H}_{n,t})_{1 \leq t \leq k_n + K_n}$ indeed forms a martingale difference array. We denote

$$
V_{n,t}^* = \mathbb{E} \left[ Z_{n,t} Z_{n,t}^\top \bigg| \mathcal{H}_{n,t-1} \right] \quad (A.1)
$$
and set

\[ S^*_n = \sum_{t=1}^{k_n + K_n} Z_{n,t}. \]  \hfill (A.2)

The conditional covariance matrix of \( S^*_n \) is exactly \( \Sigma_n \), that is,

\[ \sum_{t=1}^{k_n + K_n} V^*_{n,t} = \sum_{t=1}^{\tau_n} V_{n,t} + \xi_n = \Sigma_n. \]  \hfill (A.3)

Lemma A1, below, quantifies the approximation error between \( S_n \) and \( S^*_n \).

**Lemma A1.** Suppose that Assumption 1 holds. Then, \( \| S_n - S^*_n \| = O_p(m_n^{1/2} r_n^{1/2}) \).

**Proof of Lemma A1.** Step 1. In this step, we show that for any \( \varepsilon > 0 \), there exists a finite constant \( C_1 > 0 \) such that, for \( u^*_n = [C_1 r_n k_n] \) and \( h^*_n = k_n - u^*_n \),

\[ \limsup_{n \to \infty} \mathbb{P}(\tau_n < h^*_n) < \varepsilon. \]  \hfill (A.4)

Fix \( \varepsilon > 0 \). By Assumption 1(ii), there exists a finite constant \( C_2 > 0 \) such that for any \( h_n \leq k_n \) satisfying \( h_n/k_n \to 1 \),

\[ \limsup_{n \to \infty} \mathbb{P} \left( \lambda_{\max} \left( \sum_{t=1}^{h_n} V_{n,t} - \Sigma_n, h_n \right) > C_2 r_n \right) < \varepsilon. \]  \hfill (A.5)

Let \( \lambda > 0 \) denote a lower bound for the eigenvalues as described in Assumption 1(i). We shall show that (A.4) holds for \( C_1 \equiv C_2/\lambda \).

Since \( r_n = o(1) \) by Assumption 1(ii), we have \( u^*_n/k_n \to 0 \) and \( h^*_n/k_n \to 1 \). In particular, (A.5) holds for \( h_n = h^*_n \). Moreover, observe that

\[ \frac{u^*_n}{r_n k_n} = \frac{[C_1 r_n k_n]}{r_n k_n} \geq \frac{C_1}{\lambda}, \]

which, together with the definition of \( \lambda \), implies that

\[ C_2 r_n \leq \frac{u^*_n}{k_n} \lambda \leq \lambda_{\min} \left( \sum_{t=h^*_n+1}^{k_n} \mathbb{E}[V_{n,t}] \right). \]  \hfill (A.6)

We then observe

\[ \mathbb{P}(\tau_n < h^*_n) \leq \mathbb{P} \left( \lambda_{\max} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n, h^*_n \right) > 0 \right) \]

\[ = \mathbb{P} \left( \lambda_{\max} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n, h^*_n \right) - (\Sigma_n - \Sigma_n, h^*_n) > 0 \right) \]

\[ \leq \mathbb{P} \left( \lambda_{\max} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n, h^*_n \right) > \lambda_{\min} \left( \sum_{t=h^*_n+1}^{k_n} \mathbb{E}[V_{n,t}] \right) \right) \]

\[ \leq \mathbb{P} \left( \lambda_{\max} \left( \sum_{t=1}^{h^*_n} V_{n,t} - \Sigma_n, h^*_n \right) > C_2 r_n \right), \]  \hfill (A.7)
where the first inequality follows from the definition of $\tau_n$, the second inequality follows from the property of eigenvalues and the last inequality is by (A.6). From (A.5) and (A.7), the claim (A.4) readily follows.

**Step 2.** We now prove the assertion of Lemma A1. Note that

$$S_n - S_n^* = \sum_{t=1}^{k_n} X_{n,t} 1_{\{t > \tau_n\}} - K_n^{-1/2} \xi_n^{1/2} \sum_{t=k_n+1}^{k_n+K_n} \eta_{n,t}. $$

Hence, it suffices to show

$$\sum_{t=1}^{k_n} X_{n,t} 1_{\{t > \tau_n\}} = O_p(m_n^{1/2} r_n^{1/2}), \quad K_n^{-1/2} \xi_n^{1/2} \sum_{t=k_n+1}^{k_n+K_n} \eta_{n,t} = O_p(m_n^{1/2} r_n^{1/2}).$$

(A.8)

Recall $u_n^*$ and $h_n^*$ from step 1. By the assertion of step 1, we can assume that $\tau_n \geq h_n^*$ without loss of generality; otherwise, we can restrict attention to the event $\{\tau_n \geq h_n^*\}$ with the exceptional probability made arbitrarily small.

Since $\tau_n$ is a stopping time, $\{t > \tau_n\} \in \mathcal{F}_{n,t-1}$. Therefore, $(X_{n,t} 1_{\{t > \tau_n\}})_{t \geq 1}$ are martingale differences. It is then easy to see that

$$\mathbb{E} \left[ \left\| \sum_{t=1}^{k_n} X_{n,t} 1_{\{t > \tau_n\}} \right\|^2 \right] = \mathbb{E} \left[ \sum_{t=1}^{k_n} \|X_{n,t}\|^2 1_{\{t > \tau_n\}} \right] \leq \sum_{t=h_n^*+1}^{k_n} \mathbb{E} \left[ \|X_{n,t}\|^2 \right] = \text{Tr} \left( \sum_{t=h_n^*+1}^{k_n} \mathbb{E} [V_{n,t}] \right).$$

By Assumption 1(i), the majorant side of the above inequality is $O(u_n^* m_n / k_n) = O(m_n r_n)$. The first assertion in (A.8) then readily follows.

Turning to the second assertion in (A.8), we note that

$$\mathbb{E} \left[ \left\| K_n^{-1/2} \xi_n^{1/2} \sum_{t=k_n+1}^{k_n+K_n} \eta_{n,t} \right\|^2 \right] = \frac{1}{K_n} \sum_{t=k_n+1}^{k_n+K_n} \mathbb{E} \left[ \|\xi_n^{1/2} \eta_{n,t}\|^2 \right] = \text{Tr} (\mathbb{E} [\xi_n]) \leq \text{Tr} \left( \sum_{t=h_n^*+1}^{k_n} \mathbb{E} [V_{n,t}] \right).$$

By the same argument as above, the majorant side of the above inequality is $O(m_n r_n)$, which implies the second assertion in (A.8).

Q.E.D.

The next lemma establishes the strong approximation for $S_n^*$.

**Lemma A2.** Let $\bar{\lambda}$ denote the upper bound of the eigenvalues of $\Sigma_n$. Suppose that $K_n \geq 36 m_n^3 \lambda / B_n^2$. Then, there exists a sequence $\tilde{S}_n$ of $m_n$-dimensional centered Gaussian random vectors with covariance matrix $\Sigma_n$ such that

$$\|S_n^* - \tilde{S}_n\| = O_p((B_n m_n)^{1/3}).$$

4
**Proof of Lemma A2.** *Step 1.* We introduce some notations and outline the proof in this step. For any positive constant $C > 1$, we denote $\delta_{C,n} \equiv C(B_n m_n)^{1/3}$. We also set $\sigma_n^2 \equiv B_n^{2/3} m_n^{-1/3}$ and note that
\[
\frac{\delta_{C,n}^2}{m_n \sigma_n^2} = C^2 \quad \text{and} \quad \frac{B_n}{\sigma_n^2 \delta_{C,n}} = C^{-1}.
\] (A.9)

Below, we denote
\[
\psi_{C,n} \equiv \left( \frac{C^2}{\exp(C^2 - 1)} \right)^{m_n/2}.
\] (A.10)

Note that as $C \to \infty$,
\[
\psi_{C,n} \to 0 \text{ uniformly in } n.
\] (A.10)

In *step 2*, below, we show that the following inequality holds for any Borel subset $A \subseteq \mathbb{R}^{m_n}$:
\[
P(S_n^* \in A) \leq F_n \left(A^{3\delta_{C,n}}\right) + \frac{1}{1 - \psi_{C,n}} \left( \psi_{C,n} + 4B_n \sigma_n^2 \delta_{C,n} \right),
\] (A.11)

where $F_n$ denotes the distribution of an $\mathcal{N}(0, \Sigma_n)$ random variable and
\[
A^{3\delta_{C,n}} \equiv \left\{ x \in \mathbb{R}^{m_n} : \inf_{y \in A} \| x - y \| \leq 3\delta_{C,n} \right\}.
\]

Consequently, by Strassen’s Theorem (see, e.g., Theorem 10.8 in Pollard (2001)), we can construct a variable $\widetilde{S}_n \sim \mathcal{N}(0, \Sigma_n)$ such that
\[
P \left( \| S_n^* - \widetilde{S}_n \| > 3\delta_{C,n} \right) \leq \frac{1}{1 - \psi_{C,n}} \left( \psi_{C,n} + 4B_n \sigma_n^2 \delta_{C,n} \right)
\]
\[
= \frac{1}{1 - \psi_{C,n}} \left( \psi_{C,n} + 4C^{-1} \right).
\]

By (A.10), for any $\varepsilon > 0$ there exists a constant $C_{\varepsilon} > 1$ such that for any $C > C_{\varepsilon}$ and for any $n$, the majorant side of the above inequality is bounded by $\varepsilon$, yielding
\[
P \left( \| S_n^* - \widetilde{S}_n \| > 3C(B_n m_n)^{1/3} \right) < \varepsilon.
\]

This proves the assertion of the lemma.

*Step 2.* It remains to show (A.11). For notational simplicity, we write $\delta_n$ and $\psi_n$ in place of $\delta_{C,n}$ and $\psi_{C,n}$, respectively. With $\sigma_n$ described in step 1, we consider the following functions on $\mathbb{R}^{m_n}$:
\[
g_n(x) \equiv \max \left\{ 0, 1 - d(x, A^{\delta_n})/\delta_n \right\}, \quad f_n(x) \equiv \mathbb{E} \left[ g_n(x + \sigma_n \mathcal{N}^*) \right],
\]
where $\mathcal{N}^*$ is an $m_n$-dimensional standard normal random vector and $d(x, A^{\delta_n})$ denotes the distance between $x$ and the set $A^{\delta_n}$. By Lemma 10.18 in Pollard (2001), $f_n(\cdot)$ is three-time continuously differentiable such that for all $(x, y)$,
\[
\left| f_n(x + y) - f_n(x) - \partial f_n(x)^\top y - \frac{1}{2} y^\top \partial^2 f_n(x) y \right| \leq \frac{\| y \|^3}{\sigma_n^2 \delta_n},
\] (A.12)
and
\[(1 - \psi_n)1 \{ x \in A \} \leq f_n(x) \leq \psi_n + (1 - \psi_n)1 \{ x \in A^{3n} \}. \quad \text{(A.13)}\]

Let \( \zeta_{n,t}, \ 1 \leq t \leq k_n + K_n, \) be independent \( m_n \)-dimensional standard normal vectors and \( \tilde{\zeta}_{n,t} = (V_{n,t}^*)^{1/2} \zeta_{n,t}; \) recall the definition of \( V_{n,t}^* \) from (A.1). We set
\[D_{n,t} \equiv \sum_{1 \leq s < t} Z_{n,s} + \sum_{t < s \leq k_n + K_n} \tilde{\zeta}_{n,s}.\]

It is easy to see that
\[\int f_n(x) F_n(dx) = \mathbb{E} \left[ f_n(D_{n,1} + \tilde{\zeta}_{n,1}) \right], \quad \mathbb{E} \left[ f_n(S_n^*) \right] = \mathbb{E} \left[ f_n(D_{n,k_n + K_n} + Z_{n,k_n + K_n}) \right];\]
and
\[D_{n,t} + Z_{n,t} = D_{n,t+1} + \tilde{\zeta}_{n,t+1}, \quad 1 \leq t \leq k_n + K_n - 1.\]

Hence,
\[\mathbb{E} \left[ f_n(S_n^*) \right] - \int f_n(x) F_n(dx) = \sum_{t=1}^{k_n + K_n} \left( \mathbb{E} \left[ f_n(D_{n,t} + Z_{n,t}) \right] - \mathbb{E} \left[ f_n(D_{n,t} + \tilde{\zeta}_{n,t}) \right] \right). \quad \text{(A.14)}\]

By (A.12), we have
\[\left| \mathbb{E} \left[ f_n(D_{n,t} + Z_{n,t}) \right] - \mathbb{E} \left[ f_n(D_{n,t}) \right] - \mathbb{E} [\partial f_n(D_{n,t})^\top Z_{n,t}] \right| \leq \frac{1}{\sigma_n^2 \delta_n} \mathbb{E} [\|Z_{n,t}\|^3], \quad \text{(A.15)}\]
and
\[\left| \mathbb{E} [\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] - \mathbb{E} [\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] \right| \leq \frac{1}{\sigma_n^2 \delta_n} \mathbb{E} [\|\tilde{\zeta}_{n,t}\|^3]. \quad \text{(A.16)}\]

Since \( \tilde{\zeta}_{n,t} = (V_{n,t}^*)^{1/2} \zeta_{n,t} \) and \( \zeta_{n,t} \) is a standard normal random vector independent of \( D_{n,t} \) and \( V_{n,t}^* \), we have
\[\mathbb{E} [\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] = 0 \quad \text{and} \quad \mathbb{E} [\text{Tr}(\partial^2 f_n(D_{n,t})\tilde{\zeta}_{n,t}^\top \tilde{\zeta}_{n,t}^\top)] = \mathbb{E} [\text{Tr}(\partial^2 f_n(D_{n,t})V_{n,t}^*)]. \quad \text{(A.17)}\]

Let \( \tilde{D}_{n,t} \equiv \sum_{1 \leq s < t} Z_{n,s} + (\Sigma_n - \sum_{s=1}^t V_{n,t}^*) \zeta_{n,t}. \) We note that since \( \Sigma_n \) is nonrandom, \( \Sigma_n - \sum_{s=1}^t V_{n,t}^* \) is \( \mathcal{H}_{n,t-1} \)-measurable. We then observe that
\[\mathbb{E} [\partial f_n(D_{n,t})^\top Z_{n,t}] = \mathbb{E} [\partial f_n(\tilde{D}_{n,t})^\top Z_{n,t}] \quad = \mathbb{E} [\partial f_n(\tilde{D}_{n,t})^\top \mathbb{E} [Z_{n,t} | \mathcal{H}_{n,t-1}, \zeta_{n,t}]] \quad = \mathbb{E} [\partial f_n(\tilde{D}_{n,t})^\top \mathbb{E} [Z_{n,t} | \mathcal{H}_{n,t-1}]] = 0, \quad \text{(A.18)}\]
where the first equality holds because the conditional distribution of \( \tilde{D}_{n,t} \) given \( \mathcal{H}_{n,k_n + K_n} \) is the same as that of \( D_{n,t} \); the second equality holds because \( \sum_{1 \leq s < t} Z_{n,s} \) and \( \Sigma_n - \sum_{s=1}^t V_{n,t}^* \) are
Similarly, the last equality holds because \( Z_{n,t} | \mathcal{H}_{n,t} \) is a martingale difference array by construction. Combining the results in (A.17), (A.18) and (A.19), we have

\[
\mathbb{E}[\text{Tr}(\partial^2 f_n(D_{n,t}) Z_{n,t} Z_{n,t}^\top)] = \mathbb{E}[\text{Tr}(\partial^2 f_n(\tilde{D}_{n,t}) Z_{n,t} Z_{n,t}^\top)]
\]

\[
= \mathbb{E}[\text{Tr}(\partial^2 f_n(\tilde{D}_{n,t}) \mathbb{E}[Z_{n,t} Z_{n,t}^\top | \mathcal{H}_{n,t-1}, \zeta_{n,t}])]
\]

\[
= \mathbb{E}[\text{Tr}(\partial^2 f_n(\tilde{D}_{n,t}) \mathbb{E}[Z_{n,t} Z_{n,t}^\top | \mathcal{H}_{n,t-1}])]
\]

\[
= \mathbb{E}[\text{Tr}(\partial^2 f_n(\tilde{D}_{n,t}) V_{n,t}^* Z_{n,t}^*)] = \mathbb{E}[\text{Tr}(\partial^2 f_n(D_{n,t}) V_{n,t}^*)]. \quad (A.19)
\]

Combining the results in (A.17), (A.18) and (A.19), we have

\[
\mathbb{E}[\partial f_n(D_{n,t}) Z_{n,t}] = \mathbb{E}[\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] = 0
\]

\[
\mathbb{E}[\text{Tr}(\partial^2 f_n(D_{n,t}) Z_{n,t} Z_{n,t}^\top)] = \mathbb{E}[\text{Tr}(\partial^2 f_n(D_{n,t}) \tilde{\zeta}_{n,t} \tilde{\zeta}_{n,t}^\top)].
\]

Combining this with (A.14), (A.15) and (A.16), we deduce

\[
\left| \mathbb{E}[f(S_n^*)] - \int f_n(x) F_n(dx) \right|
\]

\[
\leq \frac{1}{\sigma_n^2 \delta_n} \sum_{t=1}^{k_n+K_n} \left( \mathbb{E}[\|Z_{n,t}\|^3] + \mathbb{E}[\|\tilde{\zeta}_{n,t}\|^3] \right)
\]

\[
= \frac{1}{\sigma_n^2 \delta_n} \sum_{t=1}^{k_n} \left( \mathbb{E}[\|X_{n,t} 1_{t \leq \tau_n}\|^3] + \mathbb{E}[\|V_{n,t}^1 \zeta_{n,t} 1_{t \leq \tau_n}\|^3] \right) + \frac{1}{\sigma_n^2 \delta_n} \sum_{t=k_n+1}^{k_n+K_n} \mathbb{E} \left[ \|K_n^{-1/2} \xi_n^1 \eta_{n,t}\|^3 \right]
\]

\[
\leq \frac{1}{\sigma_n^2 \delta_n} \sum_{t=1}^{k_n} \left( \mathbb{E}[\|X_{n,t}\|^3] + \mathbb{E}[\|V_{n,t}^1 \zeta_{n,t}\|^3] \right) + \frac{1}{\sigma_n^2 \delta_n} \sum_{t=k_n+1}^{k_n+K_n} \mathbb{E} \left[ \|K_n^{-1/2} \xi_n^1 \eta_{n,t}\|^3 \right]
\]

\[
\leq \frac{3B_n}{\sigma_n^2 \delta_n} + \frac{2}{\sigma_n^2 \delta_n} \mathbb{E} \left[ \|\xi_n^1 \mathcal{N}\|^3 \right],
\]

where \( \mathcal{N} \) is a generic \( m_n \)-dimensional standard normal random vector and the last inequality follows from (denoting by \( \Phi \) the distribution function of \( \mathcal{N} \))

\[
\mathbb{E} \left[ \|V_{n,t}^{1/2} \zeta_{n,t}\|^3 \right] = \mathbb{E} \left[ \left( \zeta_{n,t} \mathbb{E}[X_{n,t} X_{n,t}^\top | \mathcal{F}_{n,t-1}] \zeta_{n,t} \right)^{3/2} \right]
\]

\[
= \mathbb{E} \left[ \left( u^\top \mathbb{E}[X_{n,t} X_{n,t}^\top | \mathcal{F}_{n,t-1}] u \right)^{3/2} \Phi (du) \right]
\]

\[
= \mathbb{E} \left[ \left( \mathbb{E}[(u^\top X_{n,t})^2 | \mathcal{F}_{n,t-1}] \right)^{3/2} \Phi (du) \right]
\]

\[
\leq \mathbb{E} \left[ \mathbb{E}[|u^\top X_{n,t}|^3 | \mathcal{F}_{n,t-1}] \Phi (du) \right]
\]

\[
= \mathbb{E} \left[ |\zeta_{n,t}^\top X_{n,t}|^3 \right] = \sqrt{3/\pi} \mathbb{E} \left[ |X_{n,t}|^3 \right] .
\]
Note that $\Sigma_n - \xi_n$ is positive semi-definite. Hence,

\[
\mathbb{E} \left[ \left\| \frac{1}{n} \sum_{i=1}^{n} X_i \right\|^3 \right] = \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)^{3/2} \right] \leq \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i \right)^{3/2} \right] \\
\leq \lambda_{\text{max}}(\Sigma_n)^{3/2} \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i \right)^{3/2} \right] \\
\leq \bar{\lambda}^{3/2} \left( \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i \right)^2 \right] \right)^{3/4} \leq 3\lambda^{3/2} m_n^{3/2}.
\]

Hence, under the condition $K_n \geq 36\lambda^3 m_n^3 / B_n^2$,

\[
\left| \mathbb{E} [f(S_n^\star)] - \int f_n(x)F_n(dx) \right| \leq \frac{3B_n}{\sigma_n^2 \delta_n} + \frac{6\lambda^{3/2} m_n^{3/2}}{\sigma_n^2 \delta_n K_{n1/2}^{1/2}} \leq \frac{4B_n}{\sigma_n^2 \delta_n}.
\]

From (A.13) and (A.20),

\[
\mathbb{P} (S_n^\star \in A) \leq \frac{1}{1 - \psi_n} \mathbb{E} [f_n(S_n^\star)] \\
\leq \frac{1}{1 - \psi_n} \left( \int f_n(x)F_n(dx) + \frac{4B_n}{\sigma_n^2 \delta_n} \right) \\
\leq \frac{1}{1 - \psi_n} \left( \psi_n + (1 - \psi_n)F_n(A^{3\delta_n}) + \frac{4B_n}{\sigma_n^2 \delta_n} \right) \\
= F_n(A^{3\delta_n}) + \frac{1}{1 - \psi_n} \left( \psi_n + \frac{4B_n}{\sigma_n^2 \delta_n} \right),
\]

which proves (A.11) as wanted. \(Q.E.D.\)

**Proof of Theorem 1.** Let $K_n$ satisfy the condition in Lemma A2 and then define $S_n^\star$ as in (A.2). The assertion of Theorem 1 then readily follows from Lemma A1 and Lemma A2. \(Q.E.D.\)

**S.A.2 Proof of Theorem 2**

**Proof of Theorem 2.** Step 1. The proof for part (a) of the theorem is divided into 3 steps. Below, for a generic real sequence $a_n$, let $O_{pu}(a_n)$ denote a random sequence that is $O_p(a_n)$ uniformly in $x \in \mathcal{X}$. In this step, we show

\[
\frac{n^{1/2}P(x)^\top (\hat{b}_n - b_n^\star)}{\sigma_n(x)} = \frac{n^{-1/2}P(x)^\top Q_n^{-1}P_n^\top U_n}{\sigma_n(x)} + O_{pu}(\delta_{1,n} + m_n^{1/2}\delta_{3,n}),
\]

where $P_n \equiv [P(X_1), \ldots, P(X_n)]^\top$ and $U_n = (u_1, \ldots, u_n)^\top$.

By Assumption 2(ii),

\[
\sup_{x \in \mathcal{X}} \frac{||P(x)||}{\sigma_n(x)} \leq (\lambda_{\text{min}}(A_n))^{-1/2}\lambda_{\text{max}}(Q_n) \leq K.
\]

(A.22)
By Assumptions 2(ii), (iv) and (v), we have, with probability approaching one,

$$\lambda_{\max}(\hat{Q}_n) + \lambda_{\max}(\hat{A}_n) + \lambda_{\min}^{-1}(\hat{Q}_n) + \lambda_{\min}^{-1}(\hat{A}_n) \leq K.$$  \hspace{1cm} (A.23)

Let $h^*_n (\cdot) \equiv P(\cdot)^{\top} b^*_n$, $H_n = (h(X_1), \ldots, h(X_n))^\top$ and $H^*_n = (h^*_n(X_1), \ldots, h^*_n(X_n))^\top$. By the definition of $\hat{b}_n$, we can decompose

$$\hat{b}_n - b^*_n = (P_n^\top P_n)^{-1} \left( P_n^\top U_n \right) + (P_n^\top P_n)^{-1} P_n^\top (H_n - H^*_n).$$  \hspace{1cm} (A.24)

By Assumption 2(ii),

$$\mathbb{E} \left[ \left\| n^{-1} P_n^\top U_n \right\|^2 \right] = n^{-1} \operatorname{Tr}(A_n) \leq Km_n n^{-1},$$

which together with Markov’s inequality implies that

$$\left\| n^{-1/2} P_n^\top U_n \right\| = O_p(m_n^{1/2}).$$  \hspace{1cm} (A.26)

We observe

$$\sup_{x \in X} \frac{1}{\sigma_n(x)} \left| n^{1/2} P(x)^\top (P_n^\top P_n)^{-1} \left( P_n^\top U_n \right) - n^{-1/2} P(x)^\top Q_n^{-1} \left( P_n^\top U_n \right) \right|$$

$$= \sup_{x \in X} \frac{1}{\sigma_n(x)} \left| P(x)^\top Q_n^{-1} (\hat{Q}_n - Q_n) Q_n^{-1} \left( n^{-1/2} P_n^\top U_n \right) \right|$$

$$\leq (\lambda_{\min}(\hat{Q}_n) \lambda_{\min}(Q_n))^{-1} \left\| \hat{Q}_n - Q_n \right\| \left\| n^{-1/2} P_n^\top U_n \right\| \sup_{x \in X} \frac{\left\| P(x) \right\|}{\sigma_n(x)}$$

$$= O_p(m_n^{1/2} \delta_{3,n}),$$

where the inequality follows from the Cauchy–Schwarz inequality and the last line is derived from Assumption 2(iv), (A.22), (A.23) and (A.26).

By Assumption 2(ii), (A.22) and (A.23),

$$\sup_{x \in X} \frac{1}{\sigma_n(x)} \left| n^{1/2} P(x)^\top (P_n^\top P_n)^{-1} P_n^\top (H_n - H^*_n) \right|$$

$$\leq \sup_{x \in X} \frac{\left\| P(x) \right\|}{\sigma_n(x)} \left( (H_n - H^*_n)^\top P_n (P_n^\top P_n)^{-1/2} \hat{Q}_n^{-1/2} P_n (P_n^\top P_n)^{-1/2} (H_n - H^*_n) \right)^{1/2}$$

$$\leq \left( (H_n - H^*_n)^\top P_n (P_n^\top P_n)^{-1} P_n^\top (H_n - H^*_n) \right)^{1/2} \sup_{x \in X} \frac{\left\| P(x) \right\|}{\sigma_n(x)}$$

$$\leq \frac{(\lambda_{\min}(\hat{Q}_n))^{1/2}}{(\lambda_{\min}(Q_n))^{1/2}} \sup_{x \in X} \frac{\left\| P(x) \right\|}{\sigma_n(x)} = O_p(\delta_{1,n}).$$  \hspace{1cm} (A.28)

The claim in (A.21) follows by combining the results in (A.24), (A.27) and (A.28).

**Step 2.** In this step, we show that

$$\sup_{x \in X} \frac{1}{\delta_n(x)} \left| n^{1/2} P(x)^\top (\hat{b}_n - b^*_n) - n^{1/2} P(x)^\top (\hat{b}_n - b^*_n) \right| = O_p(m_n^{1/2}(\delta_{3,n} + \delta_{4,n})).$$  \hspace{1cm} (A.29)
Combining (A.32) and (A.34), we deduce
\[ \left\| \hat{\Sigma}_n - \Sigma_n \right\|_S \leq \left\| (\hat{Q}_n^{-1} - Q_n^{-1}) \hat{A}_n \hat{Q}_n^{-1} \right\|_S 
+ \left\| Q_n^{-1} (\hat{A}_n - A_n) \hat{Q}_n^{-1} \right\|_S + \left\| Q_n^{-1} A_n (\hat{Q}_n^{-1} - Q_n^{-1}) \right\|_S. \]

By the Cauchy–Schwarz inequality, Assumption 2(ii, iv) and (A.23),
\[ \left\| (\hat{Q}_n^{-1} - Q_n^{-1}) \hat{A}_n \hat{Q}_n^{-1} \right\|_S \leq \frac{\lambda_{\text{max}}(\hat{A}_n) \left\| \hat{Q}_n - Q_n \right\|_S}{\lambda_{\text{min}}(Q_n)(\lambda_{\text{min}}(\hat{Q}_n))^2} = O_p(\delta_{3,n}). \]

Similarly, \( \left\| Q_n^{-1} (\hat{A}_n - A_n) \hat{Q}_n^{-1} \right\|_S = O_p(\delta_{4,n}) \) and \( \left\| Q_n^{-1} A_n (\hat{Q}_n^{-1} - Q_n^{-1}) \right\|_S = O_p(\delta_{3,n}) \). Combining these estimates, we get
\[ \left\| \hat{\Sigma}_n - \Sigma_n \right\|_S = O_p(\delta_{3,n} + \delta_{4,n}). \tag{A.30} \]

By Assumption 2(ii), this estimate further implies that, with probability approaching one,
\[ \lambda_{\text{min}}^{-1}(\hat{\Sigma}_n) \leq K, \quad \lambda_{\text{max}}(\hat{\Sigma}_n) \leq K. \tag{A.31} \]

We then observe
\[ \sup_{x \in X} \left\| \frac{\sigma_n(x) - \hat{\sigma}_n(x)}{\hat{\sigma}_n(x)} \right\| = \sup_{x \in X} \frac{\left| \sigma_n^2(x) - \hat{\sigma}_n^2(x) \right|}{\sigma_n(x) \left( \sigma_n(x) + \hat{\sigma}_n(x) \right)} \]
\[ = \sup_{x \in X} \frac{\left| P(x) \top (\hat{\Sigma}_n - \Sigma_n) P(x) \right|}{\sigma_n(x) \left( \sigma_n(x) + \hat{\sigma}_n(x) \right)} \]
\[ \leq \frac{\left\| \hat{\Sigma}_n - \Sigma_n \right\|_S}{(\lambda_{\text{min}}(\hat{\Sigma}_n))(\lambda_{\text{min}}(\Sigma_n))^{1/2} + \lambda_{\text{min}}(\Sigma_n)} = O_p(\delta_{3,n} + \delta_{4,n}) \tag{A.32} \]

where the last line follows from Assumption 2(ii), (A.30) and (A.31).

By the Cauchy–Schwarz inequality
\[ \sup_{x \in X} \left\| \frac{n^{1/2} P(x) \top U_n}{\sigma_n(x)} \right\| \leq \frac{n^{1/2} P \top U_n}{\lambda_{\text{min}}(Q_n)} \sup_{x \in X} \left\| P(x) \right\| = O_p(m_n^{1/2}) \tag{A.33} \]

where the equality is due to Assumption 2(ii, A.22) and (A.26). By (A.21) and (A.33),
\[ \sup_{x \in X} \left\| \frac{n^{1/2} P(x) \top (\hat{b}_n - b_n^*)}{\sigma_n(x)} \right\| = O_p(m_n^{1/2}). \tag{A.34} \]

Combining (A.32) and (A.34), we deduce
\[ \sup_{x \in X} \left\| \frac{n^{1/2} P(x) \top (\hat{b}_n - b_n^*)}{\hat{\sigma}_n(x)} - \frac{n^{1/2} P(x) \top (\hat{b}_n - b_n^*)}{\sigma_n(x)} \right\| \]
\[ \leq \sup_{x \in X} \left\| \frac{\sigma_n(x) - \hat{\sigma}_n(x)}{\hat{\sigma}_n(x)} \sup_{x \in X} \left\| \frac{n^{1/2} P(x) \top (\hat{b}_n - b_n^*)}{\sigma_n(x)} \right\| = O_p(m_n^{1/2}(\delta_{3,n} + \delta_{4,n})), \right\|
which finishes the proof of (A.29).

**Step 3.** In this step, we show the assertion in part (a) of the theorem. It suffices to show that,

\[
\frac{n^{1/2} (\hat{h}_n(x) - h(x))}{\sigma_n(x)} = \frac{P(x)\top \tilde{S}_n}{\sigma_n(x)} + O_p(\delta_n), \tag{A.35}
\]

where \( \tilde{S}_n \equiv Q^{-1}_n \tilde{N}_n \) is \( \mathcal{N}(0, \Sigma_n) \) distributed.

By (A.21) and (A.29),

\[
\frac{n^{1/2} P(x)\top (\hat{b}_n - b_n^*)}{\sigma_n(x)} = \frac{n^{-1/2} P(x)\top Q^{-1}_n P_n U_n}{\sigma_n(x)} + O_p(\delta_{1,n} + \delta_{2,n} + m_n^{1/2}(\delta_{3,n} + \delta_{4,n})). \tag{A.36}
\]

By Assumption 2(ii) and (A.22),

\[
\sup_{x \in \mathcal{X}} \left\| \frac{P(x)\top Q_n^{-1}}{\sigma_n(x)} \right\| \leq K. \tag{A.37}
\]

Hence, by Assumption 2(iii), (A.36) and (A.37),

\[
\frac{n^{1/2} P(x)\top (\hat{b}_n - b_n^*)}{\sigma_n(x)} = \frac{P(x)\top Q_n^{-1} \tilde{N}_n}{\sigma_n(x)} + O_p(\delta_{1,n} + \delta_{2,n} + m_n^{1/2}(\delta_{3,n} + \delta_{4,n})). \tag{A.38}
\]

By Assumption 2(i) and (A.22),

\[
\sup_{x \in \mathcal{X}} \left\| \frac{n^{1/2} (h(x) - P(x)\top b_n^*)}{\sigma_n(x)} \right\| = \sup_{x \in \mathcal{X}} \left\| \frac{P(x)\top \left( n^{1/2} (h(x) - P(x)\top b_n^*) \right) \right\| = O(\delta_{1,n}), \tag{A.39}
\]

which combined with (A.32) yields

\[
\sup_{x \in \mathcal{X}} \left\| \frac{n^{1/2} (h(x) - P(x)\top b_n^*)}{\sigma_n(x)} \right\| \leq \sup_{x \in \mathcal{X}} \left\| \frac{\hat{\sigma}_n(x) - \sigma_n(x)}{\sigma_n(x)} \right\| \left\| \frac{n^{1/2} (h(x) - P(x)\top b_n^*)}{\sigma_n(x)} \right\|
\]

\[+
\sup_{x \in \mathcal{X}} \left\| \frac{n^{1/2} (h(x) - P(x)\top b_n^*)}{\sigma_n(x)} \right\| = O_p(\delta_{1,n}). \tag{A.40}
\]

From (A.38) and (A.40), the assertion (A.35) readily follows.

**Step 4.** Given the result in part (a), the assertion of part (b) can be shown by using similar arguments in the proof of Theorem 5.6 in Belloni, Chernozhukov, Chetverikov, and Kato (2015). We omit the proof for brevity.

**Q.E.D.**

### S.A.3 Proof of Theorem 3

We first introduce some notations and a preliminary estimate; see Lemma A3, below. Recall that the feasible estimator \( \hat{b}_n \) is given by

\[
\hat{b}_n \equiv \left[ \sum_{t=1}^{n} P(X_t) P(X_t)^\top \right]^{-1} \left( \sum_{t=1}^{n} P(X_t) g(Y_t^*; \hat{\gamma}_n) \right).
\]
We denote the corresponding infeasible estimator by
\[
\hat{b}_n \equiv \left( \sum_{t=1}^{n} P(X_t) P(X_t)^	op \right)^{-1} \left( \sum_{t=1}^{n} P(X_t) g(Y_t^*; \gamma_0) \right).
\]

**Lemma A3.** Let \( \delta_{6,n} \equiv \sup_{x \in \mathcal{X}} \| P(x) \|^{-1} \). Under Assumption 4,
\[
\sup_{x \in \mathcal{X}} n^{1/2} P(x)^	op \frac{\hat{b}_n - \hat{b}_n^\dagger}{\sigma_n(x)} = O_p(\delta_{3,n} + \delta_{5,n} + \delta_{6,n}).
\]

**Proof of Lemma A3.** Step 1. In this step, we show that
\[
\sup_{x \in \mathcal{X}} n^{1/2} P(x)^	op \frac{\hat{b}_n - \hat{b}_n^\dagger}{\sigma_n(x)} = \sup_{x \in \mathcal{X}} n^{1/2} P(x)^	op Q_n^{-1} G_n (\hat{\gamma}_n - \gamma_0) + O_p(\delta_{3,n} + \delta_{5,n}) \quad (A.41)
\]

By definition,
\[
\hat{b}_n - \hat{b}_n^\dagger = \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t)(g(Y_t^*; \hat{\gamma}_n) - g(Y_t^*; \gamma_0)). \quad (A.42)
\]

Applying the second-order Taylor expansion, we further deduce
\[
\hat{b}_n - \hat{b}_n^\dagger = \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t) g_{\gamma}(Y_t^*; \gamma_0)^	op (\hat{\gamma}_n - \gamma_0)
\]
\[
+ \frac{1}{2} \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t)(\hat{\gamma}_n - \gamma_0)^	op g_{\gamma\gamma}(Y_t^*; \hat{\gamma}_n)(\hat{\gamma}_n - \gamma_0), \quad (A.43)
\]

where \( \hat{\gamma}_n \) is a mean value between \( \hat{\gamma}_n \) and \( \gamma_0 \) that may vary across rows. By Assumption 4(iv,vi),
\[
n^{-1} \sum_{t=1}^{n} \left( \hat{\gamma}_n - \gamma_0 \right)^	op g_{\gamma\gamma}(Y_t^*; \hat{\gamma}_n)(\hat{\gamma}_n - \gamma_0) \right|^{2} = O_p(n^{-2}). \quad (A.44)
\]

Since \( P_n \left( P_n^\top P_n \right)^{-1} P_n^\top \) is a projection matrix,
\[
\left\| \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t)(\hat{\gamma}_n - \gamma_0)^	op g_{\gamma\gamma}(Y_t^*; \hat{\gamma}_n)(\hat{\gamma}_n - \gamma_0) \right\|^{2} \leq \lambda_{\min}(Q_n) n^{-1} \sum_{t=1}^{n} \left( \hat{\gamma}_n - \gamma_0 \right)^	op g_{\gamma\gamma}(Y_t^*; \hat{\gamma}_n)(\hat{\gamma}_n - \gamma_0) \right|^{2} = O_p(n^{-2}), \quad (A.45)
\]

where the rate statement follows from (A.44) and Assumption 3. Collecting the results in (A.43) and (A.45), we get
\[
\hat{b}_n - \hat{b}_n^\dagger = \left( P_n^\top P_n \right)^{-1} \sum_{t=1}^{n} P(X_t) g_{\gamma}(Y_t^*; \gamma_0)^	op (\hat{\gamma}_n - \gamma_0) + O_p(n^{-1}). \quad (A.46)
\]
By Assumptions 3 and 4(ii,vi),
\[
\left(P_n^\top P_n\right)^{-1} \sum_{t=1}^n P(X_t) g_\gamma(Y_t^*, \gamma_0) \top (\hat{\gamma}_n - \gamma_0) = \hat{Q}_n^{-1} G_n (\hat{\gamma}_n - \gamma_0) + O_p(\delta_{5,n} n^{-1/2}). \tag{A.47}
\]
For 1 \leq j \leq d, let \( g_{\gamma,j}(Y_t^*, \gamma_0) \) denote the \( j \)th component of \( g_\gamma(Y_t^*, \gamma_0) \) and let \( G_{n,j} \) denote \( j \)th column of \( G_n \). We note that
\[
G_{n,j}^\top Q_n^{-1} G_{n,j} \leq n^{-1} \sum_{t=1}^n \mathbb{E} \left[ g_{\gamma,j}(Y_t^*, \gamma_0)^2 \right], \tag{A.48}
\]
because the left-hand side is the squared \( L_2 \)-norm of the projection of \( g_{\gamma,j}(Y_t^*, \gamma_0) \) onto the column space of \( P(X_t) \) under the product measure \( \mathbb{P} \otimes P_n \), with \( P_n \) being the empirical measure. Hence, for any \( 1 \leq j \leq d \),
\[
\| G_{n,j} \|^2 \leq \frac{\lambda_{\max}(Q_n)}{n} \sum_{t=1}^n \mathbb{E} \left[ g_{\gamma,j}(Y_t^*, \gamma_0)^2 \right] \leq \lambda_{\max}(Q_n) \sup_t \mathbb{E} \left[ \| g_\gamma(Y_t^*, \gamma_0) \|^2 \right]. \tag{A.49}
\]
By Assumptions 2(ii) and 4(iv), we further deduce
\[
\| G_n \|^2 \leq K. \tag{A.50}
\]
By Assumption 3, Assumption 4(iii) and (A.50),
\[
\left\| (\hat{Q}_n^{-1} - Q_n^{-1}) G_n (\hat{\gamma}_n - \gamma_0) \right\|^2 = \left\| \hat{Q}_n^{-1} (Q_n - Q_n) Q_n^{-1} G_n (\hat{\gamma}_n - \gamma_0) \right\|^2 \leq \frac{\| \hat{\gamma}_n - \gamma_0 \|^2}{(\lambda_{\min}(Q_n) \lambda_{\min}(Q_n))^2} \left\| \hat{Q}_n - Q_n \right\|^2 \| G_n \|^2 = O_p(n^{-1} \delta_{3,n}^2),
\]
which further implies that
\[
(\hat{Q}_n^{-1} - Q_n^{-1}) G_n (\hat{\gamma}_n - \gamma_0) = O_p(n^{-1/2} \delta_{3,n}). \tag{A.51}
\]
Combining the results in (A.47) and (A.51), we get
\[
\hat{b}_n - \tilde{b}_n = Q_n^{-1} G_n (\hat{\gamma}_n - \gamma_0) + O_p(\delta_{3,n} + \delta_{5,n}) n^{-1/2}. \tag{A.52}
\]
With an appeal to the Cauchy–Schwarz inequality, we deduce (A.41) from (A.52) and (A.22).

Step 2. We now prove the assertion of Lemma A3. For \( j \in \{1, \ldots, d\} \), let \( \phi_{n,j}^* \) denote the \( j \)th column of \( \phi_n^* \); recall the definition of \( \phi_n^* \) from Assumption 4. By definition,
\[
P(x) \top Q_n^{-1} G_{n,j} - P(x) \top \phi_{n,j}^* = P(x) \top Q_n^{-1} (G_{n,j} - Q_n \phi_{n,j}^*) = n^{-1} \sum_{t=1}^n P(x) \top Q_n^{-1} (\mathbb{E} \left[ P(X_t) (H_j(X_t) - H_j^*(X_t)) \right]). \tag{A.53}
\]
where \( H_j(X_t) = \mathbb{E}[g_{\gamma,j}(Y_t^*, \gamma_0)|X_t] \) and \( H_j^*(X_t) = P(X_t)^\top \phi_n^{*j} \) (we suppress the dependence of these functions on \( n \) for notational simplicity). Using similar arguments that lead to (A.48), we can show that

\[
\left\| \mathbb{E} \left[ n^{-1} \sum_{t=1}^{n} P(X_t) (H_j(X_t) - H_j^*(X_t)) \right] \right\|^2 \leq \lambda_{\text{max}}(Q_n) \left\| \mathbb{E} \left[ n^{-1} \sum_{t=1}^{n} (H_j(X_t) - H_j^*(X_t)) \right] \right\|^2,
\]

which together with Assumption 3 and Assumption 4(iii) implies that

\[
\left\| \mathbb{E} \left[ n^{-1} \sum_{t=1}^{n} P(X_t) (H_j(X_t) - H_j^*(X_t)) \right] \right\| = O(m_n^{-\rho}). \tag{A.54}
\]

By (A.54) and the Cauchy–Schwarz inequality,

\[
\sup_{x \in \mathcal{X}} \left\| n^{-1} \sum_{t=1}^{n} P(x)^\top Q_n^{-1} \mathbb{E} \left[ P(X_t) (H_j(X_t) - H_j^*(X_t)) \right] \right\|^2 \leq \sup_{x \in \mathcal{X}} \| P(x) \|^2 (\lambda_{\text{min}}(Q_n))^{-2} \left\| \mathbb{E} \left[ n^{-1} \sum_{t=1}^{n} P(X_t) (H_j(X_t) - H_j^*(X_t)) \right] \right\|^2 \leq Km_n^{1-2\rho} \sigma_n^2 = O(1)
\]

where the last line is by (A.54), Assumption 3 and Assumption 4(v). By (A.53), we further deduce that

\[
\sup_{x \in \mathcal{X}} \left\| P(x)^\top Q_n^{-1} G_n - P(x)^\top \phi_n^* \right\| \leq K. \tag{A.55}
\]

From Assumption 4(iii,iv), it is easy to see that \( P(\cdot)^\top \phi_n^* \) is uniformly bounded. Hence, by (A.55),

\[
\sup_{x \in \mathcal{X}} \left\| P(x)^\top Q_n^{-1} G_n \right\| \leq K. \tag{A.56}
\]

Using the Cauchy–Schwarz inequality, we deduce from (A.56), Assumption 3 and Assumption 4(vi) that

\[
\sup_{x \in \mathcal{X}} \frac{n^{1/2}P(x)^\top Q_n^{-1} G_n(\hat{\gamma}_n - \gamma_0)}{\sigma_n(x)} \leq \lambda_{\text{max}}(Q_n)(\lambda_{\text{min}}(A_n))^{-1/2} \left\| n^{1/2}(\hat{\gamma}_n - \gamma_0) \right\| \sup_{x \in \mathcal{X}} \frac{\| P(x)^\top Q_n^{-1} G_n \|}{\| P(x) \|} = O_p(\delta_{6,n}).
\]

The assertion of Lemma A3 then follows from this estimate and (A.41). \( Q.E.D. \)

**Proof of Theorem 3.** By Assumption 3, we can invoke (A.32) in the proof of Theorem 2 to get

\[
\sup_{x \in \mathcal{X}} \frac{|\sigma_n(x) - \hat{\sigma}_n(x)|}{\hat{\sigma}_n(x)} = O_p(\delta_{4,n} + \delta_{4,n}), \tag{A.57}
\]
which together with Lemma A3 implies that

$$
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} P(x)^\top (\hat{b}_n - \hat{h}_n)}{\hat{\sigma}_n(x)} - \frac{n^{1/2} P(x)^\top (\hat{b}_n - \hat{h}_n)}{\sigma_n(x)} \right|
\leq \sup_{x \in \mathcal{X}} \left| \frac{\sigma_n(x) - \sigma_n(x)}{\hat{\sigma}_n(x)} \right| \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} P(x)^\top (\hat{b}_n - \hat{h}_n)}{\sigma_n(x)} \right|
\leq O_p((\delta_{3,n} + \delta_{5,n} + \delta_{6,n})(\delta_{3,n} + \delta_{4,n})).
$$

Therefore,

$$
\sup_{x \in \mathcal{X}} \frac{n^{1/2} P(x)^\top (\hat{b}_n - \hat{h}_n)}{\sigma_n(x)} = O_p(\delta_{3,n} + \delta_{5,n} + \delta_{6,n}).
$$

Let \( \hat{h}_n(x) = P(x)^\top \hat{b}_n \). Applying (A.35) with \( \hat{h}_n(x) \) replacing \( \hat{h}_n(x) \),

$$
\frac{n^{1/2} (\hat{h}_n(x) - h(x))}{\sigma_n(x)} = \frac{P(x)^\top \hat{S}_n}{\sigma_n(x)} + O_p(\delta_n).
$$

Then, by Lemma A3,

$$
\frac{n^{1/2} (\hat{h}_n(x) - h(x))}{\sigma_n(x)} = \frac{P(x)^\top \hat{S}_n}{\sigma_n(x)} + O_p(\delta_n + \delta_{5,n} + \delta_{6,n}).
$$

Under the null hypothesis (3.8) with \( h(x) = 0 \),

$$
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} \hat{h}_n(x)}{\sigma_n(x)} \right| = \sup_{x \in \mathcal{X}} \left| \frac{P(x)^\top \hat{S}_n}{\sigma_n(x)} \right| + O_p(\delta_n + \delta_{5,n} + \delta_{6,n}).
$$

Note that \( (\delta_n + \delta_{5,n} + \delta_{6,n}) (\log m_n)^{1/2} = o(1) \) under maintained assumptions. The first assertion in Theorem 3 then follows from (A.61) and the argument in the proof of Theorem 5.6 in Belloni, Chernozhukov, Chetverikov, and Kato (2015).

We now turn to the second assertion. By the triangle inequality, (A.57) and (A.60),

$$
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} \hat{h}_n(x)}{\sigma_n(x)} \right| \geq \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} h(x)}{\sigma_n(x)} \right| - \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} (\hat{h}_n(x) - h(x))}{\sigma_n(x)} \right|
= \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} h(x)}{\sigma_n(x)} \right| \left( 1 - \sup_{x \in \mathcal{X}} \left| \frac{\sigma_n(x)}{\hat{\sigma}_n(x)} - 1 \right| \right) - \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} (\hat{h}_n(x) - h(x))}{\sigma_n(x)} \right|
= \sup_{x \in \mathcal{X}} \left| \frac{n^{1/2} h(x)}{\sigma_n(x)} \right| \left( 1 + o_p(1) \right) - \sup_{x \in \mathcal{X}} \left| \frac{P(x)^\top \hat{S}_n}{\sigma_n(x)} \right| + o_p(1).
$$
By the Cauchy–Schwarz inequality and Assumption 3,
\[
\sup_{x \in \mathcal{X}} \left| P(x) \frac{\tilde{S}_n}{\sigma_n(x)} \right| \leq \frac{\lambda_{\max}(Q_n)}{(\lambda_{\min}(A_n))^{1/2}} \| \tilde{S}_n \| = O_p(m_n^{1/2}). \tag{A.63}
\]

Since \( E[g(Y_t^*, \gamma_0)|X_t = x] \neq 0 \) for some \( x \in \mathcal{X} \), there exists some constant \( c_0 > 0 \) such that \( \sup_{x \in \mathcal{X}} |h(x)| > c_0 \). Moreover, by Assumption 4(v), \( \sup_{x \in \mathcal{X}} \| P(x) \| \leq \zeta_n m_n^{1/2} \). Therefore,
\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}h(x)}{\sigma_n(x)} \right| \geq \frac{n^{1/2}\lambda_{\min}(Q_n)c_0}{(\lambda_{\max}(A_n))^{1/2}\zeta_n m_n^{1/2}}, \tag{A.64}
\]
which together with (A.62) and (A.63) implies that (recalling \( \zeta_n m_n \ll n^{1/2} \) from Assumption 4(vii)),
\[
\sup_{x \in \mathcal{X}} \left| \frac{n^{1/2}h_n(x)}{\sigma_n(x)} \right| \geq \frac{n^{1/2}\lambda_{\min}(Q_n)c_0}{(\lambda_{\max}(A_n))^{1/2}\zeta_n m_n^{1/2}} (1 + o_p(1)). \tag{A.65}
\]

Like (A.73) in Belloni, Chernozhukov, Chetverikov, and Kato (2015), we can show that the critical value \( cv_{n,\alpha} \) satisfies \( cv_{n,\alpha} = O_p((\log(m_n))^{1/2}) \), which together with Assumptions 3(ii) and 4(vii) implies that
\[
\frac{(\lambda_{\max}(A_n))^{1/2}cv_{n,\alpha}\zeta_n m_n^{1/2}}{\lambda_{\min}(Q_n)(1 + o_p(1))n^{1/2}} = o_p(1). \tag{A.66}
\]

Combining the results in (A.65) and (A.66), we deduce that
\[
\Pr \left( \hat{T}_n \leq cv_{n,\alpha} \right) \leq \Pr \left( \frac{n^{1/2}\lambda_{\min}(Q_n)c_0}{(\lambda_{\max}(A_n))^{1/2}\zeta_n m_n^{1/2}} (1 + o_p(1)) \leq cv_{n,\alpha} \right) = \Pr \left( c_0 \leq \frac{(\lambda_{\max}(A_n))^{1/2}cv_{n,\alpha}\zeta_n m_n^{1/2}}{\lambda_{\min}(Q_n)(1 + o_p(1))n^{1/2}} \right) \to 0
\]
as \( n \to \infty \). From here, the second assertion readily follows. \( Q.E.D. \)

S.A.4 Proof of Theorem 4

We first establish the martingale approximation as claimed in (4.2) and (4.3) in the main text; see Lemma A4 below. The variables \( X_{n,t}^* \) and \( \tilde{X}_{n,t} \) are defined as follows:
\[
X_{n,t}^* = \sum_{s = -\infty}^{\infty} \{ E[X_{n,t+s}|\mathcal{F}_{n,t}] - E[X_{n,t+s}|\mathcal{F}_{n,t-1}] \}, \tag{A.67}
\]
\[
\tilde{X}_{n,t} = \sum_{s = 0}^{\infty} E[X_{n,t+s}|\mathcal{F}_{n,t-1}] - \sum_{s = 0}^{\infty} \{ X_{t-s-1} - E[X_{t-s-1}|\mathcal{F}_{n,t-1}] \}. \tag{A.68}
\]

Lemma A4. The following statements hold under Assumption 5 for each \( j \in \{1, \ldots, m_n\} \)
(a) \( \sum_{s = -\infty}^{\infty} \| E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}] - E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}] \|_q \leq 4\bar{c}_n k_n^{-1/2} \sum_{s = 0}^{\infty} \psi_s; \)
(b) $\sup_{j,t,n} \|X_{n,t}^{\ast}(j)\|_q \leq 4\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s$ and $E[X_{n,t}^{\ast}|\mathcal{F}_{n,t-1}] = 0$;

c) $\sum_{s=0}^{\infty} \|E[X_{n,t+s}^{j}|\mathcal{F}_{n,t-1}]\|_q + \sum_{s=0}^{\infty} \|X_{t-s-1}^{j} - E[X_{t-s-1}^{j}|\mathcal{F}_{n,t-1}]\|_q \leq 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s$;

d) $\sup_{j,t,n} \|\tilde{X}_{n,t}^{(j)}\|_q < 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s$;

e) $X_{n,t} = X_{n,t}^{\ast} + \tilde{X}_{n,t} - \tilde{X}_{n,t+1}$ for each $t \geq 1$;

(f) $\|S_n - S_n^{\ast}\|_2 = O(\bar{c}_n m_n^{-1/2} k_n^{-1/2})$.

**Proof of Lemma A.4.** (a) We first note that

$$\sum_{s=0}^{\infty} \|E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}] - E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}]\|_q$$

$$\leq \sum_{s=0}^{\infty} \|E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}]\|_q + \sum_{s=0}^{\infty} \|E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}]\|_q$$

$$\leq \bar{c}_n k_n^{-1/2} \left( \sum_{s=0}^{\infty} \psi_s + \sum_{s=0}^{\infty} \psi_{s+1} \right)$$

$$\leq 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty.$$  \hspace{1cm} (A.69)

In addition, we have

$$\sum_{s=0}^{\infty} \|E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}] - E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}]\|_q$$

$$\leq \sum_{s=0}^{\infty} \|E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}] - E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}]\|_q$$

$$\leq \sum_{s=0}^{\infty} \|X_{n,t-s} - E[X_{n,t-s}^{(j)}|\mathcal{F}_{n,t}]\|_q + \sum_{s=0}^{\infty} \|X_{n,t-s} - E[X_{n,t-s}^{(j)}|\mathcal{F}_{n,t-1}]\|_q$$

$$\leq \bar{c}_n k_n^{-1/2} \left( \sum_{s=0}^{\infty} \psi_{s+1} + \sum_{s=0}^{\infty} \psi_{s-1} \right)$$

$$\leq 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty.$$  \hspace{1cm} (A.70)

The assertion of part (a) then follows from (A.69) and (A.70).

(b) From (A.69) and (A.70), we deduce that $\|X_{n,t}^{\ast}(j)\|_q \leq 4\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s$. It remains to verify that $E[X_{n,t}^{\ast}|\mathcal{F}_{n,t-1}] = 0$. To this end, we set

$$X_{n,t}^{\ast}(m) = \sum_{s=-m}^{m} \{E[X_{n,t+s}|\mathcal{F}_{n,t}] - E[X_{n,t+s}|\mathcal{F}_{n,t-1}]\}.$$

It is easy to see that $E[X_{n,t}^{\ast}(m)|\mathcal{F}_{n,t-1}] = 0$. We note that

$$\|X_{n,t}^{\ast}(m)\|_q \leq \sum_{s=-\infty}^{\infty} \|E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}] - E[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}]\|_q.$$
where the right-hand side of the above display is integrable by the calculations in part (a). Since \( \lim_{m \to \infty} X^{*\langle j \rangle}_{n,t} (m) = X^{*\langle j \rangle}_{n,t} \) almost surely by part (a), we deduce \( \mathbb{E} [X^{*\langle j \rangle}_{n,t} \mid F_{n,t-1}] = 0 \) by using the dominated convergence theorem.

(c) The assertion of part (c) follows from (4.1) directly. Indeed,

\[
\sum_{s=0}^{\infty} \mathbb{E} \left[ X^{\langle j \rangle}_{n,t+s} \mid F_{n,t-1} \right] \leq c_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty \quad \text{(A.71)}
\]

and

\[
\sum_{s=0}^{\infty} \| X_{t-s-1} - \mathbb{E} [X_{t-s-1} \mid F_{n,t-1}] \|_q \leq c_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty. \quad \text{(A.72)}
\]

(d) The assertion follows from part (c) and the triangle inequality.

(e) We verify the assertion of part (e) as follows:

\[
\widetilde{X}_{n,t+1} - \widetilde{X}_{n,t} + X_{n,t} = \sum_{s=0}^{\infty} \mathbb{E} \left[ X_{n,t+1+s} \mid F_{n,t} \right] - \sum_{s=0}^{\infty} \left\{ X_{t-s} - \mathbb{E} [X_{t-s} \mid F_{n,t}] \right\}
\]

\[
- \sum_{s=0}^{\infty} \mathbb{E} \left[ X_{n,t+s} \mid F_{n,t-1} \right] + \sum_{s=0}^{\infty} \left\{ X_{t-s-1} - \mathbb{E} [X_{t-s-1} \mid F_{n,t-1}] \right\} + X_{n,t}
\]

\[
= \sum_{s=1}^{\infty} \mathbb{E} \left[ X_{n,t+s} \mid F_{n,t} \right] + \sum_{s=0}^{\infty} \left\{ \mathbb{E} [X_{t-s} \mid F_{n,t}] - X_{t-s} \right\}
\]

\[
- \sum_{s=0}^{\infty} \mathbb{E} \left[ X_{n,t+s} \mid F_{n,t-1} \right] + \sum_{s=1}^{\infty} \left\{ X_{t-s} - \mathbb{E} [X_{t-s} \mid F_{n,t-1}] \right\} + X_{n,t}
\]

\[
= \sum_{s=0}^{\infty} \left\{ \mathbb{E} [X_{n,t+s} \mid F_{n,t}] - \mathbb{E} [X_{n,t+s} \mid F_{n,t-1}] \right\} + \sum_{s=1}^{\infty} \left\{ \mathbb{E} [X_{t-s} \mid F_{n,t}] - \mathbb{E} [X_{t-s} \mid F_{n,t-1}] \right\}
\]

\[
= \sum_{s=0}^{\infty} \left\{ \mathbb{E} [X_{n,t+s} \mid F_{n,t}] - \mathbb{E} [X_{n,t+s} \mid F_{n,t-1}] \right\} = X^*_{n,t}
\]

(f) The assertion follows from parts (d,e), the triangle inequality and \( \sum_k \psi_k < \infty. \) \( \text{Q.E.D.} \)

**Proof of Theorem 4.** By Theorem 1, Lemma A4(f) and the triangle inequality, there exists a sequence \( \tilde{S}^*_{n} \) of \( n \)-dimensional random vectors with distribution \( \mathcal{N}(0, \Sigma^*_{n}) \) such that

\[
\left\| S_n - \tilde{S}^*_{n} \right\| = O_p(m_n^{1/2} + (B^*_{n}m_n)^{1/3} + \bar{c}_n m_n^{1/2} n^{-1/2}), \quad \text{(A.73)}
\]

where \( \Sigma^*_{n} = \mathbb{E}[S^*_{n}S^*_{n}^\top] \). By Lyapunov’s inequality and Lemma A4(d),

\[
\left\| S_n - S^*_{n} \right\|^2 = \sum_{j=1}^{m_n} \mathbb{E} \left[ \left( \tilde{X}^\langle j \rangle_{n,t} - \tilde{X}^\langle j \rangle_{n,k_n+1} \right)^2 \right] \leq Kc_n^2 m_n k_n^{-1}. \quad \text{(A.74)}
\]

By definition, \( \Sigma_n - \Sigma^*_{n} = \mathbb{E}[S_nS_n^\top - S^*_{n}S^*_{n}^\top] \). Hence, for any \( a \in \mathbb{R}^{m_n} \),

\[
\left\| a^\top (\Sigma_n - \Sigma^*_{n}) \right\|^2 \leq K \left\| \mathbb{E} \left[ a^\top (S_n - S^*_{n})S_n^\top \right] \right\|^2 + K \left\| \mathbb{E} \left[ a^\top (S_n - S^*_{n})(S_n - S^*_{n})^\top \right] \right\|^2 + K \left\| \mathbb{E} \left[ a^\top (S_n - S^*_{n})(S_n - S^*_{n})^\top \right] \right\|^2. \quad \text{(A.75)}
\]
We now bound the terms on the majorant side of (A.75). Note that
\[
\mathbb{E} \left[ a^\top (S_n - S_n^*) S_n^\top \right] \Sigma_n^{-1} \mathbb{E} \left[ (S_n - S_n^*) a \right] \leq \mathbb{E} \left[ |a^\top (S_n - S_n^*)|^2 \right],
\]
which holds because the left-hand side is the second moment of the residual obtained by projecting \( a^\top (S_n - S_n^*) \) on the random vector \( S_n \). The first term in (A.75) can thus be bounded by
\[
\left\| \mathbb{E} \left[ a^\top (S_n - S_n^*) S_n^\top \right] \right\|^2 \leq \lambda_{\max}(\Sigma_n) \mathbb{E} \left[ |a^\top (S_n - S_n^*)|^2 \right] \leq K \|a\|^2 \|S_n - S_n^*\|^2_2 \tag{A.76}
\]
where the second inequality is by the Cauchy–Schwarz inequality and the boundedness of \( \lambda_{\max}(\Sigma_n) \).

Turning to the second term in (A.75), we use the Cauchy–Schwarz inequality to derive
\[
\left\| \mathbb{E} \left[ a^\top S_n(S_n - S_n^*)^\top \right] \right\|^2 \leq \mathbb{E} \left[ |a^\top S_n|^2 \right] \|S_n - S_n^*\|^2_2 \leq K \|a\|^2 \|S_n - S_n^*\|^2_2. \tag{A.77}
\]
For the third term in (A.75), we observe
\[
\left\| \mathbb{E} \left[ a^\top (S_n - S_n^*)(S_n - S_n^*)^\top \right] \right\|^2 \leq \|a\|^2 \left( \mathbb{E} \left[ (S_n - S_n^*)(S_n - S_n^*)^\top \right] \right)^{1/2} \leq \|a\|^2 \left( \text{Tr} \left( \mathbb{E} \left[ (S_n - S_n^*)(S_n - S_n^*)^\top \right] \right) \right)^{1/2} \leq \|a\|^2 \|S_n - S_n^*\|^2_2. \tag{A.78}
\]
Combining (A.75)–(A.78), we deduce that
\[
\sup_{\|a\|=1} a^\top (\Sigma_n - \Sigma_n^*) (\Sigma_n - \Sigma_n^*)^\top a \leq K \|S_n - S_n^*\|^2_2 + K \|S_n - S_n^*\|^4_2 = \mathcal{O}(\varepsilon_n^2 m_n k_n^{-1} + \varepsilon_n^4 m_n^2 k_n^{-2}).
\]
Hence,
\[
\|\Sigma_n - \Sigma_n^*\|_S = \mathcal{O}(\varepsilon_n m_n^{1/2} k_n^{-1/2} + \varepsilon_n^2 m_n k_n^{-1}). \tag{A.79}
\]
Let \( \bar{S}_n \equiv (\Sigma_n)^{1/2}(\Sigma_n^*)^{-1/2} \tilde{S}_n^* \), so \( \bar{S}_n \sim \mathcal{N}(0, \Sigma_n) \). By definition,
\[
\bar{S}_n - \bar{S}_n^* = (\Sigma_n)^{1/2} - (\Sigma_n^*)^{1/2} (\Sigma_n^*)^{-1/2} \tilde{S}_n^* \tag{A.80}
\]
which implies that
\[
\mathbb{E} \left[ \|\bar{S}_n - \bar{S}_n^*\|^2 \right] \leq \mathbb{E} \left[ (\Sigma_n)^{1/2} - (\Sigma_n^*)^{1/2} \right] \mathbb{E} \left[ (\Sigma_n^*)^{-1/2} \tilde{S}_n^* \right] \mathbb{E} \left[ (\Sigma_n^*)^{-1/2} \tilde{S}_n^* \right] = \mathcal{O}(\varepsilon_n^2 m_n^2 k_n^{-1} + \varepsilon_n^4 m_n^3 k_n^{-2}) \tag{A.80}
\]
where the second inequality is by Exercise 7.2.18 in Horn and Johnson (1990) (also see Lemma A.2 in Belloni, Chernozhukov, Chetverikov, and Kato (2015)) and \( \lambda_{\min}(\Sigma_n)^{-1} = \mathcal{O}(1) \), and the last line follows from \( \mathbb{E}[\bar{S}_n^\top (\Sigma_n^*)^{-1} \bar{S}_n^*] = m_n \) and (A.79). Hence,
\[
\|\bar{S}_n - \bar{S}_n^*\| = \mathcal{O}(\varepsilon_n m_n k_n^{-1/2} + \varepsilon_n^2 m_n^{3/2} k_n^{-1}). \tag{A.81}
\]
The assertion of the theorem then follows from (A.73) and (A.81).
S.A.5 Proof of Theorem 5 and Theorem 6

Lemma A5. Let $\Gamma^{(k,l)}_{X,n}(s) \equiv \mathbb{E}[X^{(k)}_{n,t}X^{(l)}_{n,t+s}]$. Under Assumption 5 and Assumption 7(i,iv),

$$\max_{1 \leq k,l \leq m_n} \sum_{s=-\infty}^{\infty} |s|^2 \| \Gamma^{(k,l)}_{X,n}(s) \| \leq K \bar{c}^2_n k^{-1}. \quad (A.82)$$

Proof of Lemma A5. For each $s \geq 0$,

$$\left\| \Gamma^{(k,l)}_{X,n}(s) \right\| = \left\| \mathbb{E}[X^{(k)}_{n,t} \mathbb{E}[X^{(l)}_{n,t+s} | \mathcal{F}_{n,t}]] \right\| \leq \left\| X^{(k)}_{n,t} \right\|_2 \left\| \mathbb{E}[X^{(l)}_{n,t+s} | \mathcal{F}_{n,t}] \right\|_2 \leq \left\| \mathbb{E}[X^{(k)}_{n,t} | \mathcal{F}_{n,t}] \right\|_q \left\| \mathbb{E}[X^{(l)}_{n,t+s} | \mathcal{F}_{n,t}] \right\|_q \leq \psi_0 \psi_s \bar{c}^2_n k^{-1}, \quad (A.83)$$

where the first equality is by repeated conditioning; the first inequality is by the Cauchy–Schwarz inequality; the second inequality follows from Lyapunov’s inequality; the last line is due to Assumption 5. Hence,

$$\sum_{s=-\infty}^{\infty} |s|^2 \| \Gamma^{(k,l)}_{X,n}(s) \| \leq 2 \sum_{s=0}^{\infty} s^2 \| \Gamma^{(k,l)}_{X,n}(s) \| \leq \left( 2\psi_0 \sum_{s=0}^{\infty} |s|^2 \psi_s \right) \bar{c}^2_n k^{-1}. \quad (A.84)$$

By Assumption 7(iv), $K = 2\psi_0 \sum_{s=0}^{\infty} |s|^2 \psi_s$ is finite. This finishes the proof. Q.E.D.

Lemma A6. Under Assumptions 5, 6 and 7, we have for any $s$ with $|s| \leq k - 1$,

$$\max_{1 \leq k,l \leq m_n} \left\| \Gamma^{(k,l)}_{X,n}(s) - \mathbb{E}[\Gamma^{(k,l)}_{X,n}(s)] \right\|_2 \leq K \bar{c}^2_n k^{-1} \quad (A.85)$$

where $K > 0$ is a finite constant that does not depend on $s$.

Proof of Lemma A6. Step 1. In this step, we derive some preliminary estimates. Let $\eta_{t,s} = X^{(l)}_{n,t+s} - \mathbb{E}[X^{(l)}_{n,t}X^{(k)}_{n,t+s}]$. We shall show that

$$|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]| \leq \begin{cases} \psi_{h-s} \bar{c}^4_n k^{-2} & \text{when } h \geq s \geq 0, \\ K \left( \psi_{h-s} + \psi_s^2 \right) \bar{c}^4_n k^{-2} & \text{when } s > h \geq 0. \end{cases} \quad (A.86)$$

We start with the case $h \geq s \geq 0$. By Assumption 7(iii), we have for all $s \geq 0$,

$$\sup_t \max_{1 \leq l,k \leq m_n} \mathbb{E}[\eta_{t,s}^2] \leq \sup_t \max_{1 \leq l,k \leq m_n} \mathbb{E}\left[\left| X^{(l)}_{n,t}X^{(k)}_{n,t+s} \right|^2 \right] \leq \bar{c}^4_n k^{-2}. \quad (A.87)$$

By the Cauchy–Schwarz inequality, Assumptions 7(ii) and (A.87), we deduce

$$|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]| = |\mathbb{E}[\eta_{t,s} \mathbb{E}[\eta_{t+h,s} | \mathcal{F}_{n,t+s}]]| \leq \|\eta_{t,s}\|_2 \|\mathbb{E}[\eta_{t+h,s} | \mathcal{F}_{n,t+s}]\|_2 \leq \psi_{h-s} \bar{c}^4_n k^{-2}, \quad (A.88)$$
as asserted.

Turning to the case with $s > h \geq 0$, we first note that by the definition of $\eta_{t,s}$ and the triangle inequality,

$$|E[\eta_{t,s}\eta_{t+h,s}]| \leq |E[\eta_{t,h}\eta_{t+h,s}]| + \left|\Gamma_{n,X}^{(k,l)}(h)\right|^2 + \left|\Gamma_{n,X}^{(k,l)}(s)\right|^2. \quad (A.89)$$

By swapping $s$ and $h$ in (A.88), we obtain $|E[\eta_{t,h}\eta_{t+h,s}]| \leq \psi_{s-h}c_n^4k_n^{-2}$. By (A.83),

$$\left|\Gamma_{n,X}^{(k,l)}(h)\right|^2 + \left|\Gamma_{n,X}^{(k,l)}(s)\right|^2 \leq K(\psi_n^2 + \psi_s^2)c_n^4k_n^{-2}.$$

The second claim in (A.86) then readily follows from these estimates.

**Step 2.** We now prove (A.85). Since $\Gamma_{X,n}^{(k,l)}(s) = \Gamma_{X,n}^{(k,l)}(-s)$, it suffices to consider $s \geq 0$. With $\eta_{t,s}$ defined in step 1, we can rewrite $\Gamma_{X,n}^{(k,l)}(s) - E[\Gamma_{X,n}^{(k,l)}(s)] = \sum_{t=1}^{k_n-s} \eta_{t,s}$. Hence,

$$\left|\frac{\Gamma_{X,n}^{(k,l)}(s) - E[\Gamma_{X,n}^{(k,l)}(s)]}{2}\right|^2 = \mathbb{E}\left[\left(\sum_{t=1}^{k_n-s-1} \eta_{t,s}\right)^2\right]
\leq 2 \sum_{h=0}^{k_n-s-1} \sum_{t=1}^{k_n-s-h} \left|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]\right| = 2(R_{1,n} + R_{2,n}), \quad (A.90)$$

where (sums over empty sets are set to zero by convention)

$$R_{1,n} = \sum_{h=s}^{k_n-s-1} \sum_{t=1}^{k_n-s-h} \left|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]\right|, \quad R_{2,n} = \sum_{h=0}^{(k_n-s-1)\wedge (s-1)} \sum_{t=1}^{(k_n-s-1)\wedge (s-1) - s-h} \left|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]\right|.$$

By (A.86),

$$R_{1,n} \leq \left(\sum_{h=s}^{k_n-s-1} \frac{k_n-s-h}{k_n} \psi_{h-s}\right) c_n^4k_n^{-1} \leq \left(\sum_{h=0}^{\infty} \psi_h\right) c_n^4k_n^{-1}, \quad (A.91)$$

and similarly,

$$R_{2,n} \leq K\left(\sum_{h=0}^{(k_n-s-1)\wedge (s-1)} \frac{k_n-s-h}{k_n} (\psi_{h}^2 + \psi_{h}^2 + \psi_{s-h}^2)\right) c_n^4k_n^{-1} \leq K\left(\sum_{h=0}^{\infty} (\psi_{h}^2 + \psi_{h}^2 + \psi_{s-h}^2)\right) c_n^4k_n^{-1}. \quad (A.92)$$

Combining (A.90), (A.91) and (A.92), we deduce

$$\left|\frac{\Gamma_{X,n}^{(k,l)}(s) - E[\Gamma_{X,n}^{(k,l)}(s)]}{2}\right|^2 \leq K\left(\sum_{h=0}^{\infty} (\psi_{h}^2 + \psi_{h}^2 + \psi_{s-h}^2)\right) c_n^4k_n^{-1}.$$ 

Since $\sum_{h=0}^{\infty} \psi_h < \infty$ and $\sup_{s \geq 0} s\psi_s^2 < \infty$ under Assumption 5 and Assumption 7(iv), the assertion of the lemma follows from the above inequality.

Q.E.D.
Proof of Theorem 5. Recall that \( \Gamma_{X,n}(s) \equiv \mathbb{E}[X_n^T X_{n,t+s}] \). By definition, we can decompose

\[
\tilde{\Sigma}_n - \Sigma_n = \sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) + \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s) \Gamma_{X,n}(s).
\]  

(A.93)

To bound the first term on the right-hand side of (A.93), we note, by the triangle inequality,

\[
\left\| \sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) \right\| \leq \sum_{s=-k_n+1}^{k_n-1} |\mathcal{K}(s/M_n)| \left\| \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right\|.
\]  

(A.94)

By (A.85),

\[
\mathbb{E} \left[ \left\| \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right\| \right] \leq \left( \sum_{k=1}^{m_n} \sum_{l=1}^{m_n} \left\| \tilde{\Gamma}_{X,n}(k,l) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(k,l) \right] \right\|^2 \right)^{1/2} \leq Kc_n^2 m_n k_n^{-1/2}.
\]

Combining this estimate with (A.94), we deduce

\[
\mathbb{E} \left[ \left\| \sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) \right\| \right] \leq Kc_n^2 m_n k_n^{-1/2} \sum_{s=-k_n+1}^{k_n-1} |\mathcal{K}(s/M_n)| \leq Kc_n^2 m_n M_n k_n^{-1/2},
\]

where the second inequality follows from Assumption 6. From here, we deduce

\[
\sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n) \left( \tilde{\Gamma}_{X,n}(s) - \mathbb{E} \left[ \tilde{\Gamma}_{X,n}(s) \right] \right) = O_p(c_n^2 m_n M_n k_n^{-1/2}).
\]  

(A.95)

We now turn to the second term on the right-hand side of (A.93). By definition,

\[
\left\| \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s) \Gamma_{X,n}(s) \right\|^2 = \sum_{k=1}^{m_n} \sum_{l=1}^{m_n} \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s) \Gamma_{X,n}^{(k,l)}(s).
\]  

(A.96)

Let \( r = r_1 \wedge r_2 \). By Assumption 6, we can fix some (small) constant \( \varepsilon \in (0,1) \) such that

\[
\frac{1 - \mathcal{K}(x)}{|x|^r} \leq \frac{1 - \mathcal{K}(x)}{|x|^r} \leq K \text{ for } x \in [-\varepsilon, \varepsilon].
\]  

(A.97)
By the triangle inequality

\[
\left| \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s)\Gamma^{(k,l)}_{X,n}(s) \right| \leq k_n M_n^{-r} \sum_{|s| \leq M_n} \left| \frac{\mathcal{K}(s/M_n) - 1}{|s/M_n|^r} \right| |s|^r \left| \Gamma^{(k,l)}_{X,n}(s) \right| \]

(A.98)

\[
+ k_n \sum_{\varepsilon M_n < |s| < k_n} |\mathcal{K}(s/M_n) - 1| \left| \Gamma^{(k,l)}_{X,n}(s) \right|. \]

By (A.97),

\[
\sum_{|s| \leq M_n} \left| \frac{\mathcal{K}(s/M_n) - 1}{|s/M_n|^r} \right| |s|^r \left| \Gamma^{(k,l)}_{X,n}(s) \right| \leq K \sum_{|s| \leq M_n} |s|^r \left| \Gamma^{(k,l)}_{X,n}(s) \right| \leq K \sum_{s=-\infty}^{\infty} |s|^r \left| \Gamma^{(k,l)}_{X,n}(s) \right|. \]

(A.99)

Since \( \mathcal{K}(\cdot) \) is bounded (Assumption 6),

\[
\sum_{\varepsilon M_n < |s| < k_n} |\mathcal{K}(s/M_n) - 1| \left| \Gamma^{(k,l)}_{X,n}(s) \right| \leq K \sum_{\varepsilon M_n < |s| < k_n} \left| \Gamma^{(k,l)}_{X,n}(s) \right| \leq K M_n^{-r} \sum_{s=-\infty}^{\infty} |s|^r \left| \Gamma^{(k,l)}_{X,n}(s) \right|. \]

(A.100)

Combining (A.98), (A.99) and (A.100), we deduce

\[
\left| \sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s)\Gamma^{(k,l)}_{X,n}(s) \right| \leq K k_n M_n^{-r} \sum_{s=-\infty}^{\infty} |s|^r \left| \Gamma^{(k,l)}_{X,n}(s) \right| \leq K \bar{c}_n^2 M_n^{-r} \]

where the second inequality is by (A.82). By (A.96) and (A.101),

\[
\sum_{s=-k_n+1}^{k_n-1} (\mathcal{K}(s/M_n) - 1)(k_n - s)\Gamma_{X,n}(s) = O(\bar{c}_n^2 m_n M_n^{-r}). \]

(A.102)

The assertion of the theorem then follows from (A.93), (A.95) and (A.102).

Q.E.D.

**Proof of Theorem 6.** By Theorem 5,

\[
\left\| \tilde{\Sigma}_n - \Sigma_n \right\| = O_p(\bar{c}_n^2 m_n (M_n k_n^{-1/2} + M_n^{-r_1 + r_2})). \]

(A.103)

To prove the assertion of the theorem, it remains to show that

\[
\left\| \tilde{\Sigma}_n - \Sigma_n \right\| = O_p(M_n m_n^{1/2} \delta_{\theta,n}). \]

(A.104)
By the definitions of $\hat{\Gamma}_{X,n}(s)$ and $\tilde{\Gamma}_{X,n}(s)$, for any $s \geq 0$, we can decompose

$$
\hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) = k_n^{-1} \sum_{t=1}^{k_n-s} \left[ g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right] \left[ g(Z_{t+s}, \hat{\theta}_n) - g(Z_{t+s}, \theta_0) \right]^\top 
+ k_n^{-1} \sum_{t=1}^{k_n-s} \left[ g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right] g(Z_{t+s}, \theta_0)^\top 
+ k_n^{-1} \sum_{t=1}^{k_n-s} g(Z_t, \theta_0) \left[ g(Z_{t+s}, \hat{\theta}_n) - g(Z_{t+s}, \theta_0) \right]^\top.
$$

(A.105)

Therefore, by the triangle inequality and the Cauchy–Schwarz inequality,

$$
\max_{|s| \leq k_n-1} \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| 
\leq k_n^{-1} \sum_{t=1}^{k_n} \left\| g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right\|^2 
+ 2 \left( k_n^{-1} \sum_{t=1}^{k_n} \left\| g(Z_t, \hat{\theta}_n) - g(Z_t, \theta_0) \right\|^2 \right)^{1/2} \left( k_n^{-1} \sum_{t=1}^{k_n} \left\| g(Z_t, \theta_0) \right\|^2 \right)^{1/2}.
$$

(A.106)

By Assumption 8(ii) and Markov’s inequality,

$$
k_n^{-1} \sum_{t=1}^{k_n} \left\| g(Z_t, \theta_0) \right\|^2 = O_p(m_n).
$$

(A.107)

By Assumption 8(i), (A.106) and (A.107), we deduce

$$
\max_{|s| \leq k_n-1} \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| = O_p(m_n^{1/2} \delta_{\theta,n}).
$$

(A.108)

By the triangle inequality, (A.108) and Assumption 6(i), we deduce

$$
\left\| \hat{\Sigma}_n - \tilde{\Sigma}_n \right\| 
\leq \sum_{s=-k_n+1}^{k_n-1} |\mathcal{K}(s/M_n)| \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| 
\leq \max_{|s| \leq k_n-1} \left\| \hat{\Gamma}_{X,n}(s) - \tilde{\Gamma}_{X,n}(s) \right\| \sum_{s=-k_n+1}^{k_n-1} |\mathcal{K}(s/M_n)| 
= O_p(M_n m_n^{1/2} \delta_{\theta,n}).
$$

(A.109)

as claimed in (A.104). This finishes the proof. Q.E.D.

S.A.6 Proof of Proposition 1

Proof of Proposition 1. By (A.21), (A.37) and Assumption 2(ii, iii),

$$
\frac{n^{1/2} P(x)^\top (\hat{b}_n - b_n^\ast)}{\sigma_n(x)} = \frac{P(x)^\top Q_n^{-1} \tilde{N}_n}{\sigma_n(x)} + O_p(\delta_{1,n} + \delta_{2,n} + m_n^{1/2} \delta_{3,n}),
$$

(A.110)
where $\tilde{N}_n \sim \mathcal{N}(0, A_n)$ and $O_{pu}(\cdot)$ denotes a uniformly (in $x$) stochastically bounded sequence. By (A.39) and (A.110),

$$\frac{n^{1/2}(\hat{h}_n(x) - h(x))}{\sigma_n(x)} = \frac{P(x)^\top Q_n^{-1} \tilde{N}_n}{\sigma_n(x)} + O_{pu}(\delta_{1,n} + \delta_{2,n} + m_n^{1/2} \delta_{3,n})$$

$$= \frac{P(x)^\top \Sigma_n^{1/2} \tilde{N}_n^*}{\sigma_n(x)} + O_{pu}(\delta_{1,n} + \delta_{2,n} + m_n^{1/2} \delta_{3,n}), \quad (A.111)$$

where $\tilde{N}_n^* \equiv A_n^{-1/2} \tilde{N}_n$ is an $m_n$-dimensional standard normal random vector. Condition NS(i)(a) in Chernozhukov, Lee, and Rosen (2013) then follows from (A.111) and condition (ii) of the proposition.

By Assumption 2(ii), $\Sigma_n$ is positive definite, with eigenvalues bounded from above and away from zero uniformly in $n$. Since $p_1(x) = 1$, we have $\|P(x)\| \geq 1$ uniformly over $x \in \mathcal{X}$. By condition (i) of the proposition, $\sup_{x \in \mathcal{X}} \|P(x)\| \leq m_n^{1/2} \zeta_n$ with $(\zeta_n^2 m_n \log(n)/n)^{1/2} = o(1)$ under condition (ii) of the proposition. By the definition of $\zeta_n^L$,

$$\|P(x) - P(x')\| \leq \zeta_n^L \|x - x'\|$$

for any $x, x' \in \mathcal{X}$, where $\log(\zeta_n^L) = o(\log(n))$ under Assumption 2(i) and condition (ii) of the proposition. This verifies Condition NS(i)(b) in Chernozhukov, Lee, and Rosen (2013).

By (A.30) and the relation between the spectral norm and the Frobenius norm of matrices,

$$\left\|\tilde{\Sigma}_n - \Sigma_n\right\| \leq m_n^{1/2} \left\|\tilde{\Sigma}_n - \Sigma_n\right\|_F = O_p(m_n^{1/2}(\delta_{3,n} + \delta_{4,n})). \quad (A.112)$$

Condition NS(ii) in Chernozhukov, Lee, and Rosen (2013) then follows from (A.112) and condition (ii) of the proposition.

Q.E.D.

S.A.7 Proof of Proposition 2

PROOF OF PROPOSITION 2. Denote $\delta_{h,n} \equiv m_n^{1/2} n^{-1/2}$ for simplicity. By Assumptions 9(i,ii),

$$n^{-1} \sum_{t=1}^n w(Z_t, h_n) = n^{-1} \sum_{t=1}^n w(Z_t, h) + O_p(\delta_n) = O_p(\delta_{h,n}). \quad (A.113)$$
Let \( \mathcal{S}_n \equiv \{ b \in \mathbb{R}^{m_n} : \| b \| = 1 \} \). By Assumptions 9(iii,iv), we have uniformly over \( b \in \mathcal{S}_n \) and for any constant \( B > 0 \),

\[
\begin{align*}
&n^{-1} \sum_{t=1}^{n} b^\top w(Z_t, P(X_t)^\top (b_n^* + B\delta_{h,n}b)) \\
&= n^{-1} \sum_{t=1}^{n} b^\top w(Z_t, h_n) + b^\top n^{-1} \sum_{t=1}^{n} \mathbb{E} \left[ w(Z_t, P(X_t)^\top (b_n^* + B\delta_{h,n}b)) - w(Z_t, h_n) \right] + O_p(\tilde{\delta}_n) \\
&= n^{-1} \sum_{t=1}^{n} b^\top w(Z_t, h_n) + B\delta_{h,n}b^\top Q_n b + O_p(\tilde{\delta}_n). \tag{A.114}
\end{align*}
\]

Note that \( \tilde{\delta}_n = o(\delta_{h,n}) \). Hence, equation (A.114) further implies that

\[
\begin{align*}
&\mathbb{P} \left( \inf_{b \in \mathcal{S}_n} n^{-1} \sum_{t=1}^{n} b^\top w(Z_t, P(X_t)^\top (b_n^* + B\delta_{h,n}b)) > 0 \right) \\
&= \mathbb{P} \left( \inf_{b \in \mathcal{S}_n} \left( \delta_{h,n}^{-1} n^{-1} \sum_{t=1}^{n} b^\top w(Z_t, h_n) + Bb^\top Q_n b \right) + o(1) > 0 \right) \\
&\geq \mathbb{P} \left( \inf_{b \in \mathcal{S}_n} \delta_{h,n}^{-1} n^{-1} \sum_{t=1}^{n} b^\top w(Z_t, h_n) > -B\lambda_{\min}(Q_n) + o(1) \right). \tag{A.115}
\end{align*}
\]

By Assumption 9(v) and (A.113), as \( B \to \infty \),

\[
\liminf_{n \to \infty} \mathbb{P} \left( \inf_{b \in \mathcal{S}_n} n^{-1} \sum_{t=1}^{n} b^\top w(Z_t, P(X_t)^\top (b_n^* + B\delta_{h,n}b)) > 0 \right) \to 1. \tag{A.116}
\]

By (A.116) and the convexity of \( \rho(z, h) \) in \( h \), we deduce that

\[
\hat{b}_n - b_n^* = O_p(\delta_{h,n}). \tag{A.117}
\]

By (A.117), \( \| \hat{b}_n - b_n^* \| \leq \log(n)^{1/2} \delta_{h,n} \) with probability approaching one. Therefore, by (A.113) and Assumptions 9(iii,iv), we have uniformly over \( \alpha \in \mathcal{S}_n \),

\[
\begin{align*}
n^{-1} \sum_{t=1}^{n} \alpha^\top w(Z_t, \hat{h}_n) &= n^{-1} \sum_{t=1}^{n} \alpha^\top w(Z_t, h_n) \\
&\quad + n^{-1} \sum_{t=1}^{n} \alpha^\top \mathbb{E} [w(Z_t, P(X_t) b) - w(Z_t, h_n)] \big|_{b = b_n} + O_p(\tilde{\delta}_n) \\
&= n^{-1} \sum_{t=1}^{n} \alpha^\top w(Z_t, h) + \alpha^\top Q_n (\hat{b}_n - b_n^*) + O_p(\tilde{\delta}_n)
\end{align*}
\]

which together with (4.10) implies that uniformly over \( \alpha \in \mathcal{S}_n \),

\[
n^{-1} \sum_{t=1}^{n} \alpha^\top w(Z_t, h) + \alpha^\top Q_n (\hat{b}_n - b_n^*) = O_p(\tilde{\delta}_n). \tag{A.118}
\]

26
By Assumption 9(v), we can replace $\alpha$ in (A.118) with $Q_n^{-1} \alpha / \|Q_n^{-1} \alpha\|$ and deduce that

$$
\alpha^\top (\hat{b}_n - b_n^*) = -\alpha^\top Q_n^{-1} n^{-1} \sum_{t=1}^n w(Z_t, h) + O_p(\bar{\delta}_n)
$$

uniformly over $\alpha \in S_n$. The assertion of the proposition then readily follows. $Q.E.D.$

### S.A.8 Technical derivations of the conditional moment restriction in the search and matching model

In this appendix, we derive the conditional moment restriction (5.5) in the main text. Recall that the equilibrium is characterized by the following Bellman equations:

\begin{align}
J_p &= p - w_p + \delta (1 - s) \mathbb{E}_p [J_{p'}], \\
V_p &= -c_p + \delta q (\theta_p) \mathbb{E}_p [J_{p'}], \\
U_p &= z + \delta \left\{ f (\theta_p) \mathbb{E}_p [W_{p'}] + (1 - f (\theta_p)) \mathbb{E}_p [U_{p'}] \right\}, \\
W_p &= w_p + \delta \left\{ (1 - s) \mathbb{E}_p [W_{p'}] + s \mathbb{E}_p [U_{p'}] \right\},
\end{align}

the free entry condition $V_p = 0$ and the Nash bargaining solution

$$
J_p = (W_p - U_p) (1 - \beta)/\beta.
$$

Taking a difference between (A.121) and (A.122) yields

$$
W_p - U_p = w_p - z + \delta (1 - s - f (\theta_p)) \mathbb{E}_p [W_{p'} - U_{p'}].
$$

Combining (A.124) with (A.123), we derive

$$
J_p = \frac{1 - \beta}{\beta} (w_p - z) + \delta (1 - s - f (\theta_p)) \mathbb{E}_p [J_{p'}].
$$

From (A.119) and (A.125), we can solve for the wage function

$$
w_p = \beta p + (1 - \beta) z + \beta \delta f (\theta_p) \mathbb{E}_p [J_{p'}].
$$

Note that the free entry condition implies

$$
\delta q (\theta_p) \mathbb{E}_p [J_{p'}] - c_p = 0.
$$

Since $f (\theta)/q (\theta) = \theta$, we can rewrite (A.126) as

$$
w_p = \beta p + (1 - \beta) z + \beta \theta_p c_p.
$$

We can rewrite (A.127) as $\delta \mathbb{E}_p [J_{p'}] = c_p/q (\theta_p)$. Plugging this and (A.128) into (A.119), we deduce

$$
J_p = (1 - \beta) (p - z) - \beta \theta_p c_p + (1 - s) \frac{c_p}{q (\theta_p)}.
$$
Finally, plugging (A.129) into (A.127) yields
\[
\delta q(\theta_p) \mathbb{E}_p \left[ (1 - \beta) \left( p' - z \right) - \beta \theta_p c_p' + (1 - s) \frac{c_p'}{q(\theta_p')} \right] - c_p = 0. \tag{A.130}
\]

In standard calibration analysis, one can solve \( \theta_p \) from this equation, and then calibrate parameters by matching certain model-implied quantities (e.g., the average market tightness, the job finding rate, etc.) with their empirical counterparts.

From an econometric viewpoint, we consider (A.130) alternatively as a conditional moment restriction on observed data. Replacing \( p \) and \( \theta \) with their observed time series yields
\[
\delta q(\theta_t) \mathbb{E}_t \left[ (1 - \beta) \left( p_{t+1} - z \right) - \beta \theta_{t+1} c_{t+1} + (1 - s) \frac{c_{t+1}}{q(\theta_{t+1})} \right] - c_t = 0, \tag{A.131}
\]

where we write \( c_t \) in place of \( c_p_t \) (recall (A.127)) and use \( \mathbb{E}_t \) to denote the conditional expectation given the time-\( t \) information.\(^1\) The conditional moment restriction (5.5) is then obtained from (A.131).

### S.B Additional technical results

This appendix collects additional technical results. Section S.B.1 verifies Assumption 1(ii) under primitive sufficient conditions. Section S.B.2 provides an explicit rate of the strong approximation. Section S.B.3 provides examples for the strong approximation result under primitive conditions. Section S.B.4 provides primitive conditions for Assumption 2 in the main text.

#### S.B.1 Sufficient conditions for Assumption 1(ii)

We illustrate how to verify Assumption 1(ii) in the following proposition. The primitive conditions mainly require that the volatility \( V_{n,t} \) is weakly dependent, here formalized in terms of strong and uniform mixing coefficients.

**Proposition B1.** Suppose (i) \( V_{n,t} = v_t / k_n \) for some process \( (v_t)_{t \geq 0} \) taking values in \( \mathbb{R}^{m_n \otimes m_n} \) such that \( \sup_{t,j,l} \| v^{(j,l)}_t \|_q \leq c_n^2 \) for some constant \( q \geq 2 \) and some real sequence \( \bar{c}_n \); either (ii) \( q > 2 \) and \( v_t \) is strong mixing with mixing coefficient \( \alpha_k \) satisfying \( \sum_{k=1}^{k_n} \alpha_k^{1-2/q} < \infty \), or (iii) \( q = 2 \) and \( v_t \) is uniform mixing with mixing coefficient \( \phi_k \) satisfying \( \sum_{k=1}^{k_n} \phi_k^{1/2} < \infty \). Then, uniformly for all sequence \( h_n \) that satisfies \( h_n \leq k_n \),
\[
\left\| \sum_{t=1}^{h_n} (V_{n,t} - \mathbb{E}[V_{n,t}]) \right\|_2 = O(r_n), \quad \text{for} \quad r_n \equiv \frac{c_n^2 m_n k_n^{-1/2}}{}.
\tag{B.132}
\]

Consequently, condition (ii) of Assumption 1 holds provided that \( r_n = o(1) \).

---

\(^1\)Since the state process is Markovian, the time-\( t \) information set is spanned by \( p_t \), that is, \( \mathbb{E}_t [\cdot] = \mathbb{E}[\cdot | p_t] \).
Therefore, \( \bar{c}_n \) bounds the magnitude of the \( k_n^{1/2} X_{n,t} \) array. It is instructive to illustrate the “typical” magnitude of \( \bar{c}_n \) in the context of series estimation, where \( X_{n,t} \) is the score process given by \( X_{n,t} = u_t P(X_t) k_n^{-1/2} \). Suppose the \( \mathcal{F}_{n,t-1} \)-conditional \( 2q \)th moment of \( u_t \) is uniformly bounded. Then

\[
\|v_t^{(j,l)}\|_q \leq C \|p_j(X_t) p_t(X_t)\|_q.
\]

Therefore, \( \bar{c}_n = \mathcal{O}(m_n) \) if \( P(\cdot) \) collects trigonometric polynomials, and \( \bar{c}_n = \mathcal{O}(m_n^2) \) if \( P(\cdot) \) consists of splines, power series, or Legendre polynomials. In these cases, \( r_n = o(1) \) is implied by \( m_n \ll k_n^{1/2} \) and \( m_n \ll k_n^{1/4} \), respectively.

**Proof of Proposition B1.** We observe that

\[
\begin{align*}
\mathbb{E} & \left[ \left| \sum_{t=1}^{h_n} \left( V_{n,t}^{(j,l)} - \mathbb{E}[V_{n,t}^{(j,l)}] \right) \right|^2 \right] \\
& = \frac{1}{k_n^2} \mathbb{E} \left[ \sum_{t=1}^{h_n} \left( v_t^{(j,l)} - \mathbb{E}[v_t^{(j,l)}] \right)^2 \right] \\
& = \frac{1}{k_n^2} \sum_{s,t=1}^{h_n} \mathbb{E} \left[ \left( v_t^{(j,l)} - \mathbb{E}[v_t^{(j,l)}] \right) \left( v_s^{(j,l)} - \mathbb{E}[v_s^{(j,l)}] \right) \right] \\
& \leq \frac{2}{k_n^2} \sum_{k=0}^{h_n-1} \sum_{t=k+1}^{h_n} \left| \mathbb{E} \left[ \left( v_t^{(j,l)} - \mathbb{E}[v_t^{(j,l)}] \right) \left( v_{t-k}^{(j,l)} - \mathbb{E}[v_{t-k}^{(j,l)}] \right) \right] \right|.
\end{align*}
\]

We then prove the assertions for the \( \alpha \)-mixing and \( \phi \)-mixing cases separately.

**The \( \alpha \)-mixing case.** By the covariance inequality for strong mixing processes (see, e.g., Corollary 14.3 of Davidson (1994)),

\[
\left| \mathbb{E} \left[ \left( v_t^{(j,l)} - \mathbb{E}[v_t^{(j,l)}] \right) \left( v_{t-k}^{(j,l)} - \mathbb{E}[v_{t-k}^{(j,l)}] \right) \right] \right| \leq K \alpha_k^{-1/2} \|v_t^{(j,l)}\|_q \|v_{t-k}^{(j,l)}\|_q.
\]

Therefore, we can further bound the terms in (B.133) as follows

\[
\begin{align*}
\mathbb{E} & \left[ \left| \sum_{t=1}^{h_n} \left( V_{n,t}^{(j,l)} - \mathbb{E}[V_{n,t}^{(j,l)}] \right) \right|^2 \right] \\
& \leq \frac{K \alpha_k^{-1/2}}{k_n^2} \sum_{k=0}^{h_n-1} (h_n - k) \alpha_k^{-1/2} \leq K \bar{c}_n^4 k_n^{-1},
\end{align*}
\]

where the second inequality is due to condition (ii). The assertion of the proposition readily follows.

**The \( \phi \)-mixing case.** The case with uniform mixing can be proved similarly. Indeed, by the covariance inequality for uniform mixing processes (see, e.g., Corollary 14.5 of Davidson (1994)) and condition (iii),

\[
\begin{align*}
\mathbb{E} & \left[ \left| \sum_{t=1}^{h_n} \left( V_{n,t}^{(j,l)} - \mathbb{E}[V_{n,t}^{(j,l)}] \right) \right|^2 \right] \\
& \leq \frac{K \bar{c}_n^4}{k_n^2} \sum_{k=0}^{h_n-1} (h_n - k) \phi_k^{1/2} \left( \sup_{j,t} \|v_t^{(j,l)}\|_2 \right)^2 \leq K \bar{c}_n^4 k_n^{-1}.
\end{align*}
\]

From here, the assertion of the proposition for the \( \phi \)-mixing case readily follows.
S.B.2 Convergence rate of strong approximation

The strong approximation rate in (3.3) can be simplified under additional—but mild—assumptions. Corollary B1, below, provides a pedagogical example of this kind. We remind the reader that the typical order of each component of the (normalized) variable \( X_{n,t} \) is \( k_n^{-1/2} \), and hence it is reasonable to assume that its fourth moment is of order \( k_n^{-2} \).

**Corollary B1.** Under the same setting as Theorem 1, if \( \sup_{t,j} E[(X_{n,t}^{(j)})^4] = O(k_n^{-2}) \) holds in addition, then \( B_n = O(k_n^{-1/2} m_n^{3/2}) \). Consequently, \( \|S_n - \tilde{S}_n\| = O_p(m_n^{1/2} r_n^{-1/2} + m_n^{5/6} k_n^{-1/6}) \).

**Comment.** This corollary suggests that \( m_n \ll k_n^{1/5} \) is needed for the validity of the strong approximation. The dimension \( m_n \) thus cannot grow too fast relative to the sample size \( k_n \).

**Proof of Corollary B1.** We can bound \( B_n = \sum_{t=1}^{k_n} E[\|X_{n,t}\|^3] \) as follows:

\[
\sum_{t=1}^{k_n} E[\|X_{n,t}\|^3] \leq \sum_{t=1}^{k_n} \left( E[\|X_{n,t}\|^4]\right)^{3/4} \\
= \sum_{t=1}^{k_n} \left( \sum_{j=1}^{m_n} \sum_{l=1}^{m_n} E\left[\left(X_{n,t}^{(j)} X_{n,t}^{(l)}\right)^2\right]\right)^{3/4} \\
\leq \sum_{t=1}^{k_n} \left( \sum_{j=1}^{m_n} \left(E\left[\left(X_{n,t}^{(j)}\right)^4\right]\right)^{1/2}\right)^{3/2} \\
= O\left(k_n^{-1/2} m_n^{3/2}\right),
\]

where the first inequality is by Jensen’s inequality; the second inequality is by the Cauchy–Schwarz inequality and the last line follows from \( \sup_{t,j} E[(X_{n,t}^{(j)})^4] = O(k_n^{-2}) \). Plugging the estimate above into (3.3), we readily deduce the assertion of Corollary B1. \( \square \)

S.B.3 Examples for Theorem 4 under primitive conditions

Condition (ii) of Theorem 4 in the main text is high-level in nature in that it is stated for the approximating martingale difference \( X_{n,t}^* \) instead of for the underlying mixingale \( X_{n,t} \) directly. In this subsection, we provide two examples so as to illustrate how to verify this high-level condition under primitive conditions. The first example concerns linear processes and is relatively simple to describe.

**Example 1 (Martingale Approximation for Linear Processes).** Let \( (\varepsilon_{n,t}, F_{n,t}) \) be a martingale difference array such that \( \|\varepsilon_{n,t}\|_q \leq \varepsilon_n k_n^{-1/2} \) uniformly for some \( q \geq 3 \). Suppose that \( X_{n,t} \) is a linear process with the form \( X_{n,t} = \sum_{|j| < \infty} \theta_j \varepsilon_{n,t-j} \), where the coefficients \( (\theta_j) \) satisfy \( \sum_{|j| < \infty} |j \theta_j| < \infty \). Then \( (X_{n,t}) \) is an \( L_q \)-mixingale that satisfies Assumption 5 with
\( \psi_k = \sum_{|j| \geq k} |\theta_j| \) (see, e.g., Example 16.2 in Davidson (1994)); in particular, the summability condition \( \sum_{k=0}^{\infty} \psi_k < \infty \) is implied by \( \sum_{|j| < \infty} |j\theta_j| < \infty \). In this case, the martingale difference component \( X_{n,t}^* \) has a closed-form expression \( X_{n,t}^* = (\sum_{|j| < \infty} \theta_j) \varepsilon_{n,t} \), which verifies the conditions in Theorem 4 if and only if \( \varepsilon_{n,t} \) satisfies Assumption 1. In the simple case when \( \varepsilon_{n,t} \) has constant covariance matrix \( \Sigma_{\varepsilon} \), the pre-asymptotic covariance matrix of \( S_n^* \) is \( (\sum_{|j| < \infty} \theta_j)^2 \Sigma_{\varepsilon} \), which is exactly the long-run covariance matrix of \( X_{n,t} \); consequently, the third error term on the right-hand side of (4.4) is absent.

The second example, which concerns mixing-type primitive conditions, is slightly more complicated. In this example, we suppose that \( X_{n,t} = k_{n}^{-1/2} \varepsilon_t \), where \((\varepsilon_t)_{t=-\infty}^{\infty}\) is an \( m_n \)-dimensional zero mean strictly stationary (strong or uniform) mixing sequence with mixing coefficients \((\varphi_s)_{s=0}^{\infty}\). Let the filtration be defined as \( \mathcal{F}_{n,t} \equiv \sigma(\varepsilon_s : s \leq t) \). We consider the following regularity condition.

**Assumption B1.** (i) \( \sup_{t,j} ||\varepsilon_{(j)}^t||_k \leq c_{\kappa,n} \) where the sequence \( c_{\kappa,n} \) is bounded away from zero, \( \kappa > 5 \) for the strong mixing case and \( \kappa > 4 \) for the uniform mixing case; (ii) \( \sum_{s=0}^{\infty} \varphi_{s}^{(k-4)/(5k)} < \infty \) for the strong mixing case and \( \sum_{s=0}^{\infty} \varphi_{s}^{1/2} < \infty \) for the uniform mixing case; (iii) the eigenvalues of \( \Sigma_n \equiv \mathbb{E} [S_n S_n^\top] \) are bounded from above and away from zero; and (iv) \( c_{\kappa,n} m_n^{5/6} k_n^{-1/6} = o(1) \).

Assumption B1(i) imposes uniform moment bounds on \((\varepsilon_t)_{t=-\infty}^{\infty}\). Assumption B1(ii) restricts the level of dependence. Assumption B1(iii) requires that the covariance matrix \( \Sigma_n \) is non-degenerate. Assumption B1(iv) mainly restricts the rate at which the dimension of \( \varepsilon_t \) grows to infinity. Under this assumption, we can verify the conditions in Theorem 4 and obtain a strong approximation for \( S_n \), as stated by the following proposition.

**Proposition B2.** Under Assumption B1, we have

\[
\left\| S_n - \tilde{S}_n \right\| = O_p\left(c_{\kappa,n} m_n^{5/6} k_n^{-1/6}\right)
\]

where \( \tilde{S}_n \) is an \( m_n \)-dimensional random vector with distribution \( \mathcal{N}(0, \Sigma_n) \).

**Comment.** We can compare this strong approximation result with Theorem 1 of Dehling (1983). For example, assuming that the strong mixing coefficient converges to zero sufficiently fast and \( c_{\kappa,n} = O(1) \), (1.13) in Dehling (1983) implies that the strong approximation error converges at a rate that is slower than \( m_n^{11/6} k_n^{-1/900} \) (this is the best-case scenario obtained by setting \( d = m_n \), \( \delta = 2/3 \), \( \varepsilon = 1 \) and \( \rho_{2+\delta} = m_n \) in that paper). Evidently, the \( m_n^{5/6} k_n^{-1/6} \) rate implied by Proposition B2 improves significantly the rate derived in Dehling (1983).

**Proof of Proposition B2.** Step 1. In this step, we verify the conditions of Theorem 4. Condition (iii) of Theorem 4 coincides with Assumption B1(iii). It remains to verify conditions (i) and (ii) of that theorem.
We first show that Assumption 5 holds for the $X_{n,t}$ array (i.e., condition (i) of Theorem 4). Let $q = 5\kappa/(\kappa + 1)$ and $q = 4$ for the strong and the uniform mixing case, respectively. Then by Assumption B1(i) and the mixing inequality (see, e.g., Theorem 14.2 and Theorem 14.4 in Davidson (1994)),
\[
\left\| \mathbb{E} \left[ X_{n,t}^{(j)} \mid \mathcal{F}_{n,t-s} \right] \right\|_q \leq 6c_{\kappa,n}k_n^{-1/2}\varphi_1^{q-1/\kappa}
\]
(B.134)
in the strong mixing case, and
\[
\left\| \mathbb{E} \left[ X_{n,t}^{(j)} \mid \mathcal{F}_{n,t-s} \right] \right\|_q \leq 2c_{\kappa,n}k_n^{-1/2}\varphi_1^{-1/\kappa}
\]
(B.135)
in the uniform mixing case. Therefore, $(X_{n,t})_{t=-\infty}^{\infty}$ is an $L_q$-mixingale array with $\tilde{c}_n = 6c_{\kappa,n}$ and $\psi_s = \varphi_1^{q-1/\kappa}$ for the strong mixing case, and $\tilde{c}_n = 2c_{\kappa,n}$ and $\psi_s = \varphi_1^{-1/\kappa}$ for the uniform mixing case. It remains to check the summability condition $\sum_{s=0}^{\infty} \psi_s < \infty$; this holds under Assumption B1(ii) because $1/q - 1/\kappa = (\kappa - 4)/(5\kappa)$ for the strong mixing case and $1 - 1/\kappa > 1/2$ for the uniform mixing case.

We now verify condition (ii) of Theorem 4, that is, the approximating martingale difference $X_{n,t}^*$ satisfies Assumption 1. Note that $X_{n,t}^* = k_n^{1/2}\varepsilon_t^*$ where
\[
\varepsilon_t^* \equiv \sum_{s=-\infty}^{\infty} \{ \mathbb{E} [\varepsilon_{t+s} \mid \mathcal{F}_{n,t}] - \mathbb{E} [\varepsilon_{t+s} \mid \mathcal{F}_{n,t-1}] \}. \tag{B.136}
\]
We denote the conditional covariance matrix of $X_{n,t}^*$ by
\[
V_{n,t}^* = \mathbb{E} \left[ X_{n,t}^* X_{n,t}^{*\top} \mid \mathcal{F}_{n,t-1} \right] = k_n^{-1}v_t^*,
\]
where $v_t^* = \mathbb{E} [\varepsilon_t^* \varepsilon_t^{*\top} \mid \mathcal{F}_{n,t-1}]$. Since $\varepsilon_t$ is stationary, $(\varepsilon_t^*)_{t \geq 1}$ is also stationary. In particular,
\[
k_n \mathbb{E} [V_{n,t}^*] = \mathbb{E} [v_t^*] = \Sigma_n^*.
\]
(B.137)
Like (A.79), we can show that
\[
\|\Sigma - \Sigma_n^*\|_S = O_p(c_{\kappa,n}m_n^{1/2}k_n^{-1/2} + c_{\kappa,n}^2m_nk_n^{-1}) = o(1), \tag{B.138}
\]
where the second equality is due to Assumption B1(iv). Hence, Assumption B1(iii) implies that the eigenvalues of $\Sigma_n^*$ is bounded away from zero and from above. In view of (B.137), we see that $X_{n,t}^*$ satisfies Assumption 1(i). Finally, we can verify that $X_{n,t}^*$ satisfies Assumption 1(ii) by using Proposition B1, with
\[
r_n = c_{\kappa,n}^2m_nk_n^{-1/2}.
\]
(B.139)

**Step 2.** By the derivations in step 1, we can apply Theorem 4 to show that
\[
\left\| S_n - \mathbb{E} S_n \right\| = O_p(c_{\kappa,n}m_n^{1/2}k_n^{-1/2} + O_p(m_n^{1/2}k_n^{-1/2} + (B_n^*m_n)^{1/3}) + O_p(c_{\kappa,n}m_nk_n^{-1/2} + c_{\kappa,n}^2m_n^{3/2}k_n^{-1})).
\]

32
Following the same argument as in Corollary B1, we deduce $B_n^* = O(c_{k,n}^2 m_n^{3/2} k_n^{-1/2})$. Using this estimate and (B.139), we can simplify the error bound above as

$$\|S_n - \tilde{S}_n\| = O_p(c_{k,n} m_n^{5/6} k_n^{-1/6} + c_{k,n} m_n k_n^{-1/4} + c_{k,n}^2 m_3^{3/2} k_n^{-1}).$$

Under the maintained assumptions, $m_n \ll k_n^{1/5}$ and $c_{k,n} \ll k_n^{-1/6} m_n^{-5/6}$, which further imply that $\|S_n - \tilde{S}_n\| = O_p(c_{k,n} m_n^{5/6} k_n^{-1/6})$ as asserted. \(Q.E.D.\)

**S.B.4Primitive conditions for Assumption 2**

In this subsection, we illustrate how to verify Assumption 2 under the following primitive condition.

**Assumption B2.** (i) $(X_t^T, u_t)_t$ is a strictly stationary strong mixing process with mixing coefficient $(\varphi_s)_{s=0}^\infty$ satisfying $\sum \varphi_s^2 \varphi_s^2(s^{-4}/(5n))$ for some finite constants $\kappa > 5$ and $r_2 > 0$; (ii) the eigenvalues of $Q_n$ and $A_n$ are bounded from above and away from zero; (iii) $\mathbb{E}[|u_t|^\kappa |X_t|] \leq C < \infty$ almost surely for any $t$; (iv) $\max_1 \leq k \leq m_n \sup_x \|p_k(x)\| \leq \zeta_n$ where $\zeta_n$ is a non-decreasing positive sequence and $\log(\zeta_n) = O(\log(m_n))$; (v) there exist $\rho > 0$ and $b_n^* \in \mathbb{R}^m$ such that

$$\sup_{x \in \mathcal{X}} |P(x)^T b_n^* - h(x)| = O(m_n^{-\rho}).$$

(vi) $\inf_{x \in \mathcal{X}} \|P(x)\| \geq c$ for all $m_n$ and some constant $c > 0$; (vii) $(\zeta_n + m_n^{1/2})M_n \zeta_n m_n n^{-1/2} + n^{1/2} m_n^{-\rho_0} + \zeta_n^{6-12/\kappa} m_n^{5} n^{-1} = o(1)$ and $\zeta_n^2 m_n M_n^{-r_1/r_2} = o(1)$ where $r_1$ is the constant defined in Assumption 6.

Assumption B2(i) imposes restrictions on the serial dependence of the data. Assumption B2(ii), which is a standard regularity condition in the series estimation literature (see, e.g., Andrews (1991), Newey (1997), Chen (2007) and Belloni, Chernozhukov, Chetverikov, and Kato (2015)). Assumption B2(iii) imposes moment bound on the residual $u_t$, which is also standard. Assumption B2(iv) defines a uniform upper bound of the series basis functions $p_k(\cdot)$. Assumption B2(v) assumes that the unknown function $h(\cdot)$ can be approximated by $P(x)^T b_n^*$ with approximation error $O(m_n^{-\rho_0})$ under the uniform metric. Assumption B2(vi) holds trivially if the basis functions include the constant function. Assumption B2(vii) specifies the growth rate of $m_n$ and the bandwidth $M_n$ in the HAC estimation.

**Proposition B3.** Assumption 2 holds under Assumptions 6 and B2.

**Proof of Proposition B3.** First, $\log(\zeta_n^L) = O(\log(m_n))$ is maintained in Assumption B2(iv). By Assumption B2(v,vi),

$$\sup_{x \in \mathcal{X}} \frac{n^{1/2} |h(x) - P(x)^T b_n^*|}{\|P(x)\|} \leq O(n^{1/2} m_n^{-\rho_0}). \quad (B.140)$$
Therefore, Assumption 2(i) holds with 
\[ \delta_{1,n} = n^{1/2}m_n^{-\rho_h}, \]
where \( \delta_{1,n} = o(1) \) under Assumption B2(vii). Assumption 2(ii) is directly assumed in Assumption B2(ii). Assumptions 2(iii), 2(iv) and 2(v) have been verified in Lemma B7, Lemma B8 and Lemma B9, respectively; \( \delta_{j,n} = o(1) \), \( j \in \{2, 3, 4\} \), holds because of Assumption B2(vii).

Q.E.D.

For many popular approximating functions (e.g., power series, splines and trigonometric series), \( \zeta_n \leq C_0m_n^{1/2} \) where \( C_0 \) is a finite constant. Therefore, Assumption B2(vii) is reduced to
\[ M_n m_n^{-1/2} + n^{1/2}m_n^{-\rho_h} + m_n^{8-6/\kappa}n^{-1} = o(1) \text{ and } m_n^2M_n^{-r_1 \wedge r_2} = o(1). \]
The above restriction requires that \( m_n = o(n^{\kappa/(8\kappa-6)}) \) which becomes \( m_n = o(n^{1/7}) \) if \( \kappa = 6 \).

Let \( a_n \) be some sequence of positive numbers satisfying \( a_n \rightarrow \infty \). In i.i.d.
setting, Belloni, Chernozhukov, Chetverikov, and Kato (2015) imposes \( m_n = o(n^{1/5}a_n^{-6/5} \log(n))^{-2/5} \) for the splines or trigonometric series and \( m_n = o(n^{1/6}a_n^{-1} \log(n))^{-1/3} \) for the power series to invoke Yurinskii’s coupling in order to establish the uniform inference of the series estimator. Our restriction on \( m_n \) is stronger, because, in order to verify the sufficient conditions in Assumption 2, we use the strong approximation of the mixingale array, which requires stronger condition.

Lemma B7. Under Assumption B2, Assumption 2(iii) holds with \( \delta_{2,n} = \zeta_n 5/6 \). The assertion of Lemma B7 then follows from Proposition B2.

Proof of Lemma B7. We use Proposition B2 to prove this Lemma. For this purpose, it is sufficient to verify Assumption B1 with \( \varepsilon_t = u_t P(X_t) \) and \( k_n = n \). By Assumptions B2(ii) and B2(iii),
\[ \sup_{t,j} \| \varepsilon_t^{(j)} \| \leq C_{1/\kappa} \sup_{t,j} \| u_t p_j(X_t) \| \leq C_{1/\kappa} \lambda_{\text{max}}(Q_n)^{1/\kappa} \zeta_n^{1-2/\kappa} \]
which verifies Assumption B1(i) with \( c_{\kappa,n} = C_{1/\kappa} \zeta_n^{1-2/\kappa} \). Assumptions B1(ii), B1(iii) and B1(iv) are implied by Assumptions B2(i), B2(ii) and B2(vii), respectively. The assertion of Lemma B7 then follows from Proposition B2.

Q.E.D.

Lemma B8. Under Assumption B2, Assumption 2(iv) holds with \( \delta_{3,n} = m_n^{2}n^{-1} \).
PROOF OF LEMMA B8. Denote \( \eta_t(j,k) \equiv p_j(X_t)p_k(X_t) - \mathbb{E}[p_j(X_t)p_k(X_t)] \). By definition,

\[
\mathbb{E}\left[\left\| \tilde{Q}_n - Q_n \right\|^2 \right] = \mathbb{E}\left[\left\| n^{-1} \sum_{t=1}^{n} \left( P(X_t)P(X_t)^\top - \mathbb{E}\left[ P(X_t)P(X_t)^\top \right] \right) \right\|^2 \right]
\]

\[
= \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \mathbb{E}\left[ \left( n^{-1} \sum_{t=1}^{n} \eta_t(j,k) \right)^2 \right]
\]

\[
= n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \sum_{s=1}^{t-1} \mathbb{E}\left[ \eta_t(j,k) \eta_s(j,k) \right] + 2n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E}\left[ \eta_t(j,k) \eta_s(j,k) \right]. \tag{B.141}
\]

By Assumptions B2(iv),

\[
n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[ \eta_t(j,k) \right] \leq n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[ p_j(X_t)^2 p_k(X_t)^2 \right] \leq m_n^2 \zeta_n^4 n^{-1}. \tag{B.142}
\]

Since \((X_t)\) is strong mixing by Assumption B2(i), \((p_j(X_t)p_k(X_t))\) is also strong mixing with the same mixing coefficient \(\left( \varphi_i \right)_{i=0}^\infty\) for any \((j,k)\). Therefore, by the covariance inequality of the strong mixing process (see, e.g., Corollary 14.3 of Davidson (1994)) and Assumptions B2(iv),

\[
\mathbb{E}\left[ \eta_t(j,k) \eta_s(j,k) \right] \leq K \varphi_{t-s}^{-2/\kappa} \|p_j(X_t)p_k(X_t)\|_\kappa^2 \|p_j(X_s)p_k(X_s)\|_\kappa \leq K \varphi_{t-s}^{-2/\kappa} \zeta_n^4. \tag{B.143}
\]

By (B.143) and the summability condition of the mixing coefficients in Assumption B2(i),

\[
n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E}\left[ \eta_t(j,k) \eta_s(j,k) \right] \leq Km_n^2 \zeta_n^4 n^{-2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \varphi_{t-s}^{-2/\kappa} = O(m_n^2 \zeta_n^4 n^{-1}). \tag{B.144}
\]

From (B.141), (B.142) and (B.144), we deduce \(\mathbb{E}[\| \tilde{Q}_n - Q_n \|^2] = O(m_n^2 \zeta_n^4 n^{-1})\), which readily implies the assertion of Lemma B8.

Q.E.D.

**Lemma B9.** Under Assumptions 6 and B2, Assumption 2(v) holds with

\[
\delta_{4,n} = \zeta_n^2 m_n (M_n n^{-1/2} + M_n^{-r_1^2 r_2}) + M_n \zeta_n m_n^{1-\rho_n} + M_n \zeta_n m_n^{3/2} n^{-1/2}.
\]

**PROOF OF LEMMA B9.** Step 1. We use Theorem 6 to prove this lemma. In order to cast the setting into that of Theorem 6, we set \(k_n = n\), \(Z_t = (Y_t, X_t^\top)\), \(\theta_0 = h(\cdot)\), \(\hat{\theta}_n = \hat{h}_n(\cdot)\) and

\[
X_{n,t} = n^{-1/2} P(X_t) u_t, \quad \hat{X}_{n,t} = n^{-1/2} P(X_t) \left( Y_t - \hat{h}_n(X_t) \right). \tag{B.145}
\]
In this step, we verify that the $X_{n,t}$ array satisfies Assumption 7. Under Assumption B2, we can use the same arguments in the proof of Proposition B2 to show that the array $(X_{n,t})$ satisfies Assumption 5 with $\bar{c}_n = 6C^{1/\kappa} \varsigma_n$, $q = 5\kappa/(\kappa + 1)$ and $\psi_s = \varphi_s^{(\kappa-4)/(5\kappa)}$. It remains to verify conditions (i)–(iv) in Assumption 7.

By (3.5), $\mathbb{E} [X_{n,t}] = 0$ for any $t$ and any $n$. Moreover, by Assumption B2(i), $\mathbb{E} [X_{n,t}X_{n,t+j}] = \kappa^{-1} \mathbb{E} [u_t u_{t+j} P(X_t)P(X_{t+j})^\top]$ only depends on $n$ and $j$. Therefore, Assumption 7(i) holds. Let $F_{n,t}$ be the $\sigma$-field generated by $\{X_s, u_{s-1} \}_{s \leq t}$. We can use the same argument in the proof of Theorem 14.2 of Davidson (1994) to deduce that

$$\left\| \mathbb{E} \left[ X_{n,t}^{(l)}X_{n,t+j}^{(k)} \right] - \mathbb{E} \left[ X_{n,t}^{(l)}X_{n,t}^{(k)} \right] \right\| \leq 6\varphi_s^{1/2 - 2/\kappa} \left\| X_{n,t}^{(l)}X_{n,t+j}^{(k)} \right\|^{\kappa/2}. \tag{B.146}$$

By the definition of $X_{n,t}^{(l)}$ and $X_{n,t+j}^{(k)}$, and Assumptions B2(iii) and B2(iv),

$$\left\| X_{n,t}^{(l)}X_{n,t+j}^{(k)} \right\|^{\kappa/2} \leq \left\| u_t u_{t+j} \right\|^{\kappa/2} c_n^{s-1} \left( C^{2/\kappa} c_n^{s} \right)^{2n-1} \leq \epsilon_n^{2n-1}, \tag{B.147}$$

which verifies Assumption 7(iii). Furthermore, this estimate and (B.146) imply that

$$\left\| \mathbb{E} \left[ X_{n,t}^{(l)}X_{n,t+j}^{(k)} \right] - \mathbb{E} \left[ X_{n,t}^{(l)}X_{n,t}^{(k)} \right] \right\| \leq 6C^{2/\kappa} \varsigma_n^{1/2 \kappa} \epsilon_n^{2n-1} \leq \epsilon_n^{2n-1} \varphi_{s}^{(\kappa-2)/(2\kappa)}. \tag{B.148}$$

Since $(\kappa - 4)/(5\kappa) \leq (\kappa - 2)/(2\kappa)$, this estimate implies Assumption 7(ii) with $\psi_s$ defined as above.

Finally, we verify Assumption 7(iv). Under Assumption B2(i), $\psi_s$ is summable. Hence, there exists a finite $\bar{s}$ such that $\psi_s \leq s^{-1}$ for any $s \geq \bar{s}$; otherwise, we could extract a subsequence from $\psi_s$ that is not summable. Therefore,

$$\sup_{s \geq 0} s^{\bar{s}} \psi_s^2 \leq 1 + \max_{0 \leq s \leq \bar{s}} s \psi_s^2 < \infty.$$

Further note that $\sum_{s=0}^{\infty} s^{\bar{s}} \psi_s < \infty$ holds by Assumption B2(i). This verifies Assumption 7(iv).

**Step 2.** In this step, we finish the proof of Lemma B9 by verifying Assumption 8 for which we note from (B.145) that the $g(\cdot)$ function is defined implicitly as $g(Z_t, h) = (Y_t - h(X_t)) P(X_t)$. Hence, by Assumption B2(iv),

$$n^{-1} \sum_{t=1}^{n} \left\| g(Z_t, \hat{h}_n) - g(Z_t, h) \right\|^2 = n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h(X_t) \right)^2 P(X_t)^\top P(X_t) \leq \epsilon_n^2 m_n \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h_0(X_t) \right)^2. \tag{B.149}$$

By Lemma B8 and Assumption B2(vii), $\| \hat{Q}_n - Q_n \| = o_p(1)$. Hence,

$$\lambda_{\min}^{-1}(\hat{Q}_n) + \lambda_{\max}(\hat{Q}_n) \leq K, \text{ with probability approaching one.} \tag{B.150}$$

Define $P_n$, $U_n$, $H_n$ and $H_n^*$ as in the proof of Theorem 2. Like (A.24) and (A.26), we can show that

$$\hat{b}_n - b_n^* = (P_n^\top P_n)^{-1} \left( P_n^\top U_n \right) + (P_n^\top P_n)^{-1} P_n^\top (H_n - H_n^*), \tag{B.151}$$

36
and \( n^{-1/2} P_n^T U_n = O_p(m_{1/2} n). \) Then, by (B.150),
\[
\left\| (P_n^T P_n)^{-1} (P_n^T U_n) \right\| \leq \lambda_{\min}^{-1}(\hat{Q}_n) \left\| n^{-1/2} P_n^T U_n \right\| = O_p(m_{1/2} n^{-1/2}). \tag{B.152}
\]

By Assumption B2(v) and (B.150),
\[
\left\| (P_n^T P_n)^{-1} P_n^T (H_n - H_n^*) \right\|^2 \leq \lambda_{\min}^{-1}(\hat{Q}_n) n^{-1} \left\| H_n - H_n^* \right\|^2 = O_p(m^{-2\rho_h}). \tag{B.153}
\]

By (B.151), (B.152) and (B.153),
\[
\left\| \hat{b}_n - b_n^* \right\| = O_p(m_{1/2} n^{-1/2} + m_n^{-\rho_h}). \tag{B.154}
\]

By Assumption B2(v), (B.150) and (B.154),
\[
n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - h(X_t) \right)^2 \leq 2n^{-1} \sum_{t=1}^{n} \left( \hat{h}_n(X_t) - P(X_t)^T b_n^* \right)^2 + 2n^{-1} \sum_{t=1}^{n} \left( P(X_t)^T b_n^* - h(X_t) \right)^2
\]
\[
\leq 2\lambda_{\max}(\hat{Q}_n) \left\| \hat{b}_n - b_n^* \right\|^2 + 2 \sup_{x \in \mathcal{X}} \left\| P(x)^T b_n^* - h(x) \right\|^2
\]
\[
= O_p(m^{-2\rho_h} + m_n n^{-1}). \tag{B.155}
\]

Combined with (B.149), this estimate further implies that
\[
n^{-1} \sum_{t=1}^{n} \left\| g(Z_t, \hat{h}_n) - g(Z_t, h) \right\|^2 = O_p(\zeta_n^2 m_{1/2} n^{-2\rho_h} + \zeta_n m_n n^{-1}), \tag{B.156}
\]

which verifies Assumption 8(i) with \( \delta_{\theta,n} = \zeta_n m_{1/2} n^{-2\rho_h} + \zeta_n m_n n^{-1/2}. \) By Assumption B2(ii,iii),
\[
\left\| g(Z_t, h) \right\|^2 \leq \mathbb{E} \left[ u_t^2 P(X_t)^T P(X_t) \right] \leq C^{2/\kappa} \text{Tr}(Q_n) \leq Km_n,
\]
which implies Assumption 8(ii). This finishes the proof. \( Q.E.D. \)

**References**


